This document was produced
by scanning the original publication.

Ce document est le produit d'une
numérisation par balayage
de la publication originale.

INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING
INTERNATIONALE GESELLSCHAFT FÜR PHOTOGRAMMETRIE UND FERNERKUNDUNG
SOCIÉTÉ INTERNATIONALE DE PHOTOGRAMMÉTRIE ET DE TÉLÉDÉTECTION

isprs

information from imagery

WITHDRAWN

THE INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES
INTERNATIONALES ARCHIV FÜR PHOTOGRAMMETRIE, FERNERKUNDUNG UND RAUMBEZOGENE INFORMATIONSWISSENSCHAFTEN
ARCHIVES INTERNATIONALES DES SCIENCES DE LA PHOTOGRAMMÉTRIE, DE LA TÉLÉDÉTECTION ET DE L'INFORMATION SPATIALE

ISSN 1682-1750

VOLUME
VOLUME
BAND
XXXIV
PART
TOME
TEIL
4
COMMISSION
COMMISSION
KOMMISSION
IV

PROCEEDINGS OF THE ISPRS COMMISSION IV SYMPOSIUM

Geospatial Theory, Processing and Applications

8 - 12 JULY 2002
OTTAWA, CANADA

Editors
C. Armenakis, Y.C. Lee

Published by:

Natural Resources Canada
Ressources naturelles Canada

In cooperation with:

CANADIAN INSTITUTE OF GEOMATICS
ASSOCIATION CANADIENNE DES SCIENCES GEOMATIQUES

INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING
INTERNATIONALE GESELLSCHAFT FÜR PHOTOGRAMMETRIE UND FERNERKUNDUNG
SOCIÉTÉ INTERNATIONALE DE PHOTOGRAMMÉTRIE ET DE TÉLÉDÉTECTION

isprs
information from imagery

THE INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES
INTERNATIONALES ARCHIV FÜR PHOTOGRAMMETRIE, FERNERKUNDUNG UND RAUMBEZOGENE INFORMATIONSWISSENSCHAFTEN
ARCHIVES INTERNATIONALES DES SCIENCES DE LA PHOTOGRAMMÉTRIE, DE LA TÉLÉDÉTECTION ET DE L'INFORMATION SPATIALE

ISSN 1682-1750

VOLUME VOLUME BAND	XXXIV	PART TOME TEIL	4	COMMISSION COMMISSION KOMMISSION	IV

PROCEEDINGS OF THE ISPRS COMMISSION IV SYMPOSIUM

Geospatial Theory, Processing and Applications

8 - 12 JULY 2002
OTTAWA, CANADA

Editors
C. Armenakis, Y.C. Lee

Published by:
Natural Resources Canada
Ressources naturelles Canada

In cooperation with:
CANADIAN INSTITUTE OF GEOMATICS
ASSOCIATION CANADIENNE DES SCIENCES GÉOMATIQUES

Copies of this book are available from:

GITC bv
P.O. Box 112
8530 AC Lemmer
The Netherlands

Fax: +31-514-56185

ISSN 1682-1750

© 2002 resides with the author(s) of a paper except where it is retained by the author's employer.

This book is published by the:

Centre for Topographic Information
Mapping Services Branch, Geomatics Canada
Department of Natural Resources Canada
Ottawa, Canada

In cooperation with the:

Canadian Institute of Geomatics
Ottawa, Canada

PREFACE

This volume contains the proceedings of the papers submitted to the Symposium of ISPRS Technical Commission IV: Spatial Information Systems and Digital Mapping. The Symposium is held in Ottawa Canada 8-12 July 2002, jointly with the 10th Spatial Data Handling Symposium and the 95th Annual Conference of the Canadian Institute of Geomatics. The proceedings of all the papers of this international joint geo-spatial event are distributed to the participants on a CD-ROM.

Today geo-information is used in many areas and in many ways both locally and globally. Its use has been expanded beyond traditional spatial applications and mapping. We see it being used to support sustainable development, in location based services, geo-spatial infrastructures, resource management, environmental databases and assessments, tourism, agriculture, disaster relief, communications, insurance, real estate, and entertainment just to mention few, while issue-oriented approaches are gaining momentum. At the same time, the complexity of spatial applications has significantly increased requiring knowledge-based solutions. Thus, is becoming apparent that the separation of geo-spatial applications from their theoretical aspects is not leading to optimal solutions. Therefore, Commission IV has grown and encompasses these days the studies of the geo-spatial concepts and theory in addition to strictly geo-applications. The theme of the Symposium "Geo-spatial Theory, Processing and Applications" was chosen to reflect these changes.

Throughout the papers submitted to the Symposium of Commission IV we truly see the integration of photogrammetry, remote sensing and spatial information sciences, thus implementing the tag line of ISPRS "Information from Imagery". The papers in this volume cover an extensive list of topics ranging from fundamentals of spatio-temporal spaces and databases, 3D GIS, urban and population management information systems, uncertainty, hazard and environmental mapping, transportation and location-based services, tourism, data warehouses, interoperability, web-based and mobile mapping applications, realization of federated databases, generalization, spatial infrastructure architectures, on-line data distribution and web-based catalogue services, image-based geospatial databases including queries and video systems, landscape modelling, terrain visualization, DEM and DSM, mapping advances using IKONOS, QuickBird, EROS, SRTM, LIDAR and InSAR, integration of data and methods, image fusion techniques, image segmentation and feature extraction, change detection and updating of databases, to mapping of Mars and other celestial bodies. We hope that the contents of this volume will enrich and advance our knowledge in our field and will be a valuable source of information.

We wish to sincerely thank all the authors for contributing their papers, the officers of ISPRS Technical Commission IV for their contributions to the technical program, and the local organizing committee for making this Symposium reality. We also thank Florin Savopol for his continuous assistance in the organization of the Symposium. The support provided by the Department of Natural Resources Canada (NRCan) is gratefully acknowledged.

Costas Armenakis and Y C Lee
Ottawa, Canada, July 2002

ISPRS Council 2000 - 2004

President	John C. Trinder	Australia
Secretary General	Ian J. Dowman	United Kingdom
Congress Director	M. Orhan Altan	Turkey
Treasurer	Ammatzia Peled	Israel
First Vice president	Lawrence W. Fritz	U.S.A.
Second Vice President	Gérald Begni	France

ISPRS Technical Commission IV
Spatial Information Systems and Digital Mapping

President – Costas Armenakis *(Canada)* Scientific Secretary - Yuk-Cheung Lee *(Canada)*

WG IV/1: Spatial and Temporal Data Modelling and Analysis
Chair: Yvan Bédard *(Canada)*
Co-Chair: Wenzhong (John) Shi *(Hong Kong)*

WG IV/2: Federated Databases and Interoperability
Chair: Jianya Gong *(China)*
Co-Chair: Rolf A. de By *(The Netherlands)*

WG IV/3: Data Generalization and Data Mining
Chair: Monika Sester *(Germany)*
Co-Chair: Dianne Richardson *(Canada)*

WG IV/4: Spatial Data Infrastructures
Chair: Parth Sarathi Roy *(India)*
Co-Chair: David Holland *(United Kingdom)*
Secretary: Shefali Agrawal *(India)*

WG IV/5: Image-Based Geospatial Databases
Chair: Peggy Agouris *(U.S.A.)*
Co-Chair: Dimitris Papadias *(Hong Kong)*

WG IV/6: Landscape Modelling and Visualization
Chair: Marguerite Madden *(U.S.A.)*
Co-Chair: Jochen Schiewe *(Germany)*
Secretary: Thomas Jordan *(U.S.A.)*

WG IV/7: Data Integration and Digital Mapping
Chair: Michael Hahn *(Germany)*
Co-Chair: Ryosuke Shibasaki *(Japan)*

WG IV/8: Global Environmental Databases
Chair: Ryutaro Tateishi *(Japan)*
Co-Chair: David Hastings *(U.S.A.)*

WG IV/9: Global Environmental Databases
Chair: Randolph L. Kirk *(U.S.A.)*
Co-Chair: Jan-Peter Muller *(United Kingdom)*
Secretary: Mark Rosiek *(U.S.A.)*

IC WG II/IV: Systems for Automated Geo-spatial Data Production and Updating from Imagery
Chair: Christian Heipke *(Germany)*
Co-Chair: Ammatzia Peled *(Israel)*

TABLE OF CONTENTS

WG IV/1: SPATIAL AND TEMPORAL DATA MODELLING AND ANALYSIS

Spatial Urban Model for Environmental Planning
Hari Shanker Gupta 3

K-Order Spatial Neighbours Based on Voronoi Diagram: Description, Computation and Applications
Renliang Zhao, Jun Chen, Zhi-lin Li 10

An Interactive System for Producing Polyhedral and Prismatic Building Models using Building Outline Segments
Jiann-Yeou Rau, Liang-Chien Chen 16

Remote Sensing and Modeling in Regional Land Cover Change Study
Y.Q. (Yeqiao) Wang 22

An Inverse Analysis of unobserved Trigger Factors of the Slope Failures based on Structural Equation Modeling
Hirohito Kojima, Shigeyuki Obayashi 27

A Data Structure for Spatio-Temporal Information Management
Yutaka Ohsawa, Wei Guo, Masayuki Sakurai 33

MapManager: The Design of a Com-Based GIS Component-ware
Xiaojun Tan, Fuling Bian 39

A Multi-Level Approach for 3D Modelling in Geographical Information Systems
Fabien Ramos 43

Data Integration using Weights of Evidence Model: Applications in Mapping Mineral Resource Potentials
H. Wang, G. Cai, Q. Cheng 48

Fuzzy Logic Integration for Landslide Hazard Mapping using Spatial Data from Boeun, Korea
Kwang-Hoon Chi, No-Wook Park, Chang-Jo Chung 54

The Development of a Spatial-Temporal Analysis System Prototype for Rice Inventory Data in Taiwan
Jung-Hong Hong, Zeal Su 60

3D GIS: Current Status and Perspectives
Siyka Zlatanova, Alias Rahman, Morakot Pilouk 66

An Empirical Evaluation of Hedonic Regression Models
Xiaolu Gao, Yasushi Asami, Chang-Jo Chung 72

GIS Data Maintenance and Management with Spatio-Temporal Model
Wei Lu, Takeshi Doihara 78

Fuzzy based Spatial Query and Analysis in an Urban Information System
M. Sarpoulaki, F. Samadzadegan, R. Ababspour 82

Handling Temporal Uncertainty in GIS Domain: A Fuzzy Approach
Jonathan Li, Simon Wu, Gordon Huang 83

A New Multi-Scale Modeling Approach for Space/Time Random Field Estimation
Kyung-Mee Choi, George Christakos, Mark L. Wilson 88

Statistical and Geo-Statistical Analysis of Wind: A Case Study of Direction Statistics
Tetsuya Shoji 94

A General Object-Oriented Spatial Temporal Data Model
Bonan Li, Guoray Cai 100

Modelling Turning Restrictions in Traffic Networks for Vehicle Navigation System
Jie Jiang, Gang Han, Jun Chen 106

Investigation on the Concept Model of Mobile GIS
Luqun Li, Chengming Li, Zongjian Lin 111

Random Topological Relations between Line and Area in GIS
Deng Min, Chengming Li, Chen Xiaoyong 113

GIS Applications using Agent Technology
Nadia Shahriari , C. Vincent Tao 118

Modeling the Conditional Probability of the Occurrence of Future Landslides in a Study Area characterized by Spatial Data
Chang-Jo Chung, Andrea G. Fabbri 124

Respecting Hierarchically Structured Taxonomies in Supervised Image Classification: A Geological Case Study from the Western Canadian Shield
E. M. Schetselaar, C. F. Chung 132

The Study on Application Mode of GIS Assistant Demarcation System
Guizhi Wang 138

Modeling Multi-dimensional Spatio-Temporal Data Warehouses in a Context of Evolving Specifications
M. Miquel, Y. Bédard, A. Brisebois, J. Pouliot, P. Marchand, J. Brodeur 142

Integrated Dynamic Model for Multi-Resolution Data in Three-Dimensional GIS
Wenzhong Shi, Bisheng Yang, Qingquan Li 148

Predictability of Individual Mass Movements from Spatial Databases for Hazard Mapping
Andrea G. Fabbri, Chang-Jo Chung 154

WG IV/2: FEDERATED DATABASES AND INTEROPERABILITY

Integration of Spatial Data within a Generic Platform for Location-Based Applications
Steffen Volz, Jan-Martin Bofinger 157

Investigation on Application of GIS in Population Management
Chengming Li, Zongjian Lin, Jie Yin 163

Update of the Canadian Road Network in a Distributed Geo-Spatial Data Framework
Denis Boutin, Marcel Sabourin 166

Integration and Access of Multi-Source Vector Data
Dan Edwards, Justin Simpson 167

Increasing Interoperability with CEONet Technology using WSDL and SOAP
Nickolas Grabovac 175

Using Schema and Data Integration Technique to Integrate Spatial and Non-Spatial Data: Developing Populated Places DB Of Turkey (PPDB_T)
Abdulvahit Torun 179

On-Line Distribution of Digital Topographic Data
Michel Gilbert 185

Geo-Spatial Data Handling for Web-Based and Mobile Applications
Michela Bertolotto, James Carswell, Liam Mcgeown, Peter Thijs 190

Extending Geo-Spatial Repositories with Geo-Semantic Proximity Functionalities to Facilitate the Interoperability of Geo-Spatial Data
Jean Brodeur, Yvan Bédard 196

Facilitating Interdisciplinary Sciences by the Integration of a CLOSi based Database with Bio-Metadata
J. L. Campos Dos Santos, R. A. De By, P.M.G. Apers, C. Magalhães 202

Concept and Realization of Federated Spatial Database
J. Gong, H. Xiong, Y. Wang, L. Shi 208

Research of Mobile GIS Application based on Handheld Computer
Liu Yong, Li Qing Quan, Xie Zhi Ying, Wang Chong 212

Semantic Heterogeneity of Geodata
Z. Xu, Y.C. Lee 216

WG IV/3: DATA GENERALIZATION AND DATA MINING

Generalization of 3D Building Data based on Scale-Spaces
Andrea Forberg, Helmut Mayer 225

Comprehensive Knowledge Discovery: Theory, Concept And Application
Sha Zong-Yao, Bian Fu-Ling 231

A Variable- Scale Map for Small-Display Cartography
Lars Harrie, L. Tiina Sarjakoski, Lassi Lehto 237

Automatic Generalization of 3D Building Models
Martin Kada 243

Automatic Derivation of Location Maps
Birgit Elias 249

Spatial Generalization: An Adaptive Lattice Model based on Spatial Resolution
Pingtao Wang, Takeshi Doihara, Wei Lu 255

An Adaptive Lattice Model and its Application to Map Simplification
Takeshi Doihara, Pingtao Wang, Wei Lu 259

Finding Spatial Units for Land Use Classification based on Hierarchical Image Objects
Qingming Zhan, Martien Molenaar, Klaus Tempfli 263

Formalising the Geographic Database Generalization Processb Means of a Conflicts/Operations Graph
Dominique Han-Sze-Chuen, Sébastien Mustière, Bernard Moulin 269

What is Spatial Context in Cartographic Generalization?
Sébastien Mustière, Bernard Moulin 274

Semantic Similarity Evaluation Model in Categorical Database Generalization
Liu Yaolin, Martien Molenaar, Menno-Jan Kraak 279

Generalization of 3D Building Data
Frank Thiemann 286

Application Dependent Generalization - The Case of Pedestrian Navigation
Monika Sester 291

Generalization of Building Polygons extracted from IKONOS Imagery
Jie Shan, D. Scott Lee 297

WG IV/4: SPATIAL DATA INFRASTRUCTURES

Design, Modeling and Data Preparation of Mines And Metals Geographical Information System
Ali Mansourian, Mohammad Javad Valadan Zoej 305

An On-Line Service to Access and Visualize the Data Alignment Layer
Babak Ameri, Ali Shankaie, Robert Orth 308

Canadian Geo-Spatial Standards in Action
Kian Fadaie, Mohammed Habbane, Cecilia Tolley 313

An Efficient Data Management Approach for Large Cyber-City GIS
Qing Zhu, Xuefeng Yao, Duo Huang, Yetin Zhang 319

Extending Geo-Information Services: A Virtual Architecture for Spatial Data Infrastructures
Javier Morales, Mostafa Radwan 324

Integrated Web Catalogue Service (IWCATS)
Serge Kéna-Cohen, Edric Keighan, Guy Rochon 331

Incremental Data Evolution - Complementary Approach to Schema Evolution of Organisation's Large Operational GIS Databases
Leena Salo-Merta 335

Developing a National Height Database
David Holland 341

Modeling Multiple Representation into Spatial Data Warehouses: A UML-based Approach
Yvan Bédard, Marie-Josée Proulx, Suzie Larrivée, Eveline Bernier 344

U.S. Geological Survey Spatial Data Access
J. Faundeen, R. Kanengieter, M. Buswell 350

Geospatial Info-Mobility Services –A Challenge for National Mapping Agencies
T. Sarjakoski, L. T. Sarjakoski, L. Lehto, M. Sester, A. Illert,
F. Nissen, B. Rystedt, R. Ruotsalainen 356

Development of Biodiversity Information System for North East India using Internet GIS
P.S. Roy, S. Saran, S. Ghosh, N. Prasad, G. Talukdar 361

WG IV/5: IMAGE-BASED GEO-SPATIAL DATABASES

Modeling Middle Urban Population Density with Remote Sensing Imagery
An-min Lu, Cheng-ming Li, Zong-jian Lin 371

Robust Haze Removal: An Integral Processing Component in Satellite-Based Land Cover Mapping
Bert Guindon, Ying Zhang 375

Segmentation of High-Resolution Remotely Sensed Data - Concepts, Applications and Problems
Jochen Schiewe 380

A Physics-Based Model for Active Contours: A Computational Algebraic Topology Approach
P. Poulin, M.F. Auclair-Fortier, D. Ziou, M. Allili 386

Design and Implementation of the High-Performance 3D Digital Landscape Server 'DILAS'
Stephan Nebiker 391

Searching And Browsing Interface For Image Databases Via Internet
Zeal Su, Jung-Hong Hong 395

Estimating Vehicle Activity using Thermal Images Sequences and Maps
Uwe Stilla, Eckart Michaelsen 401

Automated Techniques for Satellite Image Segmentation
Alberto Guarnieri, Antonio Vettore 406

Query Functionality for 3D Visual Databases
Thomas Hilker, Michael Hahn, Dietrich Schröder 411

On The Use Of Hierarchies And Feedback For Intelligent Video Query Systems
Kristin Eickhorst, Peggy Agouris 412

WG IV/6: LANDSCAPE MODELLING AND VISUALIZATION

High Resolution Digital Surface Model (DSM) Generation using Multi-View Multi-Frame Digital Airborne Images
B. Ameri, N. Goldstein, H. Wehn, A. Moshkovitz, H. Zwick 419

Design and Implementation of a GIS based Bicycle Routing System for the World Wide Web (WWW)
Manfred Ehlers, Stefan Jung, Katrin Stroemer 425

Physical Models of GIS Objects by Rapid Prototyping
Wolf-Dieter Rase 430

Tridimensional Model For A University Campus Using DTM
Claudio Marcio Penna de Moraes, Luiz Alberto Vieira Dias 434

Multimedia Visualisation of Geo-Information for Tourism Regions based on Remote Sensing Data
Alexander Almer, Harald Stelzl 436

Optimum Spatial Resolution of Digital Elevation Model for Topographical Analysis
Masataka Takagi, Hiroshi Asano, Yuki Kikuchi 442

Development of a Method Network for Object Recognition using Digital Surface Models
Guido Bohmann, Jochen Schiewe 447

Natural Colour Urban 3D Modelling: A Stereoscopic Approach with IKONOS Multispectral and Panchromatic Data
Yun Zhang 453

3D-City Model Supporting for CCTV Monitoring System
Ying Ming, Jingjue Jiang, Fuling Bian 456

Use of GIS and Multi-Temporal Imaging Spectrometer Data for Modelling and Mapping Environmental Changes in Mining Areas
Christian Fischer 460

Segmentation of TIN-Structured Surface Models
Ben Gorte 465

Analysis of SRTM DTM - Methodology and Practical Results
A. Koch, C. Heipke, P. Lohmann 470

Interactive Analysis for 3D-GIS Tools
Désiree Hilbring 476

Terrain Visualization applied to Vegetation Studies of National Parks and Historic Sites
Roy Welch, Thomas Jordan, Marguerite Madden 481

Densification of DTM Grid Data using Single Satellite Imagery
Mohammad Rajabi, Rod Blais 482

Use of AIRSAR / JERS-1 SAR Datasets in Geologic / Structural Mapping at the Northern Negros Geothermal Project (NNGP), Negros Occidental, Philippines
R. A. Camit, L.F. Bayrante, C.C. Panem, O.C. Bien, J.A. Espiridion 488

A Viewer Interface for Interactive Three-Dimensional Geo-Data Visualisation
Admire Kandawasvika, Wolfgang Reinhart, Michael Hahn 495

Web-Based Multimedia GIS for the Analysis and Visualization of Spatial Environmental Database
Shunfu Hu 496

Augmented Maps
Joachim Bobrich, Steffen Otto 502

Modelling of Visualization
L. Prisniakova, V. Prisniakov 506

Application of Differential Splines for Relief Simulation
K. Burshtynska, A. Zayac 509

A Dynamic Mutli-Resolution Model and its Application to 3D Terrain Rendering
Xu Qing, Zhang Boaming, Tan Bing, Ma Dongyang 515

Building Modeling and Visualization For Urban Environment
Efstratios Stamboulogiou, Jie Shan 521

Combination of LIDAR, Digital Photogrammetry and Terrestrial Survey to Generate High-Quality DEMs
Peter Reiss 526

Modeling Multi-Scale Landscape Structure within a Hierarchical Scale-Space Framework
Geoffrey Hay, Danielle. J. Marceau, André Bouchard 532

Aspects of Digital Elevation Model Determination
John C. Trinder, Ping Wang, Yi Lu, Zhijun Wang, Eric D. Cheng 536

WG IV/7: DATA INTEGRATION AND DIGITAL MAPPING

3D Models for High Resolution Images: Examples with Quickbird, IKONOS and EROS
Th. Toutin, R. Chénier, Y. Carbonneau 547

Image Fusion using Wavelet Transform
Zhu Shu-long 552

Linear Spectral Mixture Analysis of TM Data For Land-Use And Land-Cover Classification in Rondônia, Brazilian Amazon
Dengsheng Lu, Mateus Batistella, Emilio Moran 557

The Use of Multi-Beam RADARSAT-1 for Terrain Mapping
Eric Grunsky 563

Automatic Extraction of Buildings from LIDAR Data and Aerial Images
Franz Rottensteiner, Josef Jansa 569

Assigning the 3rd Dimension to Roads by use of LASER Scanning Data
Carsten Hatger 575

Synergy of GPS, Photogrammetry and InSAR for Coastal Zone Monitoring
S. J. Buckley, J. P. Mills, P. J. Clarke, S. J. Edwards, J. Pethick, H. L. Mitchell 581

Problems in the Fusion of Commercial High-Resolution Satelitte as well as Landsat 7 Images and Initial Solutions
Yun Zhang 587

Ubiquitous GIS, Data Acquisition and Speech Recognition
Andrew Hunter, C. Vincent Tao 593

The Design and Implementation of a Novell Object Recognition Method based on Information Fusion and Artificial Intelligent Concept
Farhad Samadzadegan, Ali Azizi, Michael Hahn 599

Evaluation of the Potential of the High Resolution Satellite Images (IKONOS) for Large Scale Map Revision
Farhad Samadzadegan, Ali Azizi, Mohammad Sarpoulaki, Ahmad Talebzadeh 601

Application of High-Resolution Satellite Images in Forestry and Habitat Mapping-Evaluation of IKONOS Images through a Hungarian Case Study
D. Kristóf, E. Csato, D. Ritter 602

Using Landsat 7 Images to Update Canadian National Topographic Data
Marie-Eve Martin, Eric Loubier 608

A High Quality Digital Orthophoto of the Mt Everest and its Surroundings
Toya Nath Baral, Michael Hahn, Frank Neidhart 613

High Accuracy Handheld Mapping System for Fast Helicopter Deployment
Jan Skaloud, Julien Vallet 614

Rigorous Generation of Digital Orthophotos from EROS-A High Resolution Satellite Images
Liang-Chien Chen, Tee-Ann Teo 620

An Accuracy Adjustment by Fusion Method with GIS Data and Remote Sensing Data
Jong Hyeok Jeong, Takagi Masataka 626

A New Epipolarity Model Based on the Simplified Push-Broom Sensor Model
Hae-Yeoun Lee, Wonkyu Park 631

3D Reconstruction From Very High Resolution Satellite Stereo and its Application to Object Identification
Jung Rack Kim, Jan-Peter Muller 637

Integration and Filtering of 3D Spatial Data using a Surface Comparison Approach
Harvey Mitchell, John Fryer, Robert Pâquet 644

Auto-Extraction of Urban Features from Vehicle-Borne LASER Data
Dinesh Manandhar, Ryosuke Shibasaki 650

A Different Approach to the Spatial Data Integration
A. Ulubay, M.O. Altan 656

Fusing Stereo Linear CCD Image and LASER Range Data for Building 3D Urban Model
Masafumi Nakagawa, Ryosuke Shibasaki, Yoshiaki Kagawa 662

Merging of Heterogeneous Data for Emergency Mapping: Data Integration or Data Fusion?
Florin Savopol, Costas Armenakis 668

WG IV/8: GLOBAL ENVIRONMENTAL DATABASES

Towards a Global Elevation Product: Combination of Multi-Source Digital Elevation Models
A. Roth, W. Knöpfle, G. Strunz, M. Lehner, P. Reinartz 675

Special-Purpose Mapping of Buried Structural Unconformities in a Basement and Sedimentary Cover
Valentina B. Sokolova 680

Elaboration of Products Derived from the EO Data And GIS Info-Layers For Flooding Risk Analysis
Gheorghe Stancalie, Corina Alecu, Elena Savin 684

Integration of Remote Sensing Data and Field Models of In-Situ Data Capture in GIS for Environmental Sensitivity Index Mapping: A Nigerian Example
Oluseyi O. Fabiyi 690

Remote Sensing and GIS: An Aid to Selecting Habitable Site in Central Himalayas, India
S. K. Sharma 697

Petroleum and Gas Research by Remote Sensing in South China Sea
Xu Ruisong 699

Projecting Global Raster Databases
E. Lynn Usery, Michael P. Finn, Douglas J. Scheidt 704

WG IV/9: EXTRATERRESTRIAL MAPPING

USGS High Resolution Topo-Mapping of Mars with Mars Orbiter Camera Narrow-Angle Images
R. Kirk, L. Soderblom, E. Howington-Kraus, B. Archinal 713

USGS and DLR Topographic Mapping of Comet Borrelly
E. Howington-Kraus, R. Kirk, L. Soderblom, B. Giese, J. Oberst 723

Photogrammetric Registration of MOC Imagery to MOLA Profile
Jie Shan, Scott Lee, Jong-Suk Yoon 728

High Precision Landing Site Mapping and Rover Localization by Integrated Bundle Adjustment of MPF Surface Images
K. Di, F. Xu, R. Li, L. H. Matthies, C. F. Olson 733

An Ortho-Image Map using Data Obtained from the Mars Orbiter Camera of Mars Global Surveyor
M. Wählisch, G. Niedermaier, S. van Gasselt, F. Scholten, F. Wewel, T. Roatsch, K-D. Matz, M. Hoyer, R. Jaumann 738

Mars Geodesy/Cartography Working Group Recommendations On Mars Cartographic Constants and Coordinate Systems
T. C. Duxbury, R. L. Kirk, B. A. Archinal, G. A. Neumann 743

IC WG II/IV: SYSTEMS FOR AUTOMATED GEO-SPATIAL DATA PRODUCTION AND UPDATING FROM IMAGERY

An Object-Wise Classification Approach
Volker Walter 749

Rice Inventory Data Update Procedure based on Remote Sensed Images: A User Interface Improvement Perspective
Hsiung-Peng Liao, Jung-Hong Hong 755

Quality Control and Updating of Road Data by GIS-Driven Road Extraction From Imagery
Felicitas Willrich 761

Extraction and Update of Street Networks in Urban Areas from High Resolution Satellite Images
Renaud Peteri, Thierry Ranchin 768

Technological Aspects of the 3-Dimensional Photogrammetric Updating of the Israeli National GIS
Yuri Raizman, Ammatzia Peled 774

Photogrammetric Management GIS System
Ammatzia Peled, Basheer Haj-Yehia 775

Updating and Specification Changes - Statistical Estimation of Workload
Dorthe Holme, Peter Højholt 776

Investigation on Land Database Updating based on High Resolution Airborne Images
Jie Yin, Guomin Zhu, Jiguo Wang, Chengming Li 778

Differential Object Extraction Methods for Automated GIS Updates
P. Agouris, A. Stefanidis, S. Gyftakis, G. Mountrakis 781

Evaluation of Automatic Road Extraction Results from SAR Imagery
B. Wessel, C. Wiedemann , O. Hellwich , W.-C. Arndt 786

Change Detection Methods for the Revision of Topographic Databases
Costas Armenakis, Isabelle Cyr, Evangelia Papanikolaou 792

Fast Parallel Image Matching Algorithm on Cluster
Zhenge Qiu, ZengBo Qian, Zhihui Gong, Qing Xu 798

Index of Authors 805

WG IV / 1:

SPATIAL AND TEMPORAL DATA MODELLING AND ANALYSIS

SPATIAL URBAN MODEL FOR ENVIRONMENTAL PLANNING

Hari Shanker Gupta

Department of Geography University College, M.D. University, Rohtak PIN-124001 (Haryana) INDIA
hari_shanker2001@rediffmail.com / haran_shanker@yahoo.co.in

Commission IV, WG IV/1

KEY WORDS: Urban Landuse, GIS Database and Probability Model

ABSTRACT:

Knowledge and creation of GIS DATABASE on area and distribution of Land Uses and its updating based on natural and man made changes plays an important role in the planning process at macro, meso, micro and local levels. An attempt has been made here to analyse the existing Land Use Pattern and Changes over a period of time with the aim of creation of the database for the process of planning at micro level. The specific objectives related to various aspects of Land Utilization are:

1. To prepare Land Use/ Land Cover Maps of the micro regions using Remote Sensing Data.
2. To create GIS database on the areas under various land use classes and change in the land Use pattern.

Bharatpur district of the state of Rajasthan (India) has been selected for the present study. Multi date Remote Sensing data, Survey of India Toposheets, Cadastral Maps and Census of India reports have been used as basic data along with Sampled Ground Truth Check Data. For analysis and creation of data ARC/INFO GIS Package has been employed. GIS package has also been used to relate the Land Use Information to the villages and arrive the tentative comparison of Land Use as reported in Census of India and as obtained from the Remote Sensing Data Analysis. Major findings of the analysis have been discussed in detail with the sole consideration, whether the GIS package employed has the quality of compatibility, reliability, cost effectiveness and time saving mechanism. Further the role of Remote Sensing Data along with the GIS as a tool has also been discussed or the future planning process..

1. INTRODUCTION

Information on the existing land use/ land cover pattern, its spatial distribution and changes in the land use pattern is a pre-requisite for planning, utilisation and formulation of policies and programmes for making any developmental plan. This information not only provides a better understanding of land utilisation aspects but also plays a vital role in developmental planning. Traditionally for regional planning exercises from macro to micro level, statistical data available under nine - fold revenue classification of land use, is used for studying and analysing the land use pattern at regional and intra regional level. Such an analysis sometimes is also supported by the spatial information through departmental maps and conventional sources. With the advent of remote sensing techniques, both aerial and satellite, it has been possible to prepare land use maps at various levels showing categories of land under different uses both in spatial and statistical form. An attempt has been made here to map land use/ land cover information with remote sensing data and to find out the changes in the land use pattern using a GIS package.

2. OBJECTIVES

The prime objective of the study is to organise an information system oriented towards regional planning at district level around a GIS package. The information system will be organised in sectors and will address issues related to all sectors relevant at district level (SAC, 1992). One of the issues is the optimum use of land resources. Towards this, the specific objectives related to various aspects of land utilisation are as follows:

To prepare Land Use/ Land Cover Maps of the micro regions using Remote Sensing Data.

B. To create GIS database on the areas under various land use classes and change in the land Use Pattern.

3. STUDY AREA

Bharatpur district of the state of Rajasthan in India, located between 26 deg. 40 min. to 27 deg. 50 min. North latitudes and 76 deg. 50 min. to 77 deg. 50 min. East longitudes has been selected for the purpose of study. The total area of the district is around 5085 sq. km. The district is divided into eight tehsils (talukas) viz. Kaman, Nagar, Nadbai, Deeg, Weir, Bharatpur, Rupbas and Bayana. Bharatpur town is the district headquarters which is well connected by rail and road network with Jaipur, Mathura and Delhi.

4. DATA USED

Multidate remote sensing data consists of Landsat - TM data (146-41) of October 1986 and IRS LISS - II data (28-48/49, 29-48/49) of March & October 1989 has been used for the study. Survey Of India Toposheets on 1:50,000/1:2,50,000 scale, Census of India reports and Cadastral maps were also used in the study.

5. METHODOLOGY

The overall methodology adopted for the land use mapping of the region is presented in fig. 1. A reconnaissance survey (pre field survey) was conducted by taking traverses in the entire region to prepare an interpretation key. Different tone, textures and shapes of the objects found on the IRS LISS -II data were checked and an interpretation key was made. Using this interpretation key remote sensing data has been interpreted using High Magnification Enlarger (HME) and the information has been transferred to base maps prepared on scale 1:50,000 scale using Survey of India topographical maps. Standard land use and land cover classification has been adopted for mapping land use categories using RS data (Anderson et al., 1992; NRSA, 1989 AND Pathan S. K., 1992). The classification and coding scheme followed for the mapping is given in Table - 1. In all seventeen land use maps were prepared on this scale using multidate RS data viz. Landsat TM data form 1986 and IRS LISS II data for 1989. These maps show the spatial distribution, extent and location of various land uses in the district of Bharatpur. The classification accuracy of these maps has been assessed on the basis of simple random sampling method. The sample size was selected on the basis of the following equation (Fitzpatrick -lins and Chambers, 1977):

$n = (p \times q) / d \times d$

where,

n = Sample Size
p = Desired percent of estimated accuracy
q = Difference between 100 and p in percent
d = standard error in percent

The sample size was calculated based on the following parameters:

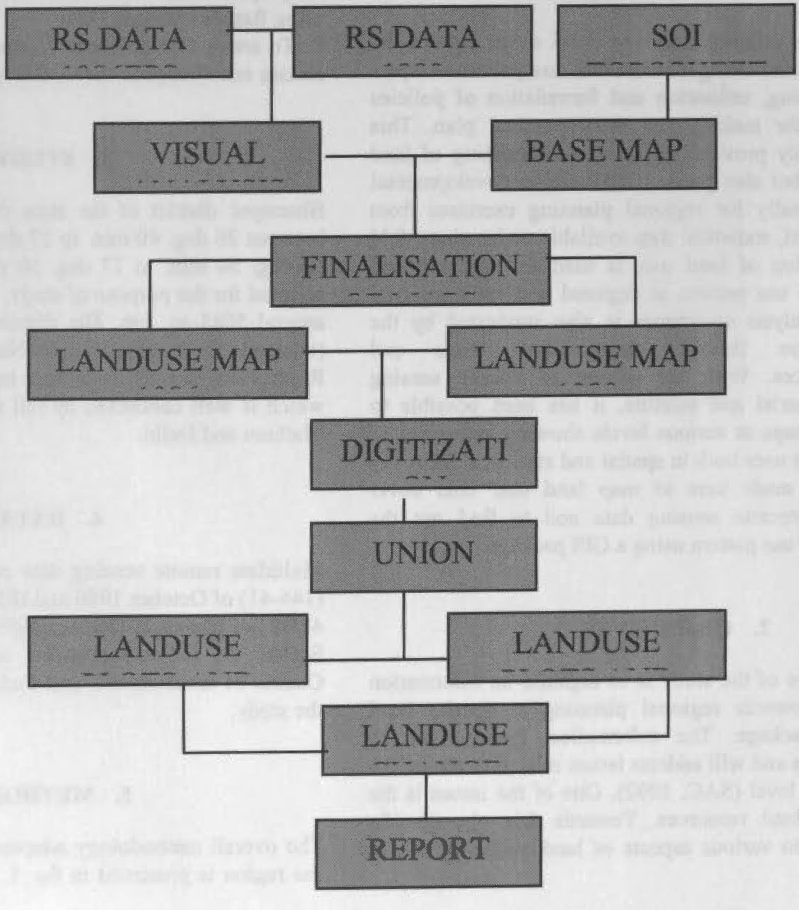
p= 85,
q= 100- 85 =15
d = 5% level of significance at 95% confidence level.

In all 51 points in different categories distributed throughout the district have been checked. A confusion matrix was drawn to assess the overall classification accuracy and mapping accuracy. It was found that the classification accuracy stands at 88% (Table - 2: Classification Accuracy Assessment).

ARC/INFO GIS package attached with VAX 11/750 available at Space Application centers has been employed for the land use analysis. The steps followed for the analyses are:

A. Formation of Titles,
B. Digitization,
C. Editing,
D. Creation of polygon topology, and
E. Union process, A spatial database was created using these steps.

FIGURE-1: FLOW OF STEPS FOR



6. LAND USE ANALYSIS:

As has been discussed in methodology, RS data has been used and land use maps have been repaired on a SOI base at 1:50,000 scale. These have been incorporated into GIS database and statistics obtained on each class on a tehsil-wise basis. Table - 3 and Table - 4 show there land use statistics for the district during 1986 and 1989 respectively. Fig. 2 shows the spatial distribution, location and extent of land uses in Bharatpur district and changes in the land use pattern during 1986 - 1989. Though the GIS data base consists of the level - II details and categories mentioned earlier (Table - 1), this information has been abstracted to level -I for the purpose of obtaining an output commensurate to 1:2,50,000 scale. However, 1:50,000 scale outputs can also be obtained.

To illustrate the concept of the capability of a spatial database, land use categories for one tehsil - Bayana has been extracted & enhanced in scale from the GIS database and is shown in Fig. 3. Based on a systematic GIS database organization, it is possible to obtain different types of outputs for the whole district, for individual tehsils, on a SOI Toposheet basis, for an area of interest defined by a polygon etc.

The changes in the landuse pattern have been determined using the GIS package where the two date land use information have been integrated. The changes in the land use classes and the extent of each land use category is presented in Fig. 2 and Table - 5.

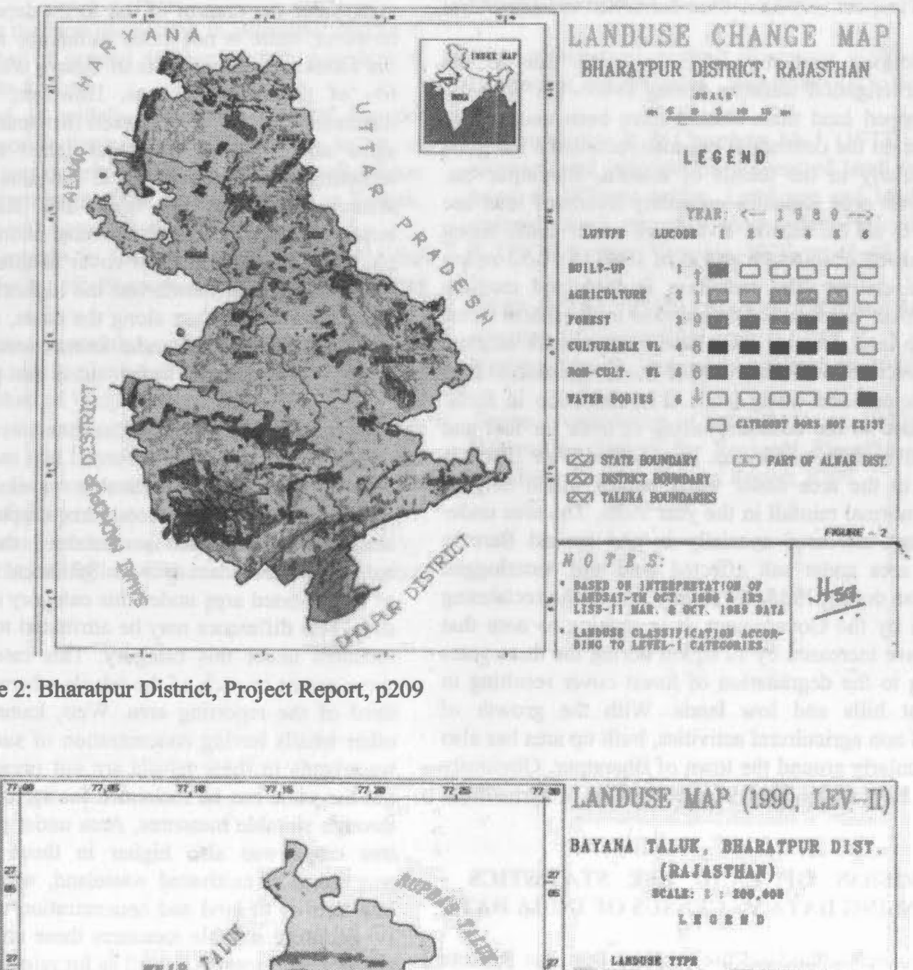

HARYANA
UTTAR PRADESH
PART OF ALWAR DISTRICT
DHOLPUR DISTRICT
INDEX MAP
INDIA
LANDUSE CHANGE MAP
BHARATPUR DISTRICT, RAJASTHAN
SCALE
LEGEND
YEAR: <— 1 9 8 9 —>
LUTYPE
LUCODE
BUILT-UP 1
AGRICULTURE 2
FOREST 3
CULTURABLE WL 4
NON-CULT. WL 5
WATER BODIES 6
CATEGORY DOES NOT EXIST
STATE BOUNDARY
DISTRICT BOUNDARY
TALUKA BOUNDARIES
PART OF ALWAR DIST.
FIGURE - 2
NOTES:
* BASED ON INTERPRETATION OF LANDSAT-TM OCT. 1986 & IRS LISS-II MAR. & OCT. 1989 DATA
* LANDUSE CLASSIFICATION ACCORDING TO LEVEL-I CATEGORIES
Hsa

Figure 2: Bharatpur District, Project Report, p209

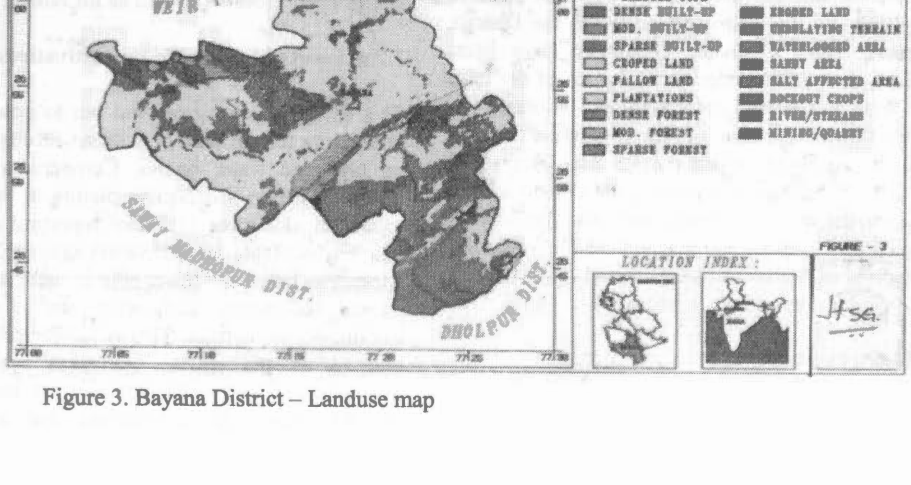

WEIR TALUK
RUPBAS TALUK
SAWAI MADHOPUR DIST.
DHOLPUR DIST.
LANDUSE MAP (1990, LEV-II)
BAYANA TALUK, BHARATPUR DIST.
(RAJASTHAN)
SCALE: 1 : 200,000
LEGEND
LANDUSE TYPE
DENSE BUILT-UP
MOD. BUILT-UP
SPARSE BUILT-UP
CROPED LAND
FALLOW LAND
PLANTATIONS
DENSE FOREST
MOD. FOREST
SPARSE FOREST
ERODED LAND
UNDULATING TERRAIN
WATERLOGGED AREA
SANDY AREA
SALT AFFECTED AREA
ROCKOUT CROPS
RIVER/STREAMS
MINING/QUARRY
FIGURE - 3
LOCATION INDEX :
Hsa

Figure 3. Bayana District – Landuse map

7. DISCUSSION ON MAJOR FINDINGS

7.1 CHANGES IN LAND USE DURING 1986 - 89.

In order to analyse the changes in land use pattern at regional and area level, time series data for a longer period would be desirable. In the absence of such time series data, changes in land use pattern have been studied during 1986 - 89 on the basis of remote sensing data which may be taken as indicatives of the trends in the emerging land use pattern. From Table - 5 it is observed that the district has registered an increase in cropped land from 3393 sq.km to 3615 sq.km during 1986 -89 which might have been facilitated by the reclaiming measures taken by the Government. The sandy area, waterlogged area and salt affected land amounting to about 164.95 sq.km, 37.53 sq.km and 7.06 sq.km respectively have been converted to agricultural use in 1989.

Increase in cropped land may also be partly due to the improvement of irrigation facilities during 1986 - 89. With the increase of cropped land there should have been decrease in fallow lands but on the contrary it has also recorded a marginal increase particularly in the tehsils of Kaman, Bharatpur etc. Decrease in forest area indicates unhealthy trends of land use pattern almost in all the tehsils. Even area under dense forest cover has decreased from 44.98 sq.km in 1986 to 10.52 sq.km in 1989 in the district. The reduction in high and medium density forest cover has resulted an increase in the sparse forest cover and scrub land. On the other hand, reduction of original sparse forest cover 1986-89 has resulted in the increase of area under rock outcrops and waste lands. The reduction in forest cover may be due to the constant felling ot trees for fuel and fodder in the hilly tehsils of Bayana, Rupbas and Weir. There is some increase in the area under water bodies which may be because of the normal rainfall in the year 1986. The area under water bodies has increased specially in and around Baretha reservoir. The area under salt affected land and waterlogged area has decrease during 1986- 89 mainly due to the reclaiming measures taken by the Government. It is striking to note that eroded lands have increased by 12 sq.km during the three years period pointing to the degradation of forest cover resulting in erosion at foot hills and low lands. With the growth of settlements and non agricultural activities, built up area has also increased particularly around the town of Bharatpur. Obviously such growth in built up land has been at the cost of agricultural land.

7.2 COMPARISION OF LAND USE STATISTICS - REMOTE SENSING DATA VS CENSUS OF INDIA DATA

As a separate exercise, the land use mapped from the Remote Sensing Data of 1986 and the land use reported in the District Statistical Abstracts, 1985 - 86 for Bharatpur have been compared. Table - 6 shows the comparative figures of the land use / land cover. This comparison is to relate and illustrate the differences in the figures because of the differences in the classification schemes adopted. While the RS data analysis is based on the land use classification Table - 1, the Census adopts the revenue classification. As a result, the plan generation exercise has to keep this in mind and account for the mismatch in the land use figures of the two sources. Based on the above analysis, the following observations are noteworthy:

7.2.1 Land Under Agriculture:

The cropped land and fallow land together constitute the extent of land under agriculture. As may be seen from the statistical data and land use figures computed from the remote sensing data in Table - 6, proportion and extent of total agricultural land in 1986 was comparable (about 80%) in the district. The distribution of cropped area was highly uneven ranging from 46% in Bayana Tehsil to about 96% in Nadbai Tehsil. In Bharatpur, Deeg, and Nagar Tehsils cropped land was more than 80% and in the remaining three tehsils of Kaman, weir, and Rupbas, it was between 60 and 80 percent.

7.2.2 Land Under Forest:

Forest area interpreted from satellite imagery relates to actual forest cover and is generally less than the forest area given in the Statistical Year Book which, by and large, indicates total area under the control of the forest department. Interestingly, however, there is not much difference in Bharatpur district in the forest area in both sets of figures which works out to about 6% of the reporting area. However, at tehsil level, some differences are noticed between two sources i. e. statistical data show about 8% in Bharatpur tehsil as against 2% of the reporting area in the RS data. Revenue figures do not record area under forest in Nagar and Kaman tehsils whereas interpretation of RS data indicates about 1% and 3.3% of the reporting area under forest cover in these tehsils, respectively, Bayana and Weir tehsils had the highest and ense forest cover spread over in patches along the nalas, roads, railways and on isolated hillocks. Area under forest cover gradually decreases as one moves from south west to north east part of the district.

7.2.3 Other Uncultivated Land:

This category includes cultivable wasteland, permanent pastures and land under miscellaneous tree crops and groves. As for as land under this class is concerned there is vast difference between the two data sources. Statistical figures reported 3.25% of the reported area under this category against 7.1% in the RS data. This difference may be attributed to the categories of land included under this category. This category was reported to some extent in each of the tehsils where Bayana tehsil has one third of the reporting area. Weir, kaman and Deeg were the other tehsils having concentration of such lands. Cultural able wastelands in these tehsils are not presently under cultivation but the same can be reclaimed for agriculture and fodder crops through suitable measures. Area under permanent pastures and tree crops was also higher in those tehsils having higher proportion of cultivated wasteland, which may be ascribed to low fertility of land and concentration of livestock population. By adopting suitable measures these areas may be turned into rich grazing grounds as well as for raising of fodder crops.

7.2.4 Land not available for cultivation:

This category consist of land put to non agricultural use i. e. undulating terrain with or without scrubs, rock out-crops, built up land and water bodies. Comparison of two sets of data sources portrays a deceptive picture in this regard. According to statistical figures 11.5% of reporting area was classified as land not available for cultivation against 7.5% in RS data which appears to be on the lower side in view of the larger number of human settlements and other non - agricultural activities existing in the district. This is further substantiated by the fact that in all the tehsils, statistical figures recorded higher percentage of land under this category in all the tehsils against the RS data. It is observed that a smaller settlements particularly rural ones, could not be interpreted clearly from RS data because of coarser spatial resolution and therefore, area under this category was reported less in RS data. There is also possibility of inter-mixing of sub classes of land use in an inappropriate major class thereby affecting the results. These areas concentrated more than 10% in tehsils of Nagar, Rupbas and Kaman and less than 10% (total geographical area of respective tehsils) in the remaining. As mentioned earlier, barren lands are more in Bayana tehsils accounting for about 15% of the total reporting area. Land under non agricultural uses including built up land, water bodies, transport network etc. was concentrated more in Bharatpur tehsil because of location of big urban settlements and other non - agricultural activities.

8. CONCLUSION

- It is observed that the land use pattern in Bharatpur district is not similar to that of general and use pattern prevalent in Rajasthan State as a whole.
- Agricultural land is widely distributed through out the district. Its concentrations however relatively lower in southern tehsils particularly in Bayana and relatively higher in the north -eastern parts of the district comparison Kaman, Deeg, Bharatpur and Nagar tehsils.
- Forest cover is mainly in the south western parts of the district, in Tehsils of Bharatpur around Ghana Bird Sanctuary Area.
- Area under pastures is mainly confined to Kaman, Bayana and Weir tehsils while plantation and tree crops are more pronounced in tehsil of Weir.
- The culturable wasteland is relatively more concentrated in Bharatpur, Kaman and Bayana Tehsils and interestingly the barren lands are also comparatively more in the latter two tehsils because of rocky terrain and poor soil conditions.
- Category wise picture of the major land uses as revealed by RS data and statistical figures (Census of India), the extent of the total agricultural land and the forest cover in 1986 comparable. However, there are differences in the areas of cultural able and non- cultural able wasteland. This may be due to different definitions adopted in classifying land uses from both sources and also due to the presence of number of smaller settlements, which would not be interpreted clearly on RS data due to its coarser spatial resolution. Because of the adoption of the GIS database, the change analysis and administrative unit wise land use and change categorization were possible in an easy manner.

9. REFERENCES:

Anderson J.R., Hardy E.E., Roach J.T. & Witmer R. E. (1976) : A landuse and landcover classification system for use with Remote sensor data, A USGS print of paper, 964, 28p.

District Statistical Abstract Bharatpur, (1986-87) : A Bureau of Economics and Statistics, Rajasthan publication, 150p.

Fitzpatrick-lins K & Chambers M. J. (1977) : Determination of accuracy and information content of land use and land cover maps at different scales. Land use and land cover maps and statistics from remotely sensed data, Edited by Anderson J. R.,USGS, Reston, Verginia, 1977, pp.41-48.

NRSA (NRSA - 1989) : Manual of National landuse / landcover mapping using satellite imagery, Part -I, A National Remote Sensing Agency, Hyderabad publication, 1989.

Space Application Centre(SAC - 1992): A GIS base Information System for Regional Planning - A case study of Bharatpur District, Project Report, p209.

TABLE 1. CLASSIFICATION SCHEME FOR LANDUSE ANALYSIS.

No.	Level I	Level II
1.	Urban or built up land	1.1 Dense Urban
		1.2 Moderate Urban
		1.3 Sparse Urban
2.	Agricultural land	2.1 Crop land
		2.2 Fallow land-
3.	Forest	3.1 Dense
		3.2 Medium
		3.3 Sparse
		3.4 Scrub/Degraded
4.	Waste lands	4.1 Salt affected land
		4.2 Gullied/ Eroded land
		4.3 Waterlogged Areas
		4.4 Undulating uplands with or without Scrubs
		4.5 Sandy Area
		4.6 Rock outcrops
5.	Water bodies	5.1 Rivers/Streams/Creeks
		5.2 Reservoirs/Tanks/Lakes
6.	Others	6.1 Habitation with plantations

TABLE - 2: CLASSIFICATION ACCURACY ASSESSMENT THROUGH COFUSION MATRIX

Catgory On ground		Category interprated from IRS LISS - II DATA										Total	Ommis-sions	Commis-sions
		1	2	3	4	5	6	7	8	9	10			
Built up	1	3	-	-	-	-	-	-	-	-	-	3	0	0
Crop	2	-	9	-	-	-	-	-	-	-	-	9	0	1
Fallow	3	-	-	7	-	-	-	-	-	-	-	7	0	1
Plantations	4	-	1	-	3	-	-	-	-	-	-	4	1	1
Forest	5	-	-	-	1	5	-	-	-	-	-	6	1	0
Undulating Terrain	6	-	-	-	-	-	7	-	-	-	-	7	0	1
Wst.,Scurb& Er.ld	6	-	-	-	-	-	1	3	-	1	-	5	2	0
Eroded land	7	-	-	-	-	-	-	-	-	-	-	-	0	0
Salt affected	8	-	-	-	-	-	-	-	3	1	-	4	1	0
Sandy	9	-	-	1	-	-	-	-	-	3	-	4	1	2
Water	10	-	-	-	-	-	-	-	-	-	2	2	0	0
Total		3	10	8	4	5	8	3	3	5	2	51	6	6

OVERALL CLASSIFICATION ACCURACY = POINTS CORRECTLY CLASSIFIED/TOTAL NO POINTS
= (45 /51) X 100 = 88%

TABLE - 3: LANDUSE STATISTICS OF BHARATPUR DISTRICT OF 1986

(Based on Landsat TM data, All figures are in sq.km)

SN.	LAND USE	KAMAN	NAGAR	DEEG	NADBAI	BH'PUR	WEIR	BAYANA	RUPBAS	TOTAL
1.	DENSE URBAN	0.98	0.47	2.23	0.94	5.42	0.58	0.45	2.47	13.54
2.	MODERATE URBAN					2.36				2.36
3.	SPARSE URBAN	0.82	1.04	2.11	1.62	9.47	1.32	0.88	----	17.26
4.	CROP LAND	497.14	397.56	358.55	333.54	706.46	356.95	275.59	464.45	3393.26
5.	FALLOW LAND	118.09	38.24	54.68	95.33	111.47	108.67	107.43	33.61	667.52
6.	DENSE FOREST	----	----	----	----	6.26	10.89	27.83	----	44.98
7.	MEDIUM FOREST	3.38	1.76	7.82	----	10.63	14.16	34.92	----	72.67
8.	SPARSE FOREST	21.38	2.84	20.61	----	3.67	36.40	70.82	12.12	167.84
9.	ERDOD LAND	2.20	----	----	0.90	2.61	1.25	21.02	0.48	28.46
10.	UNDULATING TERRAIN	14.44	0.97	4.91	----	14.65	16.78	154.24	15.93	221.92
11.	WATER LOGGED	2.92	9.30	10.90	0.09	6.89	7.30	0.39	13.13	50.92
12.	SANDY AREAS	10.71	2.10	28.60	8.11	45.34	33.14	92.16	9.62	229.78
13.	SALT AFFECTED	32.69	5.56	8.93	2.85	1.62	----	----	1.15	52.80
14.	ROCK OUT CROP	36.88	11.88	7.63	----	----	7.68	7.34	0.52	71.93
15.	WATER BODIES	0.63	0.56	0.15	----	1.68	7.91	15.34	5.50	31.77
16.	HABITATION & PLAN.	----	----	----	----	4.60	----	----	----	4.60
17.	MARSHY LAND	----	----	----	----	13.52	----	----	----	13.52
18.	MINING AREA	----	----	----	----	----	----	0.75	----	0.75
TOTAL AREA		742.26	472.30	507.12	443.38	949.65	603.03	808.40	558.94	5085.13

TABLE 4: LANDUSE STATISTICS FOR BHARATPUR DISTRICT OF 1989

(based on IRS LISS – II data,all figures are in sq. km)

SN.	LAND USE	KAMAN	NAGAR	DEEG	NADBAI	BH'PUR	WEIR	BAYANA	RUPBAS	TOTAL
1.	DENSE URBAN	0001.34	0000.46	0003.15	0000.74	0004.31	0000.57	0000.63	0002.44	0013.64
2.	MODERATE URBAN	----	----	----	----	2.46	----	----	----	2.46
3.	SPARSE URBAN	1.02	1.54	2.31	1.92	13.06	1.52	1.06	----	22.43
4.	CROP LAND	501.51	402.30	387.41	342.80	731.12	393.90	380.14	476.56	3615.74
5.	FALLOW LAND	127.50	39.87	65.19	97.08	135.56	113.66	106.64	33.88	719.38
6.	DENSE FOREST	----	----	----	----	9.08	----	1.44	----	10.52
7.	MEDIUM FOREST	2.35	----	9.10	----	8.68	4.39	19.48	----	44.00
8.	SPARSE FOREST	----	----	3.65	----	2.08	5.88	74.71	15.47	101.79
9.	ERODED LAND	3.54	----	1.59	0.91	5.42	2.99	23.59	2.72	40.58
10.	UNDULATING TERRAIN	16.97	3.19	5.82	----	11.40	32.26	152.65	14.94	237.23
11.	WATER LOGGED	2.03	4.44	5.79	----	5.36	----	----	0.70	18.32
12.	SANDY AREAS	8.54	0.95	2.49	----	2.06	23.40	11.08	4.05	52.57
13.	SALT AFFECTED	24.56	----	0.16	----	----	----	----	2.17	26.89
14.	ROCK OUTCROP	52.87	15.32	20.54	----	----	18.81	14.46	0.54	122.54
15.	WATER BODIES	----	4.16	0.13	----	5.89	5.69	21.77	5.39	43.03
16.	HABITATION & PLAN.	----	----	----	----	4.63	----	----	0.80	4.71
17.	MARSHY LAND	----	----	----	----	8.59	----	----	----	8.59
18.	MINING AREA	----	----	----	----	----	----	0.75	----	0.75
	TOTAL AREA	742.23	472.23	507.33	443.45	949.52	603.07	808.40	558.94	5085.17

TABLE - 5: CHANGE MATRIX OF LANDUSE CLASSES BETWEEN 1986 –1989

(Based on RS data, All figures are in sq.km)

SN.	1986/1989		B	D	F	DF	MF	SF	ER	UN	WL	SA	SLT	ROC	WB	TOTAL
1.	Built up land	B	37.76	--	--	--	--	--	--	--	--	--	--	--	--	37.76
2.	Crop land	C	2.25	3313.40	72.73	--	--	--	--	--	--	--	--	--	4.87	3393.30
3.	Fallow land	F	3.23	92.80	568.06	--	--	--	--	0.37	--	--	--	--	3.07	667.52
4.	Dense forest	DF	--	--	--	10.52	23.16	11.30	--	--	--	--	--	--	--	44.98
5.	Medium forest	MF	--	--	--	--	20.84	39.58	--	12.25	--	--	--	--	--	72.67
6.	Sparse forest	SF	--	--	--	--	--	50.91	--	102.67	--	--	--	14.26	--	167.84
7.	Eroded land	ER	--	--	--	--	--	--	25.14	--	--	--	--	--	3.32	28.46
8.	Undulating land	UN	--	--	62.88	--	--	--	--	63.32	--	--	--	95.72	--	221.92
9.	Water-logged	WL	--	37.53	--	--	--	--	--	--	26.91	--	--	--	--	64.44
10.	Sandy Area	SA	--	164.95	--	--	--	--	--	15.44	--	46.39	--	--	--	229.78
11.	Salt Affected	SLT	--	7.06	15.71	--	--	--	--	--	--	3.18	26.89	--	--	52.80
12.	Rock out-crop	ROC	--	--	--	--	--	--	--	58.62	--	--	--	13.31	--	71.93
13.	Water bodies	WB	--	--	--	--	--	--	--	--	--	--	--	--	31.77	31.77
Total			0043.24	3615.70	719.38	10.52	44.00	101.80	40.58	237.23	26.91	52.57	26.89	123.29	43.03	5085.17

TABLE - 6: COMPARISION OF LAND USE CATEGORIES FROM STATISTICAL AND REMOTE SENSING DATA

(All figures are in Sq.km, Figures in brackets indicate percentage)

SN	TEHSIL	TOTAL	AREA	AGRI. LAND		FOREST	LAND	OCT.	LAND	NOCT	LAND
		CENSUS	RS	CENSUS	RS	CENSUS	RS	CENSUS	RS	CENSUS	RS
1.	KAMAN	734.1	742.3	597.0	615.2	---	24.8	28.5	48.5	108.5	53.7
2.	NAGAR	471.0	472.3	416.9	435.8	---	4.6	6.3	17.0	47.8	15.0
3.	DEEG	500.9	507.1	432.8	413.3	9.5	28.4	9.3	48.4	49.3	17.0
4.	NADBAI	446.7	443.4	419.8	428.8	---	---	3.0	10.9	23.9	2.6
5.	BHARATPUR	954.8	949.7	826.8	821.0	75.8	20.7	16.8	56.8	82.9	57.1
6.	WEIR	614.0	603.0	460.0	465.6	64.8	61.5	34.9	41.7	54.1	34.3
7.	BAYANA	803.9	808.4	442.3	383.0	151.4	133.6	51.3	113.6	158.9	178.8
8.	RUPBAS	549.1	559.0	469.9	498.1	6.5	12.1	14.7	24.4	58.1	24.4
	TOTAL	5074.5 ---	5085.1 ---	4065.7 (80.1)	4060.8 (79.9)	308.0 (6.1)	285.5 (5.6)	164.8 (3.3)	361.0 (7.1)	583.5 (11.5)	382.9 (7.5)

K-ORDER SPATIAL NEIGHBOURS BASED ON VORONOI DIAGRAM: DESCRIPTION, COMPUTATION AND APPLICATIONS

Renliang Zhao[1,2], Jun Chen[3] and Zhi-lin Li[1]

[1] Department of Land Surveying and Geo-informatics , The Hong Kong Polytechnic University, Hong Kong

[2] Department of Land Surveying and Geo-informatics, The Central South University, Shangcha, China, 410083

[3] 1 National Geometrics Center of China, 1 Baishengcun, Zizhuyuan, Beijing, China, 100044
{zhaorl@sina.com, chenjun@nsdi.gov.cn, lsllz@polyu.edu.hk}

Commission IV, WG IV/1

KEY WORDS: Spatial neighbours, k-order spatial neighbours, Voronoi diagram

ABSTRACT

Spatial neighbouring relation is a very important concept in spatial information science. It has even been regarded as the base of spatial information system. Neighouring information is indispensable for many operations in spatial information system. Therefore, it is significant to describe and acquire such in an efficient way. Most of current concepts of neighbours only involve objects sharing common boundaries. However, spatial neighbours are not restricted to only this type in practical applications. There is still a lack of systematical discussion on the definition and computation of such spatial neighbours.

In this paper, a more generalized concept of neighbours and some applications in GIS fields are systematically discussed. A model is developed for a particular type of neighbours, called k-order neighbours using Voronoi diagrams. Several efficient methods for computation of k-order neighbours are developed for different cases. Then some possible applications are discussed including spatial query and neighborhood analysis.

1 INTRODUCTION

Spatial neighbour is one of very important concept in spatial information science, and is even regarded as indispensable for spatial information system (Gold, 1989). Neighbour information needs to be maintained during the whole modelling and processing procedure of spatial data. Also, many operations for spatial query and analysis in GIS are based on this concept.

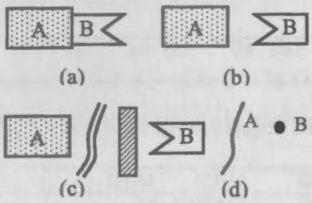


Figure 1. Various kinds of spatial neighbours

Spatial neighbours can be usually defined as objects having common boundaries. However, neighbouring relations are not restricted to such a type in real geographical space. In Figure 1, spatial objects A and B in (a) are neighbours, but in (b) and (c), they are not be neighbouring because their boundaries don't touch each other. But spatial interaction still remains between these two objects and the relational information in the cases (b) and (c) is very useful in some applications of spatial query and analysis. The reality is that due to the fact that they have no common boundaries, it is difficult to define and compute the relational information.

For such spatial information, several authors have also mentioned such concept as neighbours among points using the dual graph of Voronoi diagrams, i.e., Delaunay triangular nets, in some practical applications such as generalization and spatial analysis (Okabe 1992, Broke 1996, Zhang and Murayama 2000). But such a method can be applied to only point objects. An alternative method is presented by Chen and Zhao (Chen et al., 2000, Zhao et al., 1999), using Voronoi distance based on Voronoi diagrams. In this method, two objects are k-order neighbors if there are at least k Voronoi regions in between. But there is still a lack of deep and systematical analysis of the concept and search methods of spatial neighbours. In this paper, spatial neighbours, especially k-order spatial neighbours are systematically discussed, i.e. on its definition, characteristics, computation and applications.

The remainder of this paper is organized as follows: Section 2 analyzes the current methods for defining spatial neighbours. Section 3 gives definitions and Section 4 describes the algorithms for computing k-order spatial neighours and then exploits its characteristics. Section 5 discusses some potential new applications of k order neighbours to GIS, concentrating on spatial query, analysis, and constructing neighbourhood on the Voronoi distance. Section 6 gives some conclusions.

2 CURRENT METHODS FOR DEFINTING SPATIAL NEIGHBOURS

Generally, spatial neighbouring relations can be classified into two types: one type is referred as the relation between two objects touching each other; the other is as the relation between disjoint objects.

The former can be easily and clearly defined by common parts of neighbouring objects, and can be obtained directly since they are often explicitly stored in most GIS databases with the support of topology. This type of neighbouring relations can be called topological adjacency.

The other type of neighbouring relations exists among disjoint objects. Statistically, the distribution of disjoint relation prevails much over that of other relations in geographical space (Florence, 1996). So, it is very significant to describe this type of neighbours among discrete object. But this type of neighbors are often fuzzy and complex, and related to a certain definition of distance, are also dependent on the context within which they are applied (Gahen, 1995).

Currently, there are two main methods, i.e., metrical method and topological method, to define and compute the two types of spatial neighbours including this type of neighbours among disjoint objects.

The metrical method is widely used in GIS. In this method, neighbouring relations are usually defined by distance and neighbours are defined as objects within a specified distance (Hernandez et.al. 1995; Huang and Sevonson 1993, Hong, 1995). Distance relations are quantitative and often easily got. But neighbouring relations are qualitative, so the conversion between quality and quantity has to be done in the method. In some cases, the computation of distances is complex and difficult, so the cost becomes high.

The topological method mainly refers to the fact that spatial neighbours are defined by topological concepts. In this method, spatial neighbours will have common boundaries but no interiors. The advantage of this method lies in: the high distance computation can be avoided and related operations will be simple and clear. So it is a qualitative method different from the metric method.

If topological method is extended to Voronoi diagrams, it forms another useful method, Voronoi method. Gold (1992) gives the definition of neighbors based on Voronoi diagram, i.e., two objects are neighbouring if they share common Voronoi boundaries. Chen et al. (1991) also describe this type of neighboring relations by using Voronoi 9-Intersection model for spatial relations. Furthermore, it can be distinguished into more detailed neighboring relations such as lateral neighbor, nearest neighbor and second-order neighbor. Voronoi method actually transforms the neighbouring problems of spatial objects into those of their Voronoi regions. In this way, use of the various advantages of Voronoi diagram [Aurenhammer, 1991] can be made, particularly semi-quantity nature [Edward, 1993; Gold, 1989].

3 DESCRIPTION OF K-ORDER SPATIAL NEIGHBOURS USING VORONOI DIAGRAM

3.1 Description of K-order neighbour based on Voronoi diagram

From the viewpoint of computational geometry, Voronoi diagram is essentially 'a partition of the plane into N polygonal regions, each of which is associated with a given point. The region associated with a point is the locus of points closer to that point than to any other given point (Lee and Drysdale 1981). A point associated with Voronoi regions is called generation point. Of course, these generators can be also line and area objects (Okabe, 1992).

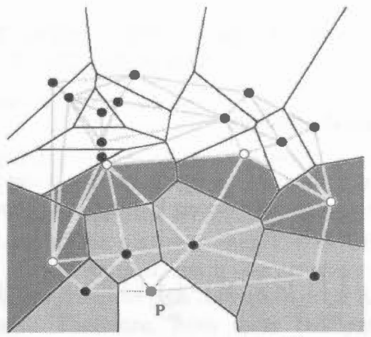


Figure 2 k-order neighbors to object P

On Voronoi diagrams, k-order neighbour can be defined by using Voronoi regions as the following:

(1) Two objects are 1-order neighbours, if two objects share boundaries of Voronoi region;

(2) An object is a k-order neighbour of a given object, if the object is an immediate neighbour of one k-1 order neighbour of the given object, only if it is not a k-2 order neighbour.

In Figure 2, the bold points are *1*-order neighbours to generator *P*, the hole points are 2-order neighbours to generator *P*.

3.2 Mathematical definitions of K-order neighbours

Mathematically, k-order neighbour can be described formally in terms of point set theory as follows.

K-order neighbour: Let P is a set of objects P_1, P_2, ……, Pn in a finite convex in IR^2, P_i, $P_j \in P$ ($i \neq j$, i, j=1,……n). K-order neighbour of Pi can be described:

(1) *1*-order neighbour:

$$N_1(P_i) = \{P_j \mid \exists x, d(x,P_j) = d(x,P_i)\} \quad \text{………. ..} \quad (1)$$

(2) k-order neighbour:

$$N_k(P_i) = \{P_j \mid \exists x, d(x,P_j) = d(x, N_{k-1}(P_i), k > 1, if\, k > 2, P_i \notin N_{k-1}(P_i)\}$$
$$. \quad \text{…………………} \quad (2)$$

where, *d* denotes a distance function.

If the order is described as Voronoi distance (denoted by *vd)*, k-order spatial neighbour of object *Pi* can be defined as the following equation.

$$N_k(P_i) = \{P_j \mid vd(P_i,P_j) = k, k > 0\} \quad \text{…………………..} \quad (3)$$

Using the concept of point set topology, k-order spatial neighbour can be classified as:

(1) Topological neighbour (*TN*): two objects have common boundaries

$$TN(P_i) = \{P_j \mid \partial P_j \cap \partial P_i \neq \varnothing \; and \; P_i^0 \cap P_j^0 = \varnothing\} \quad \text{………} \quad (4)$$

(2) Geometrical neighbour (*GN*):

$$GN_k(P_i) = \{P_j \mid P_i \cap P_j = \varnothing \; and \, vd(P_i,P_j) = k, k > 0\} \quad \text{……} \quad (5)$$

4 COMPUTATION OF K-ORDER SPATIAL NEIGHBOURS

4.1 General principle

From the definition of k order neighbors, it can be derived that a certain relations hold among k order, k-1 order and k-2 order neighbors, as follows:

$$\bigcup_{i\geq 0} S_i^{n+1} = \bigcup_{i\geq 0}\bigcup_{j\geq 0} (S_i^n)_j^1 - \bigcup_{i\geq 0} S_i^{n-1} \qquad \text{............ (6)}$$

$\sum S_i^{n+1}$ -----object S's all n+1 order neighbors;
$\sum\sum (S_i^n)_j^1$ ---all immediate neighbors to object S's all n order neighbors;
$\sum S_i^{n-1}$ ----object S's all n-1 order neighbors;
$\sum S_i^n$ ----object S's all n order neighbors;
S_i^n ---a n order neighbor of object S.

This forms the basis of the computation of k-order neighbors. According to this relation, the computation of k-order is composed of two parts: (1) compute and representing 1-order neighbour; (2) derive k-order neighbour.

4.2 Computing and Representing 1-order Neighbour

According to the definition, two objects will be 1-order neighbours if their Voronoi regions share common parts. Therefore, if all edges of a Voronoi region are searched once, all 1-order neighbours to the object (to which the Voronoi region belongs) will immediately be found.

The procedure for the computation can be described as the following,

Let E_V be the set of the edges of Voronoi boundaries; L_O the list of objects as generators; L_{N1} the list of 1-order neighbours; O_T the set of objects related to the current edge; and P_{N1} the pointer of each object to its 1-order neighbours in the list of L_{N1}.

Step 1: O_T = empty, L_{N1} = empty, for each object P_{N1} = *NULL*;

Step 2: Take an edge from E_V, find all objects which take the edge as Voronoi boundaries, put them into O_T;

Step 3: For each object *T* in O_T, record 1-order neighbours. If P_{N1} is *NULL*, allocate relative position to P_{N1} in L_{N1}; Record the others of O_T as its 1-order neighbours, insert them into L_{N1} at the position pointed by P_{N1}. If P_{N1} is not NULL, check if each object in the others of O_T recorded as 1-order neighbour to *T*, insert the object into the L_{N1} at the position pointed by P_{N1}; O_T = *NULL*;

Step 4: if all edges in E_V are searched, then stop; or go to step 2

4.3 Computing K-order Neighbours

Several methods for searching k-order neighbours have been designed for different cases.

4.3.1 Waved front Method

If one wants to know all k-order neighbours to a given object, the search can start with this object in a waved way. The procedure for the computation is the following.

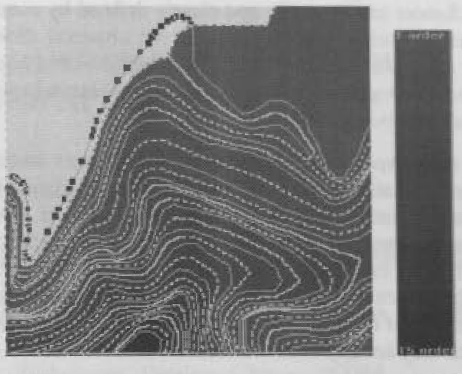


Figure 3. All k order neighbors to a given object obtained with wave front method (up to 15-order).

Let L_N be the list of found k-order neighbours to a given object T_0; *T* the current object; L_{N1} the list of 1-order neighbours to *T*; Q_N the queue of neighbours to be used for searching higher order neighbours; *P* the beginning position of I order objects and *K* a given order value of neighbours, then

Step 1: $L_N \leftarrow T_0$, Q_N = empty, I = 0, T=T_0, P= T_0;

Step 2: IF L_{N1} is not *NULL*, I < *K*, THEN Check all objects in L_{N1}, and put those objects not in L_N into L_N ELSE stop.

Step 3: IF I < *K* and *T* = *P*, THEN I=I+1, move the pointer *P* to the end of L_N, ELSE Stop and output L_N,;

Step 4: Move the pointer *T* to the next object in L_N, goto step 2.

Figure 3 is an illustration of a 15-order neighbours of a contour line, computed by this method.

4.3.2 Meeting Method

If two objects are given, the computation can implemented with a meeting method in which searching starts with the two objects at the same time (Figure 4).

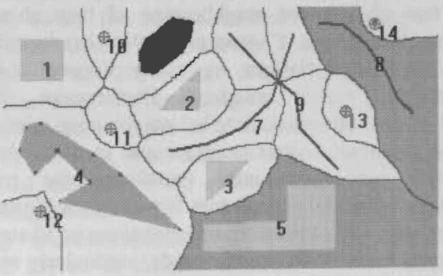


Figure 4. Computing the order of two objects using meeting method: object 4 and 14 are 3 order neighbors

Let T_0, and T_1 be the two specified objects; L_0 k-order neighbours to T_0; L_1: k-order neighbours to T_1, then

Step 1: L_0 and L_1 are set empty, i = 0;

Step 2: IF objects pointed to by P_{N1} of T_0 include *T1*, THEN *K*=1, stop; ELSE i =1;

Step 3: Perform the waved front method respectively for T_0 and T_1, search i-order neighbours: Search for i-order neighbours to T_0, record them into L_0, and search for i-order neighbours to T_1, record them into L_1.

Step 4: Check L_0 and L_1. IF L_0 and L_1share at least one object, then *K* = 2*i stop. IF there is an object in L_0 which is 1-order neighbour to an object in L_1, THEN $K = 2*i + 1$, stop; ELSE L_0 and L_1 are set empty, i = i+1, go to Step 3.

4.3.3 Directional Method

This method can be used for searching k-order neighbours to given two objects along a specified routine.

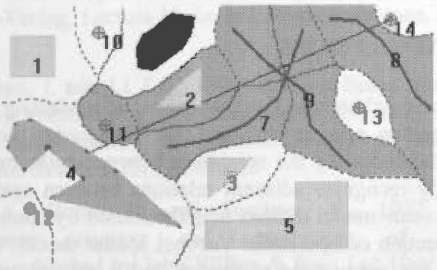


Figure 5. Computing k-order neighbors along a specified line: object 4 reaches 14 through 11,2,7,9,8.

There exist only two cases between two neighboring objects of an specific object: (a) two objects are not immediate neighbors but separated by the Voronoi region of the specific object, so the two objects are 2-order neighbors; and (b) two objects are immediate neighbors. Therefore, the searching procedure must be traced back to check if each new object in current step is one immediate neighbour of the got object. If they are not immediate neighbours, the order *K* should be increased by 1, else *K* value retains as the old order.

The procedure is implemented as follows: (a) Suppose an auxiliary straight line starting from an object to another one as the desired routine; and (b) Trace k-order neighbours along the line at the beginning of an end of the line, stop till finding another object (Figure 5).

The algorithm can be implemented in terms of the following steps.

Let K be the order to be determined between two specified objects; *T* the current object whose Voronoi boundary intersects the auxiliary line; *N* the found neighbour; *A* the source object, *B*: the destination object; *L* the straight line between *A* and *B; then*

Step 1: Select one of them as the source object *A*, then the other is the destination object *B*, construct the straight line *L* between *A* and *B*, *T* = *A*, *K*=0;

Step 2: Compute the intersection point between line *L* and *T* 's Voronoi boundary, get the edge of the Voronoi boundary intersecting the auxiliary line, then the neighbours sharing this edge with *T* can be found immediately. The found neighbours are denoted by *N*;

Step 3: IF $K > 1$, check whether N are the (k-1)-order neighbours to *T*. ELSE, THEN *K* = *K*+1;

Step 4: *T* = *N*;

Step 5: IF *N* is the destination object *B*, THEN stop and output the value of *K*, ELSE, go to step 2.

5 POTENTIAL APPLICATIONS

K-order neighbours can be applied into spatial query and analysis in GIS. Several authors have applied spatial neighbours into spatial analysis and map generalization based on Delaunay nets (Broke 1996, Zhang and Murayama 2000). Here it is attempted to discuss some potential other applications in GIS.

5.1 Nearest neighbour query

Since (k+1)-order neighbours are 'farther' to the generator than k-order neighbours and located behind k-order neighbours, there must exist an object of k-order neighbours, which is nearer to the generator than anyone of (k+1)-order neighbours. If one wants to reach one generator from (k+1)-order neighbours, he must pass through k-order neighbours to the generator. Therefore, from these two properties and Euclidean distance, one very important fact can be inferred: the *kth* nearest object to a given object must be within from 1-order neighbours to k-order neighbours to the object. For example, the nearest object to a given one must be within 1-order neighbours, it impossibly belongs to the 2-order neighbours to the given object. Or, it will violate the above two properties. So the query for kth nearest object can be implemented with the combination of k-order neighbours and Euclidean distance. It reduces the searching range to a local range, but at the same it can guarantee the completeness and accuracy of the result.

5.2 Buffering Query

Buffer zone is a kind of influence or service range of a spatial object or engineering planning item. It is one of important functions for searching for information in generalization of map and spatial analysis of GIS. Many algorithms for buffer zone construction have been presented in order to improve the efficiency of buffer analysis in GISs. Currently, the general method for searching based on buffer zone based on traditional space model is implemented in the whole geographical space so as to guarantee the completeness of searching result. The cost of time is very large. From the definition, k-order neighbours and their Voronoi regions are very similar to buffer zones, but it is not based on traditional distance. So it need integrate traditional distance to make some spatial query in a specified buffer.

The introduction of k-order spatial neighbour can make the query based on buffer zone become simpler. Furthermore it still guarantees the completeness of the result though the range is reduced from the whole geographic space to adjacent areas of some particular objects.

5.3 Neighboring tree among contours

It is necessary and important to construct and recognize the neighboring relations among contours. K order neighbors can be used to construct the relations tree based on k order neighbors among contours. The tree can be possibly used for forming contour tree and automatic elevation assignment or check. Figure 6 shows a k order relation graph of contours.

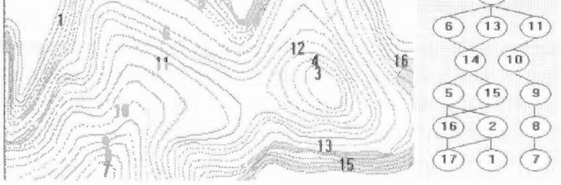


Figure 6. A k order relation graph of contours

5.4 Neighbourhood analysis and reasoning

It is important for spatial reasoning and analysis to construct a suitable neighborhood. Voronoi distance is a kind of means for measuring how far an object is from another in terms of the number of objects, so it can be used for neighborhood analysis and reasoning.

Coefficient for related nearness: Let P is a set of spatial objects in geographical space, if the largest Voronoi distance related to object Pi is n and P_j is k-order neighbour to P_i, the coefficient for related nearness of P_j related P_i (denoted by $reln\,(P_i, P_j)$) is :

$$reln\,(P_i, P_j) = k / n \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

In the above equation, if P_j is the farthest from Pi then $reln\,(P_i, P_j)$ will equal 1, if P_j is the nearest to P_i, then $reln\,(P_i, P_j)$ will equal 0. The coefficient for related nearness is asymmetric.

Nearness predicate $near\,(P_i, P_j)$: taking the value true when object P_j is deemed to be near to object P_i. Assume a fixed parameter p ($0 < p < 1$) and define $near\,(P_i, P_j)$ as follows:

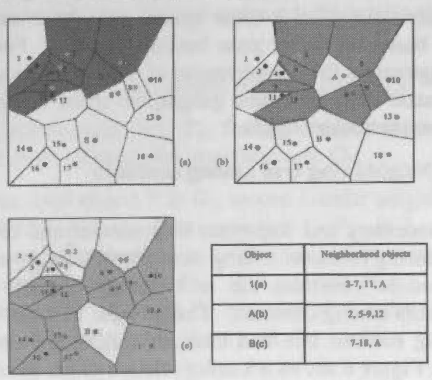
$$near(P_i, P_j) = TRUE \quad if\ and\ only\ if\ reln(P_i, P_j) > p \quad \ldots\ldots\ldots (8)$$

If $near\,(P_i, P_j)$ = TRUE, we will denote it by P_i *near* P_j and say that P_j is *near* to P_i. Note that for $P_i = P_j$, $reln(P_i, P_j) = 0$, so *near* is reflexive, that is for all spatial object P_i belonging P, P_i *near* P_j. But it does not satisfy symmetry or transitivity. However, by introducing a concept of neighbourhood based on k-order neighbors, we are able to get transitivity.

Neighbourhood relationship: Let G be a geographic space in which a adjacent predicate *near* is defined, and a set P of spatial objects is embedded in G. Define a neighbourhood relationship nr on G, so that for all spatial objects P_i, P_j belonging P, P_i nr P_j if and only if, both of the following conditions hold:

1) P_i *near* P_j;

2) For all P_k, if P_i *near* P_k then Pj *near* P_k.



Object	Neighborhood objects
1(a)	2-7, 11, A
A(b)	2, 5-9,12
B(c)	7-18, A

Figure 7: Neighbourhood analysis based on Voronoi distance

Using this Neighbourhood relationship, one can get neighbourhood of P_i as follows,

K-order nearness based neighbourhood: Let G be a geographic space in which a adjacent predicate *near* is defined, and a set P of spatial objects is embedded in G, then for any a object P_i, the neighbourhood of P_i is defined to be the set of all objects neighbour to P_i, denoted by U_{Pi},

$$U_{P_i} = \{P_k \mid P_k \in P, and P_i\ nr\ P_k\} \quad \ldots\ldots\ldots\ldots\ldots (9)$$

The concept of neighbourhood based on k-order nearness U_{Pi} has the properties similar to point-set topological neighbourhood. The result can be further used for the analysis of spatial distribution and determination of graph structure. Figure 7 gives all neighbourhood objects of object *1*, *A* and *B* when *reln* is specified no greater than 0.5.

6 CONCLUSIONS

Spatial neighbour is the foundation of modelling any spatial information system such as GIS. The raster model defines neighbourhood by the neighbour between cells, but it can not directly recognize adjacent relations between spatial objects. The vector model defines neighbourhood by complicated line intersection computation. Voronoi spatial model, on the other hand, has many good properties and is very promising in the application of GIS community. Voronoi neighbours imply both qualitative and quantitative properties. The introduction of Voronoi-based K-order provides a hybrid model for neighbour relations. This study provides a new attempt for spatial query and analysis.

REFERENCES

Aurenhammer, F., 1991, Voronoi diagram—a survey of a fundamental geometric data structure. ACM Computing Surveys,.23(3), pp345-405.

Broke, J. T., 1996, Framework for entropy-based map evaluation. Cartography and Geographic Information systems, 23(2), 78-95

Chen, Jun, Li, C., Li, Z. and Gold, C., 2001, A Voronoi-based 9-intersection model for spatial relations. International Journal of Geographical Information Science, 15(3), 201-220.

Chen, Jun, Zhao, R.L., and Li., Zhi-Lin, 2000, Describing spatial relations with a Voronoi diagram-based Nine Intersection model. In: SDH'2000, Forer, P., Yeh, A. G. O. and He, J. (eds.), pp4a.4-4a.14

Edwards, G., 1993, The Voronoi model and cultural space: applications to the social sciences and humantities. Springer-Verlag, Lecture Notes in Computer Science, No.716: pp202-214

Florence III, John and Egenhofer, M. J., 1996, Distribution of topological relations in Geographic data sets. ASPRS/ACSM. Annual convention and exposition technical papers, pp315-325

Gahegan, Mark, 1995, Proximity operations for qualitative spatial reasoning, In: Spatial Information Theory, Springer-Verlag, Lecture Notes in Computer Science, No. 988, pp31-44

Gold, C. M., 1989, Spatial adjacency-a general approach, Autocart 9, pp298-312

Gold, C. M., 1991,Problems with handling spatial data –the voronoi approach. CISM Journal ACSGC, 45(1), 65-80

Gold, C. M., 1992, The meaning of " Neighbour". In: Theory and methods of spatial-temporal reasoning. Springer-Verlag, Lecture notes in computer science, No.639, pp220-235

Hernandez, Daniel, Clementini, Eliseo and Felice, Paolino Di, 1995, Qualitative distances. Spatial Information Theory, Springer-Verlag, Lecture Notes in Computer Science, No.988, pp45-57

Hong , Jung Hong , Egenhofer, Max J. and Frank, Andrew U., 1995,On the robustness of qualitative distance and direction-Reasoning.Annual convention & exposition technical papers. Auto Carto 12, 301-310

Huang, Zhexue and Svensson, Per, 1993, Spatial query language and analysis In: Advances in Spatial data Bases, Springer-Verlag, Lecture Notes in Computer Science, No.692, 413-436

Li, C.,Chen, J. and Li Z.L., 1999, A raster-based method for computing Voronoi diagrams of spatial objects using dynamic distance transformation. International Journal of Geographic Information Science, 13(3), 209-225

Okabe, A., Boots, B. and Sugihara, K., 1992, Spatial tessellations: concepts and applications of Voronoi Diagrams,published by John Willey & Sons Ltd 1992. pp335-363

Wong, D. W. S., 1999, Several fundamentals in implementing spatial statistics in GIS: using centrographic measures as examples. Geographical Information Sciences, 5 (2), 163-174

Wright, D.J., and Goodchild., M. F., 1997, Data from the deep: implications for GIS community. International Journal of Geographical Information Science, 11, 523-528

Zhang, C. and Murayama, Y., 2000, Testing local spatial autocorrelation using k-order neighbours. International Journal of Geographical Information Science, 14, 681-692

Zhao, R.L. Chen, J. and Li, Zhi-lin, 1999, defining and describing k-order adjacency with Vroonoi distance. In Proceedings of the 2nd International Workshop on dynamic and Multi-dimensional GIS (Beijing: National Geomatics Center of China), pp.77-82.

ACKNOWLEDGEMENTS:

The work described in this paper was substantially supported by a grant from Research Grants Council of the Hong Kong Special Administrative Region (Project PolyU 5048/98E) and partly by the National Science Foundation of China (under No. 69833010).

AN INTERACTIVE SYSTEM FOR PRODUCING POLYHEDRAL AND PRISMATIC BUILDING MODELS USING BUILDING OUTLINE SEGMENTS

Jiann-Yeou Rau[*] Liang-Chien Chen

Center for Space and Remote Sensing Research, National Central University, Chung-Li, Taiwan, R.O.C.
(jyrau, lcchen)@csrsr.ncu.edu.tw

Commission IV, WG IV/1

KEY WORDS: Split-Merge-Shape, Building Modelling, 3-D Line-Segments.

ABSTRACT

This paper presents an interactive system for producing polyhedral and prismatic building models. The core technology, called "Split-Merge-Shape" (SMS) method, reconstruct building models from photogrammetric 3-D line-segments of buildings, i.e. the roof-edges. The method may treat those complete line-segments as well as partial line-segments due to image occlusions. The proposed method comprises four major parts: (1) initialization of a building model and pre-processing of 3-D line-segments, (2) splitting the model by 3-D line-segments to construct a combination of roof-primitives, (3) merging those connected roof-primitives to complete the boundary of each building, and (4) shaping each building rooftop by coplanar fitting. The SPLIT and MERGE processes sequentially reconstruct the topology between any roof edges of the buildings and then reform them as enclosed regions. The SHAPE process uses height information to determine the shapes and heights of the roofs. The results are the polyhedral building models. For most of the 3-D GIS applications the prismatic building models is preferred. So, after modelling the polyhedral building models, prismatic building models can also be created by using an interactive procedure. Experimental results indicate that the proposed method can soundly rebuild the topology among those 3-D line-segments and reconstruct the building models with up to 98% of successful rate. The other 2% of SHAPE failure can be recovered interactively. The proposed system based on the SMS technology can achieve high reliability with high degree of automation even groups of connected buildings and complex buildings are processed. The proposed SMS method can simplify the work of topographic mapping for buildings, and integrated with three-dimensional building modelling thus reducing the operational costs dramatically.

1. INTRODUCTION

1.1 Motivation

Due to progress in computer technology in the establishment, management, and application of city planning, a move toward the combining of 3-D building models with 3-D Geographic Information Systems (GIS) is being made. Traditionally, third dimensional information has been denoted by contour lines or by the number of building stories, which has limited the reliability and feasibility of its applications. Now, with the assistance of computer technology, users can easily analyse spatial data on the fly. Third dimensional ground feature information can be described by geometric models of objects, such as Digital Terrain Models (DTM), Digital Building Models (DBM), or Digital Surface Models (DSM). The primary data for three-dimensional spatial information analysis is the DSM, which denotes all objects on the ground (e.g. trees, buildings, terrain, etc.). The most important feature in urban modelling is the 3-D building model. It is essential in many applications, such as true-orthophoto generation (Rau & Chen, 2002), map revision, cartographic database compilation, change detection, transportation, urban planning, environmental planning (Lange, 1999), flight simulations, noise and air pollution simulations, microclimate studies, wireless networks telecommunication planning (Siebe & Buning, 1997; Leberl, *et al.*, 1999), virtual tourism information querying (Volz & Klinec, 1999), etc. An effective solution for the generation of reliable and accurate building models is thus more urgent than ever.

1.2 Related Works

Since 3-D mapping is increasing in importance, some Digital Photogrammetric Workstations (DPWs) have developed new functions that will create DBMs directly, and integrate them into topographic mapping. However, image occlusion is still the biggest problem in the manual delineation process. The operator is responsible for the measurement of roof corners and the structuring of building models. Whenever image occlusions occur, the operator has to estimate the hidden corners from conjugate images. The whole process is time-consuming and labour intensive, as well as tedious, inefficient and produces with a limited accuracy, especially for connected buildings in densely built-up areas.

Except for the ***manual operations*** necessary for building modelling in some DPWs, most of the related work can be categorized into two approaches, according to the degree of human intervention required for topological reconstruction. The two approaches are the ***automatic approach*** (Baillard. and Zisserman, 2000; Fischer, *et al.*, 1998; Henricsson, 1998) and the ***semi-automatic approach*** (Gülch, *et al.*, 1999; Grün & Wang, 1998).

Gülch, *et al.* (1999) proposed the use of monocular images and building-based measurements for 3-D city model creation. Their approach is based on a ***semi-automatic*** scheme. In manual modelling, the operator's task is to fit a wire-frame building model, which is selected from a predefined model database, to the images, via monoscopic viewing. The operator has to adjust the wire-frame model to fit the corresponding image features. A complex building is decomposed into some basic building types and constructed using a Constructive Solid Geometry (CSG) tree. The operator is also responsible for handling the CSG tree structures. Although the approach is innovative, the operator takes too heavy a responsibility, necessitating a qualified operator. The approach may be efficient for simple structure and specific type of building, but not for groups of connected buildings in densely built-up areas where occlusions and shadows frequently occur.

Grün & Wang (1998) proposed a topology builder, i.e. the CC-Modeler system, for the generation of building models. The system is based on a ***semi-automatic*** approach, utilizing manually measured 3-D point clouds. The measurement of 3-D point clouds, denoting all roof corners must be complete, including any hidden ones. The system is a model-based approach, but it can be applied to general objects, such as roads, rivers, parking lots, ships, etc. During data acquisition, 3-D point clouds are manually coded as boundary points and interior points, according to their functionality and structure. Although the system is operational, some of their limitations cannot be ignored. Those limitations are stated as follows. (1) The measurement of hidden corners caused by building occlusions is necessary to assure the completeness of a roof unit. (2) The structuring may fail if the processed roof unit or object is not in the pre-defined model database. (3) The boundary point digitizing sequence must be restricted to being point-wise. (4) Each roof unit or object needs to be processed independently, which leaves connection problems between two buildings to the operator.

2. POLYHEDRAL BUILDING MODELLING

In the real world, it is difficult to describe all types of buildings using a single comprehensive building model database. In our approach, a building model may be decomposed into several planar ***roof-primitives***. A roof-primitive may be a part or a complete building. Each roof-primitive is a planar rooftop, (e.g. a horizontal or oblique plane), with its boundary projected onto the ground as a polygon. One roof-primitive, or a combination of roof-primitives, can be reformed as a ***polyhedral building model***.

2.1 Initial Building Model

The key for the realization of the whole idea is to create an initial building model, which is the first roof-primitive with a known topology. The initial building model is simply built in such a way that an operator needs only to specify the Area Of Interest (AOI) with a polygon. By the incorporation of a reasonable height, a volumetric representation of the initial building model, which covers all of the 3-D line-segments in a process. Briefly speaking, the creation of initial building model has the following two important meanings: (1) the first building model with a known topology, and (2) the selection of working line-segments.

2.2 Pre-Processing Functions

There are two reasons for pre-processing the manually measured 3-D line-segments. One refers to the adjustment of geometric irregularities due to stereo measurement errors, the other to obtaining a topology error-free solution. Many geometric irregularities due to the errors of manual stereo measurement can happen, such as: (1) two collinear lines are misaligned, (2) rectangular buildings are skewed, (3) two consecutive line-segments intersect and cause overshooting, and (4) gaps due to image occlusions, especially in a densely built-up area, may cause incorrect modelling. These kinds of situations should be solved before building modelling. Since the roof-edges are measured in the 3-D object space, these pre-processing steps are also performed in the object space.

2.3 Split-Merge-Shape

A novel building modelling method, called "SPLIT-MERGE-SHAPE" Method, was proposed by Rau (2002). A brief discussion of the method is described below.

The SPLIT and MERGE processes can sequentially reconstruct the topology between two consecutive line-segments and then reform the areas as enclosed regions. In splitting, one line-segment is chosen as a reference. If any roof-primitives contain this line-segment, we SPLIT them into two. For successive line-segments, a combination of the possible roof-primitives is constructed. The splitting action is similar to the manual inference of hidden corners. The merging procedure is also worked on the 2-D horizontal plane. Every two connected roof-primitives are analysed successively. If the boundary shared between them does not correspond to any 3-D line-segments, the two roof-primitives will be merged into one. The SHAPE process is worked in 3-D object space. The first step of shaping is to assign a possible height for each roof-edge from its corresponding 3-D line-segment. Every roof-edge is automatically labelled as a *shared edge* or an *independent edge* at first. The height information for an *independent edge* can then be assigned and fixed from its corresponding 3-D line-segment. The second step is to define the shape of a rooftop according to the height of the independent edges. If more than two independent edges exist and are sufficient to fit into a planar face, then least-squares coplanar fitting can be applied. Otherwise, the system will provide the most possible solution by ***consecutive-coplanar analysis***, which is used to find a possible planar rooftop using consecutive line-segments or any two non-consecutive but coplanar ones.

2.4 Post-Processing Functions

The proposed SMS system is designed to have some *post-processing functions* for visual inspection and modification. In the visual inspection stage, each generated roof-primitive is examined both on the ground view and 3-D view.

1. The user can *select* the roof-primitive of interest on the ground view. The selected roof-primitive together with the original 3-D line-segments will be shown in the 3-D view.
2. Whenever the selected roof-primitive doest not fit the original line-segments, *reshaping* is necessary. The system will provide all possible solutions by the consecutive-coplanar analysis and show them on the screen one by one. The operator can thus choose the correct one by comparing the original stereo-pair.

3. PRISMATIC BUILDING MODELLING

Complex buildings may be decomposed into many coplanar roof-primitives when a polyhedral model is employed. In general, the attributing of a building is in "house" unit, or belonging to the same "owner". It is practical to relate the roof-primitives that belong to the same owner as one attribute when a complex building is described with a polyhedral model. However, a generalization of this polyhedral models will simplify the above attributing task. As discussed earlier, prismatic models can describe complex buildings with flat roofs, with polygonal ground plans and with vertical walls. It is a generalized form of the polyhedral model, but the attributes of a complex building can thus be directly assigned. Here, we provide a solution to create prismatic models interactively from the existing polyhedral models. The techniques used are similar to the SMS method. The generation of prismatic model includes the following three steps:

3. Manually select the neighbouring roof-primitives that belong to the same "house". Since the selection of roof-primitives that belong to the same "house" unit needs to be accurate, an operator needs to have a priori information about the house unit. Cadastral maps, for instance, are needed when ownership is considered. This step is performed on the ground view. For example, the operator selects five roof-primitives that belong to the same house unit, i.e. denoted in green in figure 1(a). In the 3-D view, i.e. in figure 1(b), the selected roof-primitives are chosen and also depicted in green.
4. After selection, the operator automatically fuses them into one house unit, using a technique similar to the SMS MERGE step. The major difference is that shared boundaries should be totally removed. In the meantime, the system provides an estimation of the possible height value, i.e. the green one in figure 1(c). The height estimation is designed to find the lowest height for the greatest area. However, this may not meet the topographic mapping protocol, since this estimation may influence the calculation of the number of stories. Therefore, we provide an interactive mechanism for the operator to make the best choice.
5. In the third step, i.e. shape, the operator can then change the height iteratively, using an interactive mechanism similar to the SMS shaping process. The provided height values are estimated automatically from the original five roof-primitives. Notice that roofs are forced to be flat.

Using of the same procedures the other parts can be performed completely, as shown in figures 1(d), 1(e), and 1(f). The previous fused result is depicted in pink. Finally, we can use the 3-D visualization tool to verify the final prismatic models, i.e. figure 1(g). For the purpose of comparisons, the original polyhedral models are illustrated in figure 1(h). The effect of the generalization from a polyhedral model is quite obvious. The provided interactive procedures are easy to use. No complicated designs and no tedious operations are required.

4. CASE STUDY

In theory, the proposed SMS method can cope with partial occlusion problems. Except for the manually measured visible roof-edges from the quadruplet, a digital topographic map of building outlines are also tested to evaluate the robustness, efficiency, and accuracy of the proposed method.

4.1 Line-Segments from Visible Roof-Edges

The test data set was measured manually using a DPW. The content of this data set can be abstractly categorized into three parts. Part (I) is a university campus, i.e. Fu-Zen University, TAIWAN. The buildings are large with complex boundary, and are separated to each other with a distance. Part (II) is a high-density built-up area with groups of connected and rectangular buildings. Part (III) is a high-density built-up area with groups of connected, complex rooftops, and less-orthogonal buildings. Figure 2 depicts the above three areas on one of the original aerial-photos superimposed by manually measured 3-D line-segments. The number of measured roof-edges is 6,363. Figure 3 depicts the reconstructed polyhedral building models for this data set displayed using the 3-D visualization tool.

The number of roof-primitives created using the SMS method was 1,809. The splitting and merging process was totally successful after the correction of blunder measurements. However, 38 roof-primitives failed at the shaping stage, giving a success rate of 98%. The 2% failure rate was totally recovered using the provided *post-processing functions,* without any manually editing of building models. These failures occurred mostly in Part (III), where the buildings were connected with a less-orthogonal structure and the rooftops were complex that the delineation of roof-edges is not coplanar in nature. The performance was totally satisfactory for Part (I) and Part (II), where the buildings were rectangular in structure, although connected, or with complex boundaries. This investigation demonstrated the robustness of the proposed method that diversified types of buildings could be modelled.

In order to investigate the robustness in detail, two types of building group are demonstrated. The first type of buildings, depicted in figure 4(a), contains a complex building with a combination of circular and rectangular boundaries. The building is a hip roof, gable roof, flat roof composite, and also has two donut components with courtyards inside. Under a PC with Pentium III 800Mhz CPU, the process time of modelling for this group of buildings is less than one second. For the circular boundary, a series of consecutive line-segments are measured. The second type of building, shown in figure 4(b), contains a group of connected and mixed rooftop types. The number of 3-D line-segments is 209, which generates 69 roof-primitives at last. The process time for this group of buildings takes only 1.5 seconds. The hidden effects are problematic for manual measuring and building modelling. The SMS method is proven to be robust for the second type of buildings, where image occlusions frequently happen. It is also efficient even for a group of buildings is processed and which is suitable for a semi-automatic approach of building modelling.

The accuracy of the generated building models mostly depends on the accuracy of the manual measurements. The estimated locations of hidden corners were less accurate than the direct measurement of visible ones. In order to evaluate the modelling error, we considered the manual measurements of visible corners as reference data, so evaluations could be performed accordingly. We measured the parts-of-the-roof-edges that makes the totally visible roof-edges incomplete. The total number of visible corners being evaluated is 163. After applying the SMS method for building modelling, it achieves a MEAN error of 1.06 *cm*, 1.22 *cm*, and 2.73 *cm* on the X, Y and Z-axis, respectively. In the mean time, a Root-Mean-Square Error (RMSE) of 13.5 *cm*, 14.5 *cm*, and 34.9 *cm* on the X, Y and Z-axis, respectively, is achieved. Since the original stereo-pair has a nominal ground sampling distance of 12.5 cm and a base-height-ratio of 0.3, the RMSE is close to be one pixel on the image scale, which falls into the range of random errors. It is proven that the proposed SMS method is accurate in modelling.

4.2 Creating Prismatic Models

Here we use the polyhedral building models generated from the previous section to create the prismatic models. Here we present a case study of building models imported to the Arc View 3D Analysis module. A 3-D building stories query procedure is demonstrated. The determination of the number of stories needs information about the *real height* of a building, which is a subtraction of the *absolute height* from the *terrain*. Therefore, DTM data is a must in the conversion process. The DTM we used is part of a topographic database on Taiwan. We assume the buildings have an average height of 3.3 *m* per story. Thus, the number of stories in a building is determined by dividing the real height by 3.3 *m*. This attribute, together with the prismatic models, can be transformed into the Shapefile format and analysed using the ArcView software, as shown in figure 5. The user can pan or zoom using 3-D roaming. The users can also query the number of stories in a building, e.g. the upper-left window in figure 5. The number of stories in each building is differentiated by colour, e.g. the middle-left window in figure 5.

4.3 Building Outlines from a Digital Topographic Map

In the creation of digital topographic maps, the stereo-measurements are mapped onto the ground plane to produce planimetric maps. This means that the measurements are made in three-dimensional space. Figure 6, for instance, shows a digital topographic map of Kao-Hsiung City, TAIWAN. The map scale is 1:1,000 and it contains a number of detailed ground features such as buildings, roads, trees, fence, and so on. We select the building layer and project it back onto the original aerial photo, as shown in figure 7. Since this is a large-scale urban topographic map, the delineation of building outlines is very detailed. By applying the proposed SMS method we can generate a 3-D city model. The result is illustrated in figure 8 using the 3-D visualization tool. This experiment depicts the following important issues.

1. Conventional stereo-measurement of buildings in the original 3-D line-segments form can be efficiently and directly converted to make a city model, which means that existing urban topographic maps with 3-D building outlines are useful for 3-D city model generation.
2. The protocol for the topographic mapping of buildings may be evolved to only measure the visible part of roof-edges. The manual post-editing of a topographic map on the X-Y plane can be replaced by the proposed SMS method automatically. This may reduce the workload of the operator during post-editing phase. Moreover, geometrically regular buildings can be restored automatically by the pre-processing functions.
3. We provide the possibility of integrating topographic mapping with 3-D city modelling, to make the 3-D mapping more cost-effective.

5. CONCLUSIONS

This paper presents a novel method for geometrical building modelling. The proposed Split-Merge-Shape method that adopts manually measured visible roof-edges for building modelling is innovative. Table 1 compares the SMS system with two other semi-automatic approaches as described in section 1.2. From the comparison, one finds that our method appears advantages in many aspects. The major features of the method are:

1. It can handle a group of buildings simultaneously to avoid possible topology errors in neighbouring buildings.
2. It can cope with partial occlusion problems.
3. It can deal with diverse types of buildings.
4. It is proven to be robust, efficient, and with high accuracy.
5. Due to only visible roof-edges are required and the digitising sequence is free, the operator's workload is reduced as well as the cost of production.
6. The designed SMS system can create polyhedral and prismatic building models in an efficient way. It fulfills the requirements for many applications not only in the field of photogrammetry, but also in GIS, remote sensing, computer vision, environmental studies, etc.
7. An existing digital topographic map that includes buildings outlines can be directly converted to be a 3-D city model without any additional manual measurement.
8. By means of the SMS technology, the integration of topographic mapping with 3-D building modeling is possible. A cost-effective environment could be established for 3-D mapping.

6. ACKNOWLEDGEMENT

This work was partially supported by the National Science Council, Taiwan, under project number NSC 89-2211-E-008-086. The authors like to thank Prof. S.C. Wang, National Cheng-Kung University, and Mr. J.S. Gao, the Eagles Flying Survey Technology Co., for providing the test data.

7. REFERENCES

Baillard, C. and A. Zisserman,(2000). A plane sweep strategy for the 3D reconstruction of buildings from multiple images, *Proc. of International Archives of Photogrammetry and Remote Sensing,* Vol.33, B3, Amsterdam, Netherlands, pp.56-62.

Fischer, A., T. H. Kolbe, F. Lang, Cremers, A. B., Förstner, W., Plümer, L. & Steinhange,V.,(1998). Extracting Buildings from Aerial Images Using Hierarchical Aggregation in 2-D and 3-D, *Computer Vision and Image Understanding,* Vol.72, No.2, Nov. pp.185-203.

Grün, A. and X. Wang,(1998), CC-Modeler: A Topology Generator for 3-D Building Models, International Journal of Photogrammetry and Remote Sensing, Vol.53, pp.286-295.

Gülch, E., H. Muller & T. Labe,(1999). Integration of Automatic Processes Into Semi-Automatic Building Extraction, *Proc. of ISPRS Conference "Automatic Extraction Of GIS Objects From Digital Imagery",* September 8-10, (INVITED).

Henricsson, O.,(1998). The Role of Color Attributes and Similarity Grouping in 3-D Building Reconstruction, *Computer Vision and Image Understanding,* Vol.72, No.2, pp.163-184.

Lange, E.,(1999). The degree of realism of GIS-based virtual landscapes: implications for spatial planning, *Photogrammetric Week* (Fritsch, D. and Spiller, R., Eds.), Wichmann, Karlsruhe, pp. 367-374.

Leberl, F., W. Walcher, R. Wilson and M. Gruber,(1999). Models of urban areas for line-of-sight analyses, *Proc. of International Archives of Photogrammetry and Remote Sensing,* Vol.32, Part 3-2W5, Munich, Germany, pp.217-226.

Rau, J.Y. and L.C. Chen,(2002), True Orthophoto Generation of Built-Up Areas Using Multi-View Images, *PE & RS,* (Accepted).

Rau, J.Y.(2002), Geometrical Building Modeling and Its Application to the Ortho-Rectification for Aerial Images, Ph.D. Thesis, National Central University, R.O.C.

Siebe, E. and U. Buning,(1997). Application of digital photogrammetric products for cellular radio network planning, *Photogrammetric Week* (Fritsch, D. and Spiller, R., Eds.), Wichmann, Karlsruhe, pp.159-164.

Volz, S. and D. Klinec,(1999). Nexus: the development of a platform for location aware application, *Proc. of the third Turkish-German Joint Geodetic Days,* Vol.2, Istanbul, Turkey, pp. 599-608.

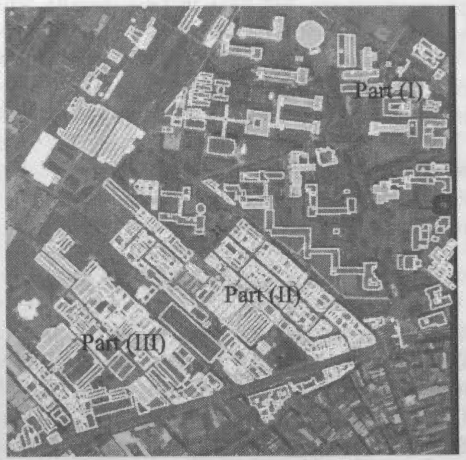


Figure 2. Original aerial-photo with delineated roof-edges.

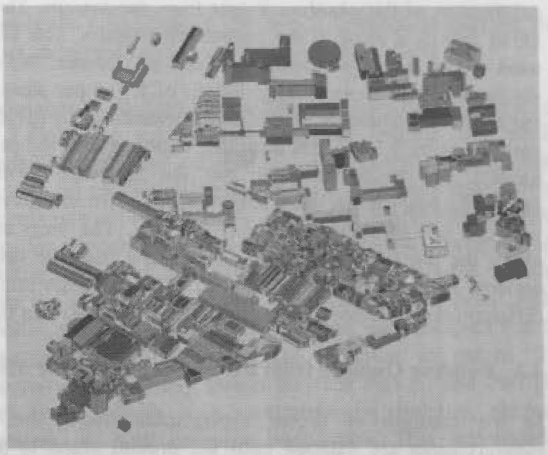
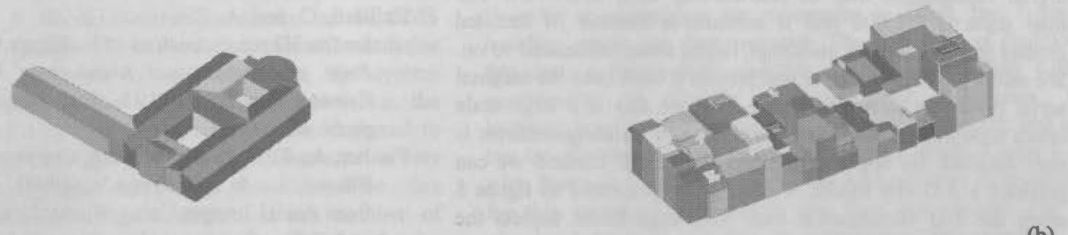


Figure 3. The generated 3-D building models in a 3-D view.



(a)

(b)

Figure 4. Two detail example of generated building models.

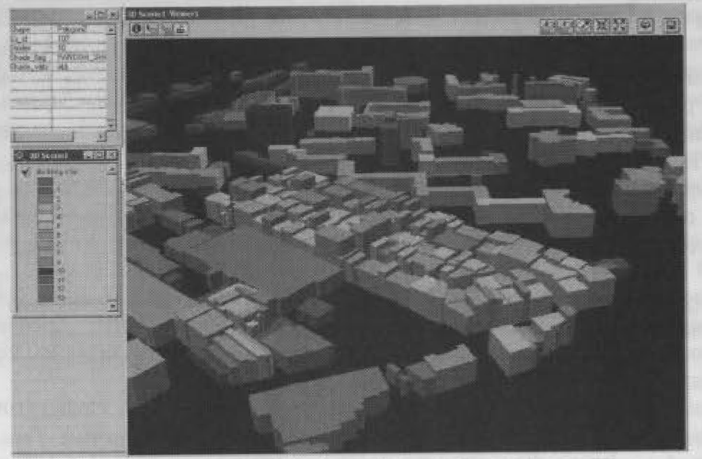


Figure 5. An example of 3-D querying in the ArcView environment.

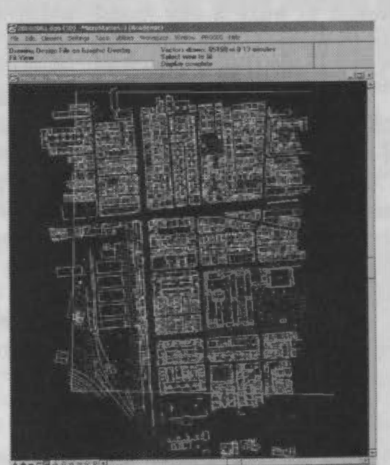


Figure 6. Example of a digital topographic map.

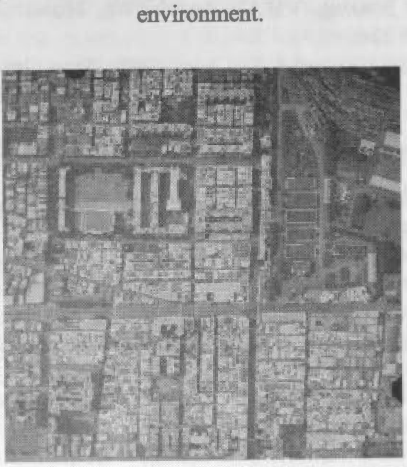


Figure 7. A digital topographic map of building outlines superimposed onto the aerial photo.

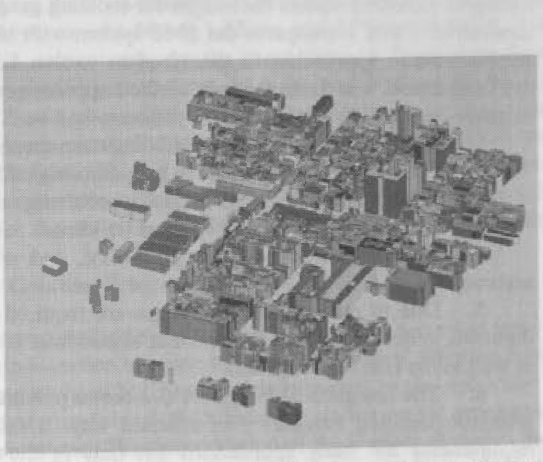


Figure 8. 3-D city model generated from a digital topographic map of building outlines.

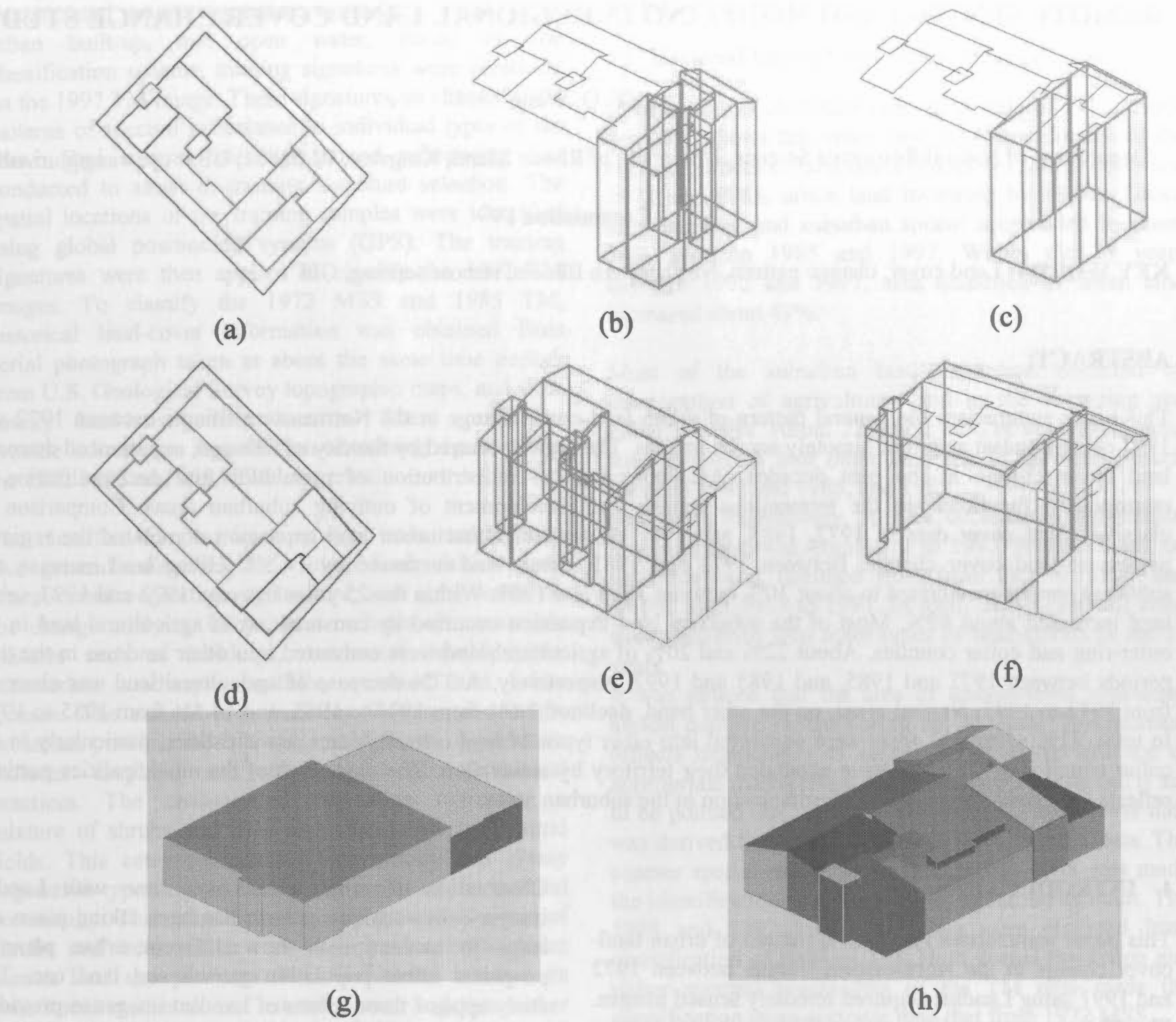


Figure 1. Procedure for creating prismatic models.

Table 1. Comparisons of the proposed SMS system with two semi-automatic approaches.

	SMS	Grün & Wang (1998)	Gülch, et al. (1999)
Viewing	***Stereo***	Stereo	Mono In Multiple Images
Measurement	***Incomplete 3-D Roof-Edges***	Complete 3-D Roof Corners	Building Primitives
Digitizing Sequence	***Free***	Point-Wise for Boundary Points	Free
Hidden Corners	***Automatically Generated***	Needs Measurement	Not Measured
Can Handle a Group of Connected Buildings	***Yes***	No	No
Model-Based	***No***	Yes	Yes
Automation	***Building Model Reconstruction***	Object Model Reconstruction	Height Determination Parameter Estimation
Accuracy	***Photogrammetry***	Photogrammetry	Photogrammetry
Outputs	***Polyhedral & Prismatic Models***	Polyhedral Objects	CSG Structure of Buildings
Post Processing	***Needed***	Needed	Needed

REMOTE SENSING AND MODELING IN REGIONAL LAND COVER CHANGE STUDY

Y. Q. Wang

Department of Natural Resources Science, University of Rhode Island, Kingston, RI 02881, USA - yqwang@uri.edu

Commission IV

KEY WORDS: Land cover, change pattern, Northeastern Illinois, remote sensing, GIS

ABSTRACT:

This paper summarises the general pattern of urban land-cover change in the Northeastern Illinois between 1972 and 1997 using Landsat acquired remotely sensed images. The region, centred by the city of Chicago, experienced dramatic land cover change in the past decades. Accompanying the redistribution of population and decentralisation of metropolitan functions are the tremendous growth and development of outlying suburban areas. Comparison of classified land cover data of 1972, 1985, and 1997 demonstrates that urban land expansion dominated the regional pattern of land cover change. Between 1972 and 1985, urban land increased by 14.5%. Urban land increase and suburban sprawl accelerated to about 30% between 1985 and 1997. Within the 25 years between 1972 and 1997, urban land increased about 49%. Most of the suburban land expansion occurred by consumption of agricultural land in the outer-ring and collar counties. About 22% and 20% of agricultural land were converted into other land use in the time periods between 1972 and 1985, and 1985 and 1997, respectively. A 37% decrease of agricultural land was observed from 1972 to 1997. Natural areas, on the other hand, declined 7.6% from 1972 to 1985, and 14.5% from 1985 to 1997. In total, 21% of natural areas were converted into other types of land covers. Many municipalities, particularly in the collar counties of Chicago, have expanded their territory by annexation. The expansion of the municipalities partially reflects the extent and potential of urbanization in the suburban areas.

1. INTRODUCTION

This paper summarises the general pattern of urban land-cover change in the Northeastern Illinois between 1972 and 1997 using Landsat acquired remotely sensed images. The Chicago metropolitan region, like most of the other metropolitan areas in the country, experienced dramatic land cover change in the past decades. Accompanying the redistribution of population and decentralization of metropolitan functions are the tremendous growth and development of outlying suburban areas. An important step towards a regional analysis is to understand the pattern and the quantitative results of urban land change. The objectives of this study are to obtain a quantitative description of the spatial pattern of urban land cover changes; to help in understanding the relations between urban land cover change and its driving factors; and to provide technical protocol and background data for a regional impact study. Eight northeastern Illinois counties, which encompass the city of Chicago, its suburbs and the surroundings, were the study area, including Cook, DuPage, Grundy, Kane, Kendall, Lake, McHenry, and Will counties.

Remote sensing has been broadly applied in land cover mapping and land cover change detection (Green et al., 1994; Clarke and Gaydos, 1998; Vogelmann, et al., 2001; Wang and Moskovits, 2001). Landsat remote sensing images provide a critical data source for this study. Observation of urban areas over time with Landsat imagery shows where growth has been taking place and helps in evaluation of how different urban planning programs affect population growth and land use. This study applied three scenes of Landsat images to provide a historical Landsat view of the urban land change in the Chicago metropolitan region. Based on the processed Landsat images, the quantitative descriptions of the landscape and interpreted patterns of urban land-cover change were obtained.

2. METHOD

Land cover maps of the region were derived from classification of the Landsat Multispectral Scanner (MSS) imagery data acquired on October 2, 1972 and the Thematic Mapper (TM) data acquired on May 2, 1985 and October 10, 1997. The 1972 MSS data was the first cloud-free Landsat imagery available for the Chicago region. These digital images were geometrically rectified and georeferenced into the Universal Transverse Mercator (UTM) map coordinates. An interpolation process was applied to resample the MSS data into 30-m pixels so that the same spatial resolution was maintained for the MSS and TM imagery data.

This study applied a classification scheme that has eleven (11) land cover categories including forest, woodland, savanna, prairie, wetland, unassociated woody vegetation, unassociated grassy vegetation, agriculture, urban grass, urban built-up, and open water. Based on the classification scheme, training signatures were identified on the 1997 TM image. These signatures tie characteristic patterns of spectral reflectance to individual types of the above land covers. Intensive ground referencing was conducted to assist in training signature selection. The spatial locations of the training samples were identified using global positioning systems (GPS). The training signatures were then applied to classify the 1997 TM images. To classify the 1972 MSS and 1985 TM, historical land-cover information was obtained from aerial photograph taken at about the same time period, from U.S. Geological Survey topographic maps, and from county land-management records. These references provide information for selection of training signatures on the 1972 MSS and 1985 TM imageries.

The classified Landsat imagery data were recoded into five general categories, *i.e.:* urban land, natural area, unassociated vegetation, agriculture, and open water for the change analysis. Urban land includes all man-made features such as buildings, residential developments, cemeteries, roadways, landfills, quarries, and urban grasses. The natural area includes forests, woodlands, savannas, prairies, shrub lands, and wetlands. Agriculture includes cropland and other types of agricultural practices. The unassociated vegetation represents a mixture of shrubs and trees and abandoned agricultural fields. This category includes both woody and grassy vegetation types and is considered a part of the cultural community. The unassociated woody growth is a mix of shrubs and trees which owe their existence to recent human land use practices. It is so named because its constituent species do not naturally occur together, either historically, or as associates in long-term self-perpetuating communities. The unassociated grassy areas are relatively open fields, which have developed after weedy alien and native herbaceous species colonize barren soils or decimated natural areas. Most unassociated grassy areas occur as re-colonization of plants in the years following the cessation of agricultural cultivation or grazing.

To eliminate the influence by possibly misclassified individual pixels, image processing was applied to filter the classified land cover data. The area of three adjacent pixels was defined as the minimum spatial unit. Therefore, any patches of land cover types that contained less than 3 pixels in size (smaller than 0.5 acres) were eliminated from the change analysis.

Accuracy assessment of land cover classifications indicates that high overall accuracy was achieved for the generalized land cover data. About 90% accuracy was obtained for the 1972 land cover data and about 93% accuracy was obtained for the 1985 and 1997 land cover data. For urban land, about 87% accuracy was achieved.

3. RESULT

3.1 Regional Urban Land Cover Change

Comparison of classified land cover data of 1972, 1985, and 1997 shows that urban land expansion dominated the regional pattern of land cover changes (Table 1). Between 1972 and 1985, urban land increased by 14.5%. Urban land increase and suburban sprawl accelerated to about 30% between 1985 and 1997. Within the 25 years between 1972 and 1997, area classified as urban land increased about 49%.

Most of the suburban land expansion occurred by consumption of agricultural land in the outer-ring and collar counties. About 22% and 20% of agricultural land were converted into other land use in the time periods between 1972 and 1985, and 1985 and 1997, respectively. A 37% decrease of agricultural land was observed from 1972 to 1997. Natural areas, on the other hand, declined 7.6% from 1972 to 1985, and 14.5% from 1985 to 1997. In total, 21% of natural areas were converted into other types of land covers in the 25 years.

It was noticed that the area of unassociated vegetation almost doubled in size during the 25 years. This increase reflects degradation of natural lands in the absence of appropriate management and ecological restoration. It has to be pointed out, however, that the 1972 land cover data was derived from classification of Landsat MSS data. The coarser spatial resolution of the original MSS data made the identification of unassociated vegetation difficult. The 1985 and 1997 land cover data were obtained from classification of Landsat TM. High spatial resolution and richer spectral bandwidths of the TM data made the classification more accurate than that from 1972 MSS.

	1972–1985 (hectares)	1985-1997 (hectares)	1972–1997 (hectares)
Urban Land	+33,011 (14.5%)	+78,114 (30%)	+11,1125 (48.85%)
Natural Area	-17,251 (7.6%)	-30,735 (14.65%)	-47,986 (21.13%)
Agriculture	-130,658 (21.8%)	-93,267 (19.89%)	-223,925 (37.35%)
Unassociated Vegetation	+118,711 (147.2%)	+45,361 (22.75%)	+164,072 (203.45%)

Table 1. Land cover change in metropolitan Chicago region between 1972 and 1997.

3.2 Spatial Pattern of Urban Land Cover Change

Distance from downtown Chicago is an important factor controlling the region's landscape change. A half-ring pattern is clearly observable from both Landsat images and the derivatives of land cover maps. To quantify the spatial pattern of urban land distribution and the changes, the region was divided into five concentric zones. This zones separated the region into 0-15 km (Zone I), 15-30 km (Zone II), 30-45 km (Zone III), 45-65 km (Zone IV), and > 65 km (Zone V) areas measured from the center of Chicago (Figure 1). The proportion of land converted into urban land use in different zones was compared by calculating the percentage of increase and density of change of the urban land (Table 2).

The result indicates that Zone III, the area between 30 km and 45 km from downtown Chicago, had the highest change density (11.54 ha./km^2) and a high rate of urban land increase (62%) between 1985 and 1997. Zone IV experienced about 76% urban land increase and ranked second in change density (7.26 ha./km^2). This pattern agreed well with the facts that many suburban developments fell into these two zones. The two zones contain many residential and commercial development sites in the northwest Cook County, the southern part of Lake and McHenry counties, most of DuPage County, the eastern Kane County, and the northern Will County. About 22% of urban increase in Zone III and 37% in Zone IV were converted from agriculture land.

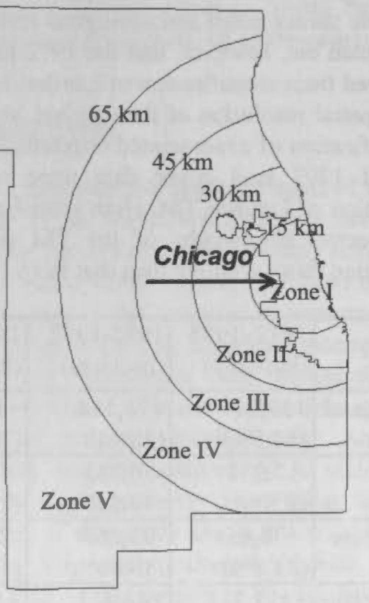


Figure 1. Study area and distance zones.

Land cover change	Urban Land Change (%) 1985-1997	Change Density (ha./km^2)
Zone I (0-15 km)	2.03	1.63
Zone II (15-20 km)	16.92	6.84
Zone III (30-45 km)	62.01	11.54
Zone IV (45-65 km)	76.11	7.26
Zone V (> 65 km)	52.81	4.04

Table 2. Urban land cover change by distance from downtown Chicago

3.3 Changes in Municipality Boundaries

Many municipalities, particularly in the collar counties, have expanded their territory by annexation in the past 25 years (Figure 2). The expansion of the municipalities partially reflects the extent and potential of urbanization in the suburban areas. As a result of expansion, more planned development has been carried out within these municipalities. To depict the patterns of land expansion in the region, the changes of municipality boundaries were extracted from the census data. The relations between the urban land changes and the expansions of the

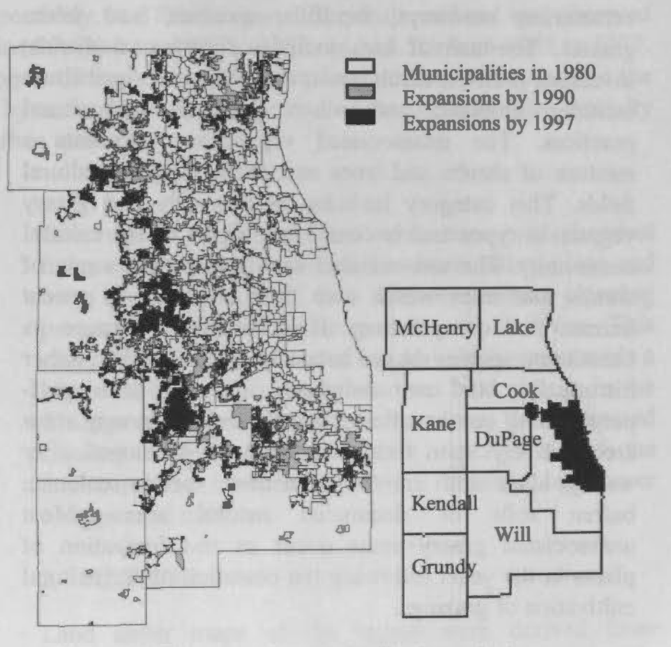


Figure 2. Changes of municipality boundaries in Chicago area between 1980 and 1997.

Municipality Name	Area Changed (km^2)	Population Changed	P/A Ratio
Crystal Lake	18.99	6,106	321.54
Long Grove	18.31	2,734	149.32
Naperville	16.36	43,476	2,657.46
Aurora	15.09	18,263	1,210.27
Frankfort	13.49	2,823	209.27
Antioch	11.93	1,686	141.32

Gurnee	11.37	6,536	574.85
Wadsworth	10.95	722	65.94
Bolingbrook	9.48	3,582	377.85
Bartlett	9.08	6,141	676.32
Vernon Hills	8.91	5,492	616.39
Orland Park	8.86	12,675	1,430.59
Crest Hill	8.79	1,747	198.75
Lake in the Hills	8.67	249	28.72
Buffalo Grove	8.43	14,187	1,682.92
Woodridge	7.64	4,037	528.4
Joliet	7.56	-739	-97.75
Waukegan	7.17	1,828	254.95
Lakemoor	7.03	442	62.87

Table 3. Selected municipalities that changed the most in areas between 1980 and 1990

Municipality Name	Area Changed (km^2)	Population Changed	P/A Ratio
Bolingbrook	20.86	10,469	501.87
Huntley	20.72	576	27.8
Joliet	19.04	9,532	500.63
Naperville	16.71	21,195	1,268.4
Plainfield	14.38	2,783	193.53
Orland Park	13.4	9,937	741.57
Grayslake	12.6	6,982	554.13
Tinley Park	11.22	6,195	552.14
West Chicago	11.2	2,084	186.07
Romeoville	10.82	1,395	128.93
New Lenox	10.38	3,390	326.59
Lake in the Hills	10.11	10,989	1,086.94
Algonquin	10.01	6,326	631.97
Aurora	9.96	16,849	1,691.67
St. Charles	7.89	3,076	389.86
Channahon	7.66	2,074	270.76
Montgomery	6.85	790	115.33
Harvard	6.69	882	131.84

Table 4. Selected municipalities that changed the most in areas between 1990 and 1997

Municipality Name	Area Changed (km^2)	Population Changed	P/A Ratio
Naperville	33.07	64,671	1,955.58
Bolingbrook	30.34	14,051	463.12
Joliet	26.61	8,793	330.44
Aurora	25.05	35,112	1,401.68
Crystal Lake	23.53	13,590	577.56
Huntley	22.89	1,383	60.42
Orland Park	22.26	22,612	1,015.81
Lake in the Hills	18.78	11,238	598.4
Long Grove	18.11	4,045	223.36
Grayslake	17.38	9,110	524.17
Plainfield	17.15	3,573	208.34
New Lenox	16.53	7,296	441.38
Gurnee	16.41	15,310	932.97
West Chicago	16.32	4,342	266.05
Wadsworth	15.4	1,208	78.44
Frankfort	15.38	5,250	341.35
Romeoville	15.26	-23	-1.51
Algonquin	14.08	12,185	865.41
St. Charles	13.46	8,204	609.51
Antioch	13.26	2,979	224.66

Table 5. Selected municipalities that changed the most in areas between 1980 and 1997

municipalities was then obtained using multiple year land cover data and the changes of the municipality boundaries. Population changes within the municipalities were extracted as well (Table 3-5).

The results show that municipalities undergoing the greatest expansion lay within and around the areas of Zone III and Zone IV. The changes of the municipality boundaries were more evident between 1990 and 1997 than 1980 and 1990. Between 1980 and 1990 the municipalities of Crystal Lake, Long Grove, Naperville, Aurora, Frankfort, Antioch, Gurnee, and Wadsworth ranked near the top increasing by over 10 km^2 in areas. Crystal Lake ranked first in area increase (about 19 km^2) but ranked only 30th in population increase (6,106). Naperville, on the other hand, ranked third in area expansion (16.36 km^2) but first in population increase (43,476). To reflect the intensity of urban land development, the ratio between population change and the area increase (P/A) was calculated. The P/A ratio served as an indicator for the intensity of urban development. For example, both Naperville and Wadsworth had over 10 km^2 land expansion. Naperville experienced a very intensive urban development (P/A=2657.46/km^2), while Wadsworth had a relatively slower pace in development (P/A=65.94). Therefore, the municipalities with lower P/A ratio should have a greater potential or available land for urban development in the future. The lower P/A ratio also indicates that the area is in a lower density pattern of suburban development rather than an intensified pattern of new urban development.

Bolingbrook ranked 9th in area expansion between 1980 and 1990. However, this municipality ranked the first in area expansion between 1990 and 1997 with 10,469 population increase (Table 4). The second ranked Huntley, on the other hand, had about the same area expansion as the Bolingbrook but its population increased by only 576 during the same time period. The big difference in P/A ratio between the top ranked two municipalities in area expansion indicated that they were taking different development actions in the time period.

Examination of the changes for the municipalities from 1980 to 1997 shows that Naperville, Aurora, Orland Park, and Gurnee are the top four in P/A ratio (Table 5). Satellite derived land cover maps agreed that the most intensive urban development happened in these areas. The municipalities that ranked high in P/A ratio are mostly located within or beyond Zone IV. The combination of satellite derived land cover maps and the changes of municipalities show that the most intense urban land consumption is occurring in the suburban areas most outlying from the central city.

4. DISCUSSION

This study depicts the land cover change and the pattern within the greater eight-county metropolitan Chicago region during the 25 years between 1972 and 1997. Urban land expansion and suburban sprawl dominated the regional land cover change in this time period. Quantitative comparison and visual maps provide strong evidence of the expansion of suburban sprawl in the region.

This study agreed well with the result derived by other institutions using different methodology and technical approaches. A recent report by the Northeastern Illinois Planning Commission (NIPC) states that developed land in metropolitan Chicago increased 49% between 1970 and 1990, whereas its population grew by 4%. NIPC projected that the population of the region will increase by 25% by the year 2020 (NIPC, 1999). It is certain that suburban sprawl and other urbanization processes will continue there.

Historical Landsat remotely sensed data and the Landsat-7 ETM+ imagery data are critical in development of quantitative analysis models and in simulation of the effects of urban land cover change on regional natural and cultural systems. Based on the research finding from this study, coupled with the population and employment projections made by the census and planning agencies (NIPC, 1998), simulation of the region's land cover change to the year 2020 became possible (Wang and Zhang, 2001). Modeling analysis by integration of driving factors and land cover maps derived from remote sensing provide reliable information to predict the effects of suburban sprawl and the potential impact.

REFERENCES

Clarke, K.C., Gaydos, L.J., 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore, *International Journal of Geographical Information Sciences*, 12, 699-714.

Green, K., Kempka, D., and Lackey, L., 1994. Using remote sensing to detect and monitor land-cover and land-use change, *Photogrammetric Engineering and Remote Sensing*, 60(3):331-337.

NASA, 1999. NASA Landsat-7 Press Kit, http://pao.gsfc.nasa.gov/gsfc/newsroom/pkits/landsat7.pdf.

NIPC, 1999. Population, household and employment forecasts for northeastern Illinois 1990 to 2020, Northeastern Illinois Planning Commission, Chicago, 14p.

Vogelmann, J.E., St.M. Howard, L. Yang, C.R. Larson, B.K. Wylie, and N.V. Driel, 2001. Completion of the 1990s National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources, *Photogrammetric Engineering & Remote Sensing*, 67(6): 650 - 662.

Wang, Y. and D.K. Moskovits, 2001. Tracking Fragmentation of Natural Communities and Changes in Land Cover: Applications of Landsat Data for Conservation in an Urban Landscape (Chicago Wilderness), *Conservation Biology*, 15(4): 835-843.

Wang, Y. and X. Zhang, 2001. Dynamic Modeling Approach to Simulating Socioeconomic Effects on Landscape Change, *Ecological Modelling*, 140(1-2): 141-162.

ACKNOWLEDGEMENT

The research project upon which this paper is based was supported by National Aeronautics and Space Administration (NASA) grant NAG5-8829 (NIP in Earth Science).

AN INVERSE ANALYSIS OF UNOBSERVED TRIGGER FACTORS OF THE SLOPE FAILURES BASED ON STRUCTURAL EQUATION MODELING

Hirohito KOJIMA* and Shigeyuki OBAYASHI*

* Remote Sensing Lab., Science University of Tokyo, 2641 Yamazaki Noda-City, JAPAN 278-8510
kojima_h@rs.noda.sut.ac.jp

KEY WORDS: slope stability evaluation, SEM, trigger factor, inverse analysis, satellite data, and geographical information

ABSTRACT:

This paper presents an inverse analysis of the unobserved trigger factors with respect to the slope failures based on the Structural Equation Modeling (SEM). The quantitative prediction models generally construct the relationship between the past slope failures and the causal factors (i.e. lithology, soil, slope, aspect, etc.), but does not deal with the trigger factors(i.e. rail fall, earthquake, etc.), due to the difficulties of pixel-by-pixel observation of the trigger factors itself. As a measure, in this study, an inverse analysis algorithm on the "trigger factors" is proposed, which consists of the following steps:

- **Step1:** The relationship between the slope failures (as the endogenous variables), the causal and trigger factors (as the exogenous variables) are delineated on the path diagram used in SEM.
- **Step2:** The regression weights in the path diagram are estimated in minimizing the errors between the observed and reemerged "variance-covariance matrix" by the model.
- **Step3:** As an inverse estimation, through the measurement equation in SEM between the causal and the trigger factors, a "Trigger Factor Influence map (TFI map)" is newly produced.

As an application, the TFI maps are produced with respect to the "slope failures" and the "landslides," respectively. Furthermore, as a final product, the difference of those TFI maps are delineated on a "Difference map (DIF map)." The DIF map and its interpretation are indeed useful not only for assessing the hazardous area affected by the trigger factors, but also as a "heuristic information" for locating places for setting the field measuring systems.

1. INTRODUCTION

"When, Where and What scale" of the slope failures and the landslides are the important aspects to endure the person's life as well as the social- and economical-infrastructures against the unpredictable risk. Due to the limitation of the detail field investigation, the research approaches applying the satellite remote sensing data and the various kinds of geographical information (termed "causal factor") are highly expected for identifying the hazardous area affected by the slope failures and landslides as well. However, under the present situation, the quantitative prediction models for the slope failures deals with only those causal factors (Carrara et al., 1995; Chung et al., 1995; Kasa et al., 1991; Obayashi et al., 1999), but not applying the "trigger factors", such as the local downpour, earthquake, weathering, etc., because of the difficulties of pixel-by-pixel observation of the trigger factors itself. As another viewpoint against the previous researches, the trigger factors should be treated as "unobserved factors" in terms of time and space in prediction. To estimate such trigger factors, the substantial questions are

- How can we incorporate the "trigger factors" in prediction modeling? ; and,
- Is it possible to estimate the "trigger factors," quantitatively?

With those issues as background, we have tackled the following outstanding subjects:

- To construct an inverse-analysis algorithm for the "trigger factors", based on the Structural Equation Modeling (SEM).
- To produce the Trigger Factor Influence maps (termed TFI map) with respect to the slope failures and the landslides, as well as to consider its application for the landslide hazard assessment.

2. STUDY AREA AND PREDICTION MODEL

2.1 Study Area and Spatial Input Data Set

The study area is located on Futtu in Chiba prefecture, Japan. In the rainy season between July and August in 1988, the local-downpour with continuous rainfall had caused the slope failures and landslides in this study area. Through the field investigation and the aerial photographs, those occurrences were precisely plotted on the topographical map as the training data sets for constructing the prediction model.

The quantitative prediction model constructed the relationship between those past occurrences and the following nine "causal factors": (1) Soil, (2) Surface geology, (3) Vegetation, (4) Land cover, (5) Vegetation index, (6) Slope, (7) Aspect, (8) Elevation, and (9) Drainage. Each map consists of 100×50 pixels (3.0 Km ×1.5 Km, 30m/pixels corresponding to the ground resolution of the Landsat TM data). The latter four factors were produced based on the Digital Elevation Model (DEM). The experts in each research field have made the Soil-, Surface geology- and the vegetation-map. The land cover map is made through the maximum likelihood classification for the Landsat TM data. The vegetation-index map is also produced by calculating the Normalized Vegetation Index (NVI) given by

$$NVI = (B7 - B5) / (B7 + B5) \tag{1}$$

where $B5$ and $B7$ are the digital numbers in each pixel corresponding to TM-Band 5 and TM-Band 7, respectively.

2.2 Quantitative Prediction Model

Figure 1 shows the inverse analysis concept of the trigger factors expanding the previous quantitative prediction model. Chung and Fabbri (1999) have adopted the formulas for geologic hazard zonation as a part of "favorability function" approaches, and the various procedures have been applied to the landslide prediction. To promote those prediction models as well as optimizing prediction, the practical analytical procedures have been presented as follows; i) Comparative strategy of the prediction models (Kojima et al., 1998, 1999), ii) Analysis of the landslide types (Kojima et al., 2000), iii) Testing on the time-robustness in prediction (Kojima et al., 2001), and iv) Sensitivity analysis of the prediction models with respect to the causal

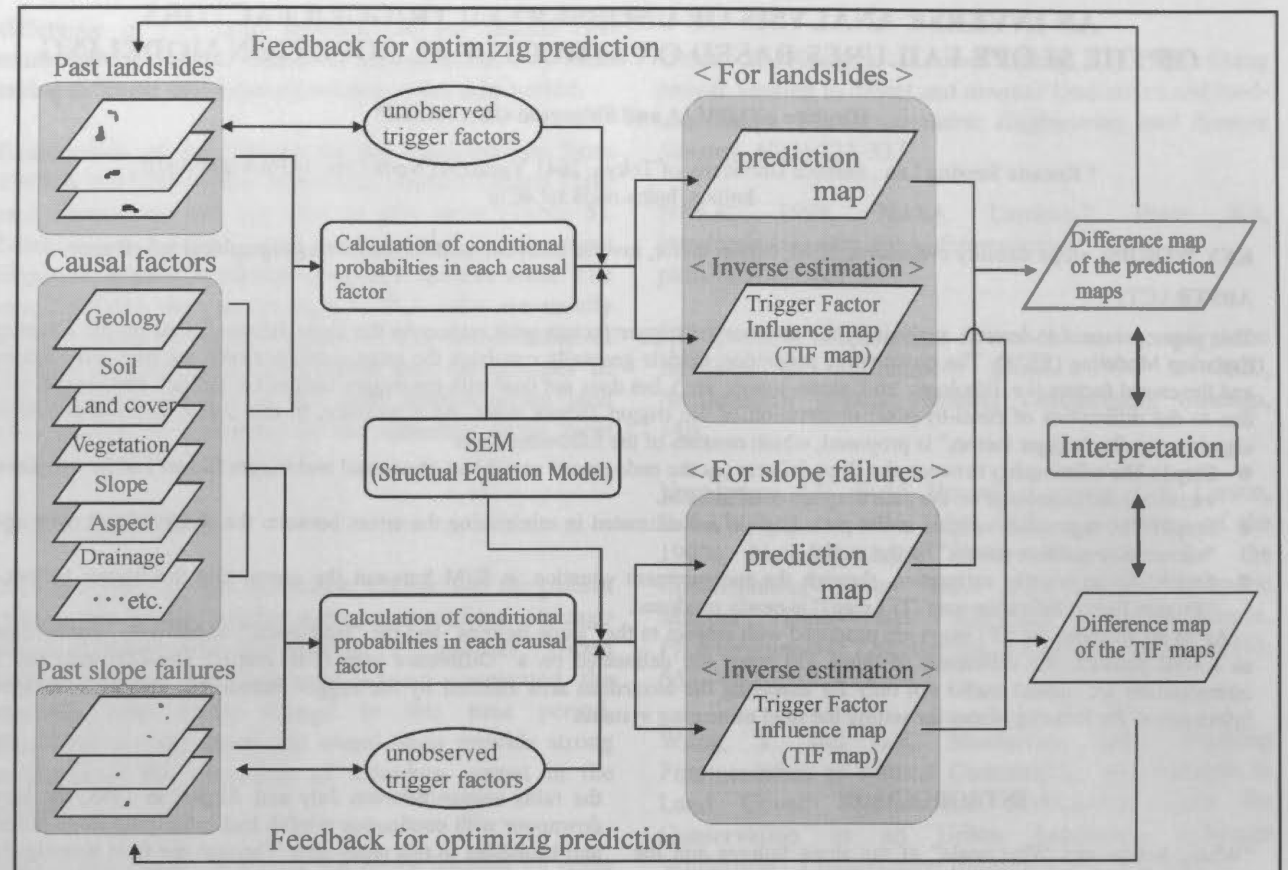

Feedback for optimizig prediction
Past landslides
unobserved trigger factors
< For landslides >
prediction map
Calculation of conditional probabilties in each causal factor
Causal factors
< Inverse estimation >
Difference map of the prediction maps
Geology
Soil
Land cover
Vegetation
Slope
Aspect
Drainage
· · · etc.
Trigger Factor Influence map (TIF map)
SEM (Structual Equation Model)
Interpretation
< For slope failures >
prediction map
Calculation of conditional probabilties in each causal factor
Difference map of the TIF maps
Past slope failures
< Inverse estimation >
Trigger Factor Influence map (TIF map)
unobserved trigger factors
Feedback for optimizig prediction

Figure 1. Proposed inverse-analysis algorithm of the "trigger factor" with respect to the slope failures and landslides.

factors (Chung et al., 2002).

Those analytical procedures are crucial components in the prediction models, and are indeed useful not only for the experts working on the landslides, but also for the end-users of the prediction models. However, the conventional prediction models generally construct the relationship between the past slope failures and the causal factors, but could not apply the "trigger factors," because of the difficulties in observing the trigger factors "pixel-by-pixel". As a measure in this study, an inverse analysis algorithm on the "trigger factors" is presented, based on the Structural Equation Modeling (termed SEM) that was originally proposed by Joreskog and Lawley (1968). SEM is also well known as the "analysis of covariance structures," in a word, SEM could be regarded as a model integrated the regression analysis jointly with the exploratory factor analysis. The detail of SEM itself is available for the references (Joreskog et al.1968).

3. INVERSE ANALYSIS OF TRIGGER FACTOR

Figure 1 shows the proposed inverse-analysis algorithm of the "trigger factor," which consists of the following steps:

3.1 Conditional Probabilities as the Input Data

To construct a probability model for slope failure hazard, consider the following proposition:

> F_p : " a pixel p will be affected by a future slope failure of a given type D."

The conditional probabilities in each causal factor given by

$$\boldsymbol{Prob(F_p|C_{ij}) = T_{ij} / N_{ij}} \qquad (2)$$

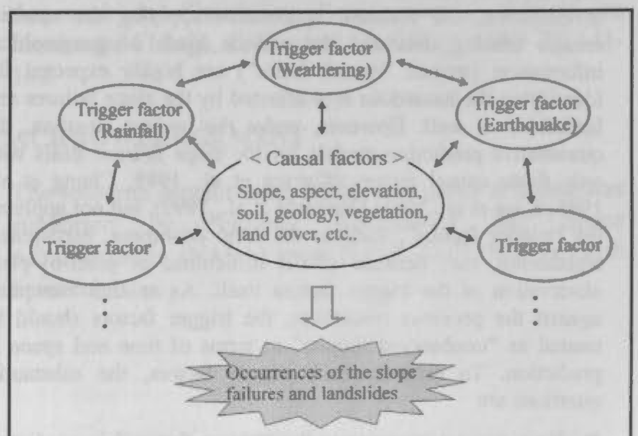

Trigger factor (Weathering)
Trigger factor (Rainfall)
Trigger factor (Earthquake)
< Causal factors >
Slope, aspect, elevation, soil, geology, vegetation, land cover, etc..
Trigger factor
Trigger factor
Occurrences of the slope failures and landslides

Figure 2. Causal and Trigger factors with respect to the slope failures and landslides.

where C_{ij} is the ith category of the jth causal factor; N_{ij} is the number of pixels of C_{ij}; and T_{ij} is number of pixels of the past slope failures or landslides that had occurred in the area corresponding to C_{ij}. $\boldsymbol{Prob(F_p|C_{ij})}$ are used as the input data for the SEM-based analysis.

3.2 Path Diagram

To make a prediction model, the relationship between the slope failures (as the endogenous variables), the causal and trigger factors (as the exogenous variables) should be delineated on the path diagram used in the SEM. **Figure 2** shows the intricate relation between the causal and trigger factors with respect to the

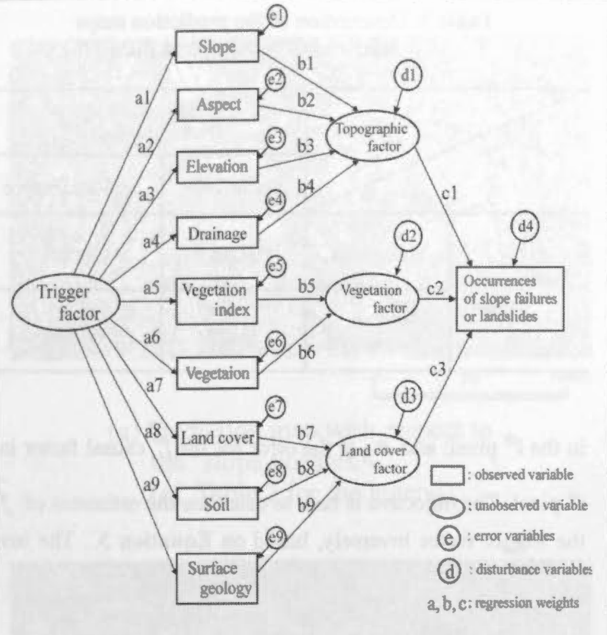


Figure 3. Path diagram for the SEM-based analysis.

slope failures and landslides. As a counterpart of **Figure 2**, let us consider the path diagram as shown in **Figure 3** that is called a recursive model. $\boldsymbol{Prob(F_p \mid C_{ij})}$ of **Equation 2** are the input data as the exogenous variables, while the pixels corresponding to occurrences and non-occurrences of the slope failures as well as landslides are assigned to the value "1" or "0", respectively, that are used as the endogenous variables.

3.3 Hypothesis Testing

Not knowing the trigger factors, the program is how to estimate the path weights of $\{a_1, \ldots, a_n, b_1, \ldots, b_n, c_1, \ldots, c_n\}$ in **Figure 3**. Through the estimation procedure in SEM, those are estimated by minimizing the errors between the observed variance-covariance matrix and the reemerged one. Among various estimation procedures, i.e., maximum likelihood estimation, asymptotically distribution-free estimation, generalized least squares estimation, 'scale free' least squares estimation, unweighted least squares estimation, etc., Maximum likelihood estimation procedure is selected in this study, which is generally reported as a better estimator for the large population than the others.

To make clear the significance introducing the "unobserved trigger factors" in **Figure 3**, let us consider the following path models for the slope failures and landslides, respectively:

- Model A: using both causal and trigger factors for the "slope failures," as shown in **Figure 3**.
- Model B: only using causal factors for the "slope failures."
- Model C: using both causal and trigger factors for the "landslides," as shown in **Figure 3**.
- Model D: only using causal factors for the "landslides."

As the hypothesis testing for those models, the Chi-square value, the Goodness of Fit Index (GFI), the Adjusted Goodness of Fit Index (AGFI), the Akaike Information Criterion (AIC), and the Root Mean Square Error Approximation (RMSEA) are applied as the statistical measures of fit. **Table 1** shows the results of the hypothesis test and the fit measures.

The Chi-square value for Model A is higher than that for Model B, also the probability level in Model A indicates more than 0.05, but not in Model B, which means that Model A could not be rejected under the significance level "0.05", while there is no reliability on the identification of Model B. As the other measures of fit, the GIF and the AGFI need to be more than 0.9, conversely, the RMSE should be less than 0.08 for the model selection. Furthermore, among those models, the only model with the lowest AIC should be selected.

Table 1. Hypothesis test results and the fit measures.

Measures of fit		For the slope failures		For the landslides	
		Model A	Model B	Model C	Model D
Chi-square test	Chi-square	18.3	358.5	58.8	519.1
	Degree of freedom	28	37	28	37
	Probability level	0.316	0.000	0.227	0.000
GIF		0.993	0.786	0.975	0.801
AGIF		0.986	0.778	0.951	0.705
AIC		24.5	400.5	65.0	523.6
RMSEA		0.025	0.147	0.04	0.167

Notes)
Model A: using both causal and trigger factors for "slope failures."
Model B: using only causal factors for "slope failures."
Model C: using both causal and trigger factors for "landslides."
Model D: using only causal factors for the "landslides."

Based on the above experiments, for the slope failures in the study area, we would say at least that Model A is better than Model B on the statistical grounds. Furthermore, for the landslides, **Table 1** indicates that Model C would be better fit against Model D. Those results imply that the path diagram introducing the "trigger factor" shown in **Figure 3** might be meaningful for constructing the prediction model.

3.4 Path Parameter Estimation

Table 2 shows the standardized regression weights of estimated for the selected Model A and Model C side by side. The difference of those parameter weights between Model A and Model C is obvious, which corroborates that the trigger factors with respect to the slope failures and the landslides may be different.

Focusing on the path-relation between the trigger factor and the causal factors, in Model C for the landslides, the highest weight "0.869" comes from the path between the trigger factor and the land-cover. While in Model A for the slope failures, it is interesting to note that the negative weights come from the path between the elevation, drainage and the land-cover. Comparing such differences of the path weights, we can evolve the factor analysis from various points of view on the slope failures and the landslides.

3.5 Prediction Map

Based on the path model of **Figure 3**, the prediction maps could be produced. **Plate 1(a)** and **Plate 1(b)** show the prediction maps with respect to the slope failures and the landslides, respectively. **Table 3** shows the description of those prediction maps, which are made by the mini-max discriminate method that classifies the pixels into two groups as "occurrence and non-occurrence (Kasa, Kojima, et al., 1991)." The classified results of these kinds are indeed useful for supporting the decision-making of the landslide

Table 2. Results of the standardized parameter estimates.

Path discription			Parameters in Figure 3	Model A for slope failures	Model C for landslides
Trigger factor	→	Slope	a1	0.232	0.705
	→	Aspect	a2	0.011	0.007
	→	Elevation	a3	-0.030	0.159
	→	Drainage	a4	-0.052	0.447
	→	Vegetaion Index	a5	0.001	0.010
	→	Vegetation	a6	0.247	0.459
	→	Land cover	a6	-0.203	0.869
	→	Soil	a7	0.294	0.273
	→	Surface geology	a8	0.260	-0.075
Slope	→	Topographic factor	b1	0.411	-0.004
Aspect	→		b2	0.748	0.703
Elevation	→		b3	0.451	0.163
Drainage	→		b4	0.271	0.680
Vegetaion Index	→	Vegetaion factor	b5	0.662	0.637
Vegetation	→		b6	0.749	0.768
Land cover	→	Land cover factor	b7	-0.034	0.953
Soil	→		b8	0.856	0.122
Surface geology	→		b9	0.449	-0.094
Topographic factor	→	Occurrences of slope failures or landslides	c1	0.105	0.100
Vegetaion factor	→		c2	0.065	0.093
Land cover factor	→		c3	0.051	0.061

Table 3. Description of the prediction maps shown in Plate 1(a) and Plate 1(b).

		Training data sets (Past situation)	
		Occurrence	Non-occurrence
Prediction map	Hazardous	Red	Green
	Non-hazardous	Blue	White

prevention plans, against the general prediction map ranked with several hazardous levels.

In **Plate 1(a)** and **Plate 1(b)**, it is obvious that the prediction patterns are fairly different, which means the causal and the trigger factors between the slope failures and the landslides might be different. In the next section, as further investigation, let us discuss about the inverse estimation of the trigger factor for the slope failures and landslides as well.

3.6 Inverse Estimation of Trigger Factor

Note that the path components connecting "unobserved variables to each other" and "observed variables to unobserved variables" are often called the "structural equation" and "measurement equation," respectively. In this study, through the measurement equation, the influence of the trigger factor are inversely estimated pixel-by-pixel, and those are delineated on a "Trigger Factor Influence map (termed TFI map)." In the path diagram shown in **Figure 3**, the measurement equation between the trigger factors (as unobserved variables) and the causal factors (as observed variables) is given by

$$z_{ji} = a_j f_i + e_{ji} \tag{3}$$

where z_{ji} is the conditional probability of the jth causal factor in the ith pixel as shown in **Equation 2**; a_j is the path parameter in **Table 2** linked the jth causal factor with the trigger factor; f_i is the real value of the degree on the trigger factor influence in the ith pixel; and e_{ji} is the error for the jth causal factor in the ith pixel. The objective is how to calculate the estimates of $\hat{f}_i$ of the trigger factor inversely, based on **Equation 3**. The inverse function is given by

$$\hat{f}_i = \sum_{j=1}^{p} b_j z_{ji} \tag{4}$$

The parameters of $\{b_1, \ldots, b_n\}$ are determined by minimizing the following least square error:

$$Q = \sum_{i=1}^{n} (f_i - \hat{f}_i)^2 = \sum_{i=1}^{n} (f_i - \sum_{j=1}^{p} b_j z_{ji})^2 \rightarrow min$$

$$\frac{\partial Q}{\partial b_{j'}} = -2 \sum_{i=1}^{n} z_{j'i}(f_i - \sum_{j=1}^{p} b_j z_{ji}) = 0 \tag{5}$$

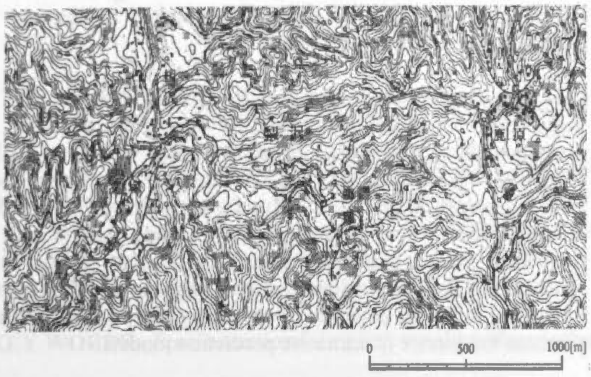
Note that the average and the variance for z_{ji} and f_i are standardized to "0" and "1", respectively. So, the error parameter could be ignored in **Equation 5**. Hence, $\{b_1, \ldots, b_n\}$ are simply given by

$$b_j = \sum_{j'=1}^{p} a_{j'} r^{jj'} \tag{6}$$

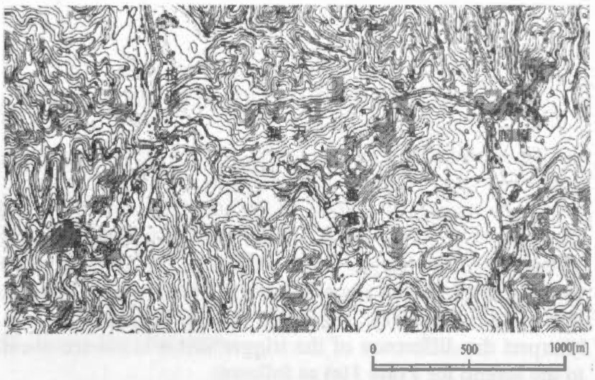
where $r^{jj'}$ is the element (j, j') of the inverse matrix for the correlation matrix. Through **Equation 4**, $\hat{f}_i$ is calculated pixel-by-pixel, and is delineated on the Trigger Factor Influence map (TIF map). **Plate 1(c)** and **Plate 1(d)** show the TIF maps with respect to the slope failures and the landslides, respectively. From those TIF maps jointly with the prediction maps, the following points could be indicated:

- At a glance, it is found that the estimated patterns between the TIF maps with respect to the slope failures and the landslides are clearly different.
- In **Plate 1(c)** on the slope failures, the steeper the slope is, the higher the trigger factor influence is for the most part. On the other hand, in **Plate 1(d)** on the landslides, the gentler the slope is, the higher the trigger factor influence is. Those results corroborate that the trigger factors could be estimated "pixel-by-pixel" through the path model shown in **Figure 3**.
- However, in **Plate 1(c)** on the slope failures, we can also see the areas where the trigger factor influence is high in spite of relatively gentle slope. Especially, for those areas, the detail field investigation should be carried out.

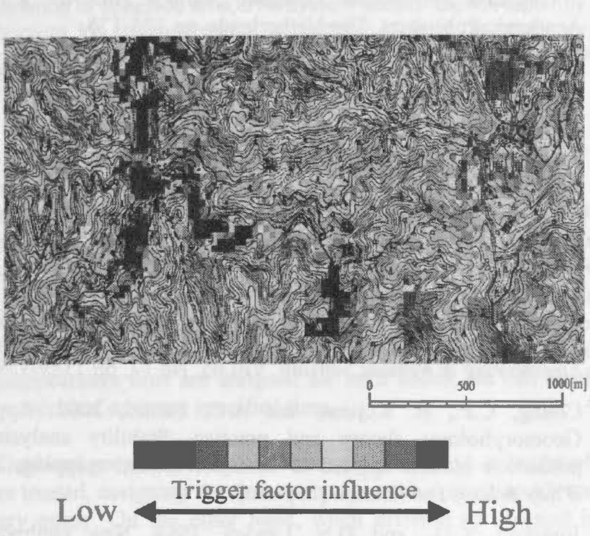
For the practical utilization of the TIF maps, let us consider the



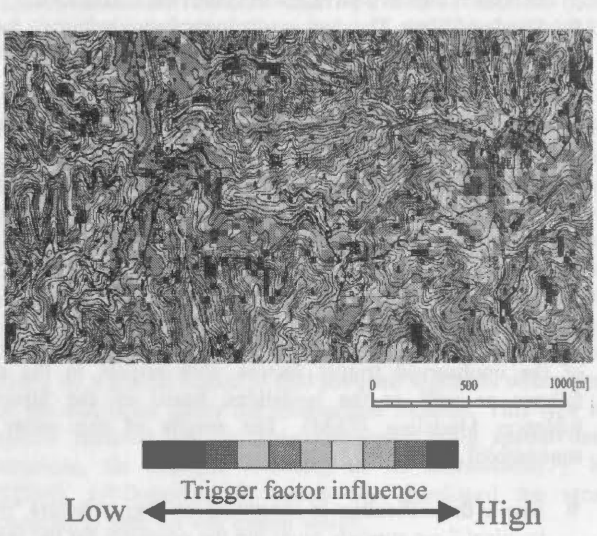
(a) Prediction map with respect to the "slope failures." (cf. Table 3 for the legend)



(b) Prediction map with respect to the "landslides." (cf. Table 3 for the legend)

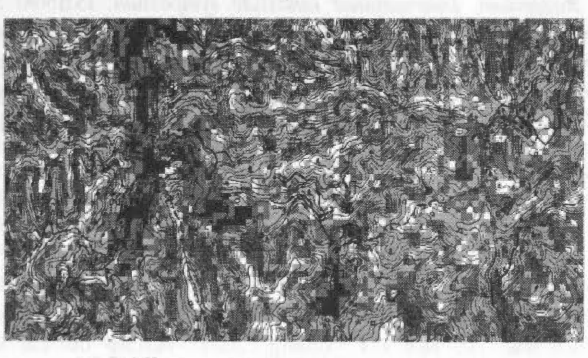

0
500
1000[m]
Trigger factor influence
Low
High

(c)Trigger Factor Influence map (TFI map) with respect to the "slope failures."


0
500
1000[m]
Trigger factor influence
Low
High

(d)Trigger Factor Influence map (TFI map) with respect to the "landslides."

Visualization of differences



(e) Difference map (DIF map) between the trigger factor influence maps with respect to the slope failures and the landslides, respectively.

< Legend of Plate 1(e) >

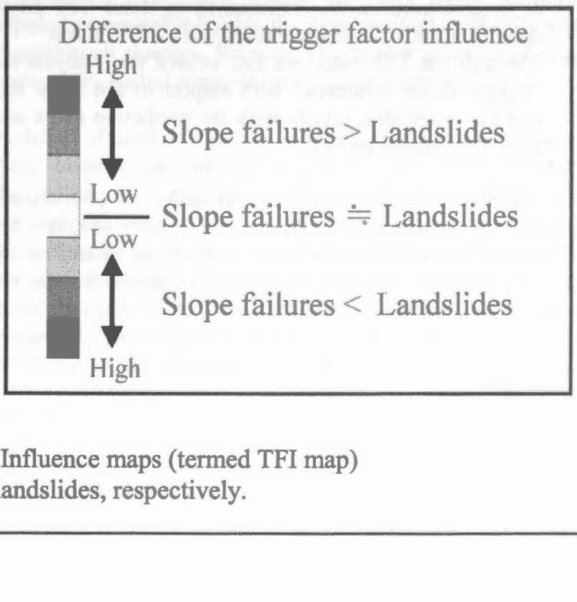

Difference of the trigger factor influence
High
Slope failures > Landslides
Low
Slope failures ≒ Landslides
Low
Slope failures < Landslides
High

Plate 1 Prediction maps and the Trigger Factor Influence maps (termed TFI map) with respect to the slope failures and the landslides, respectively.

interpretation for differences of the TIF maps with respect to the slope failures and the landslides, respectively.

4. INTERPRETATION OF THE TRIGGER FACTOR INFLUENCE MAP (TIF MAP)

As the conventional interpretation of the prediction maps, the difference map (termed DIF map) is produced between two prediction maps with respect to the slope failures and landslides (Kojima, et al. 1998, 2000, 2001). Similarly, to make clear the difference of the trigger factor influence, the DIF map between two TIF maps for the slope failures and landslides is produced as shown in **Plate 1(e)**. Attention should be paid that we can interpret the difference of the trigger factor influence according to the legend for **Plate 1(e)** as follows:

- **Shade of red:** The degree of trigger factor influence for the slope failures is higher than that of the Landslides.
- **White:** The degree of trigger factor influence for the slope failures is almost equivalent to that of the Landslides.
- **Shade of blue:** The degree of trigger factor influence for the slope failures is lower than that of the Landslides.

Such "heuristic information" might be useful not only for assessing the hazardous area affected by the trigger factors in terms of the types of the slope failures and landslides, but also for improving the cost-effectiveness in locating the places for setting the field measuring systems, i.e. the tensiometer, the rain gage, etc.

5. CONCLUDING REMARKS

In this contribution, we have discussed about an inverse analysis of the unobserved trigger factors with respect to the slope failures as well as the landslides, based on the Structural Equation Modeling (SEM). The results of this study are summarized as follows:

- Due to the difficulties in observing the trigger factors "pixel-by-pixel," we strongly point out the necessity for the inverse estimation of the "unobserved trigger factors" itself. As a measure, through the measurement equation between the causal factors (as observed variables) and trigger factors (as unobserved variables), a "Trigger Factor Influence map (termed TFI map)" is newly produced.
- As an application of the proposed inverse-analysis algorithm, the TFI maps are produced with respect to the slope failures and the landslides, respectively. Furthermore, as a final product, the differences of those TFI maps are delineated on a "Difference map (termed DIF map)."
- Through the DIF map, we can evolve the analysis on the "trigger factor influence" with respect to the slope failures and the landslides, jointly with the prediction maps and the expert's opinions as well.

As for the subsequent subjects, in order to corroborate the practicality of the inverse-analysis algorithm **(Figure 1)**, the additional investigation for other study areas should be carried out. As occasion demands of the investigators, we can readily add the training data sets of other types of slope failures and landslides in analysis. In this point of view, the analytical procedure shown in **Figure 1** is expected to contribute to the landslide hazard assessment as one of the standards.

Furthermore, as for the structure of the path diagram in **Figure 3**, a single "exogenous variable" is considered as the "trigger-factor." However, in practice, either slope failures or landslides are caused by various trigger factors as shown in **Figure 2**. So, the modified path models adding the several exogenous variables as the trigger factors should be investigated to improve the identification of the models itself.

It is no exaggeration to say that the precise estimation of the trigger factor itself is impossible. As one of the measures, the inverse-analysis algorithm presented in this study, as well as "heuristic information" on the DIF map of the trigger factor influence maps, might be effective for identifying the hazardous area affected by the different types of slope failures and the landslides. Such a systematic analysis procedure, as against the limitation of the conventional research approaches for prediction modeling, might be essential to optimize prediction as well as to promote the former quantitative prediction models.

REFERENCES

Carrara, A., M. Cardinali, F. Guzzetti, and P. Reichenbach, 1995. GIS Technology in mapping landslide hazard, *Geographical Information Systems in Assessing Natural Hazards*, Kluwer Academic Publishers, The Netherlands, pp.135-175.

Carrara, A., 1998. Current limitations in modeling landslide hazard, *Proceedings of The International Association for Mathematical Geology*, No.4, pp.195-203.

Chung, C.F., A.G. Fabbri, and C.J. van Westen, 1995. Multivariate regression analysis for landslide hazard zonation, *Geographical Information Systems in Assessing Natural Hazards*, Kluwer Academic Publishers, The Netherlands, pp.107-133.

Chung, C.F., and A.G. Fabbri, 1999. Probabilistic prediction models for landslide hazard mapping, *Photogrammetric Engineering & Remote Sensing*, Vol.65, No.12, pp.1389-1399.

Chung, C.F., H. Kojima, and A.G. Fabbri, 2002. Applied Geomorphology: theory and practice, Stability analysis of prediction models applied to landslide hazard mapping, *John Wiley & Sons Publication (in press).*

Joreskog, K.G., and D.N. Lawley, 1968. New methods in maximum likelihood factor analysis, *British J. Math. Statist. Psycol.*, 21, 85-96.

Kasa, H., M. Kurodai, H. Kojima, and S. Obayashi, 1991. Study on the landslide prediction model using satellite data and geographical information, Landslides, *Bell(ed.), Balkerma, Rotterdam, International landslide symposium*, ISBN90 5410 032 X, pp.983-988.

Kojima, H., C.F. Chung, S. Obayashi, and A.G. Fabbri, 1998. Comparison of strategies in prediction modeling of landslide hazard zonation, *Proceedings of The International Association for Mathematical Geology*, No.4, pp.195-203.

Kojima, H., C.F. Chung, and C.J. van Westin, 2000. Strategy on the landslide type analysis based on the expert knowledge and the quantitative prediction model, *International Archives of Photogrammetry & Remote Sensing*, Vol.33, Part-B7, pp.701-708.

Kojima, H., and C.F. Chung, 2001. Testing on the time-robustness of a landslide prediction model, *Proceedings of The International Association for Mathematical Geology*, Vol.7, pp.41-46.

Obayashi, S., H. Kojima, and C.F. Chung, 1999. Comparison of the accuracy of the slope stability evaluation models and its practical application, *Journal of the Construction Management and Engineering, Japan Society of Civil Engineers (in Japanese)*, No.630, 6-44, pp.77-89.

A DATA STRUCTURE FOR SPATIO-TEMPORAL INFORMATION MANAGEMENT

Y. Ohsawa *, W. Guo, M. Sakurai [a]

Department of Infomation and Computer Sciences, Faculty of Engineering
Saitama University, Shimo-okubo, 255, Saitama, 338-8570, Japan
ohsawa@mm.ics.saitama-u.ac.jp

KEY WORDS: Spatio-temporal, Implicit topology description, STIMS, Free GIS, GBD tree

ABSTRACT:

This paper describes about spatio-temporal geographic information system, named STIMS(Spatio-Temporal Information Management System). STIMS handle temporal information assigned each geographic object and can execute several kinds of spatial operations on the world of specified time. The system adopts implicit topology data model to handle such temporal data. By this data model, the data structure for spatio-temporal geographic objects becomes very simple. This paper describes about the data structure and how to handle spatio-temporal data on the system. STIMS is now released as free software on the WEB.

1. INTRODUCTION

On geographic information system, to handle transition of real world, time slice or version management of whole map have been adopted. On this method, the change of each geographic entity cannot be described. The change is known to compare two temporally neighbouring time slice maps. On the other hand, if time stamps which shows the generation and disappearance time are assigned for each entity, we can catch the world of arbitrary specified time.

If isolated geographic object, for example building or landmark are treated, assigning the time stamps to each entity is achieved very easily. On the other hand, when division of a parcel is concerned, this time management is not trivial. Because the topology of the neighbour of the parcel has been changed by this division. On conventional geographic information systems, region have been managed as network composed by nodes and links. Then, if a parcel is divided into two, this affects to the organization of the network.

On usual geographic information systems, topological information has been described explicitly using relational tables or pointers[1]. This method was very natural and optimum choice when GIS came into existence (1970s). Because the computer processing speed at that time was very slow and the amount of main memory was limited. Then, to describe topological information explicitly was mandatory for obtaining good response.

However, during later two or three decades, computing speed has increased more than 1,000 times. By this computing power progress, new methodology which restore the topological information by calculation when it is necessary becomes actual[2],[3]. We call this methodology as implicit topology description method.

Not only for region data, but also for network data, like road and river, topology description is very important. When we want to know the shortest path between two specified points (start and destination) on a road network, the path can be found by successive following the connected road segments under the control of A* algorithm[4] or Dijkstra algorithm[5]. When the topological information of road network is given explicitly, the set of connected road segments at an intersection can be known very easily by following pointers or inquiring relational tables. On the other hand, on implicit topology model, the topological information is restored by spatial retrieval at the each intersection. The connecting road segments at an intersection have the same end point, then these road segments can be found by a spatial search to find the road segments whose one of the end point meets the intersection location. This type of spatial retrieval can be done efficiently using spatial data structure, for example R-tree[6] or its descendants[7]. In STIMS, GBD-tree[8],[9], which was developed the same purpose, is used for this spatial search.

By possessing topological information implicitly, temporal information management becomes easy[10]. For example, suppose the situation to manage the transition of administrative borderline for 50 years time span. During the time span, cities may have been merged and split, and the names of the cities may have been changed. When explicit topological expression is used for managing these changes, the structure becomes very complicated like spaghetti. On the other hand, on implicit topology description model, the data structure stays very simple despite such changes. Because, city border at specified time is restored by spatial retrieval using spatio-temporal key.

A defect of implicit topology model is in the longer calculation time, because the topology is always restored by calculation. To solve this problem, STIMS adopts topology cache method[11]. Once topology is restored, the restored topology is kept in cache for the next use. By this method, calculation time reduced drastically. When a part of data is changed, the cache data which is affected by the change is discarded, then the topology is re-calculated. When the time view is changed, the cache which affect the time view change is also discarded.

STIMS is implemented by C++ language on Windows PC and it is released as free software.

In section 2, data model used in STIMS is described. In section

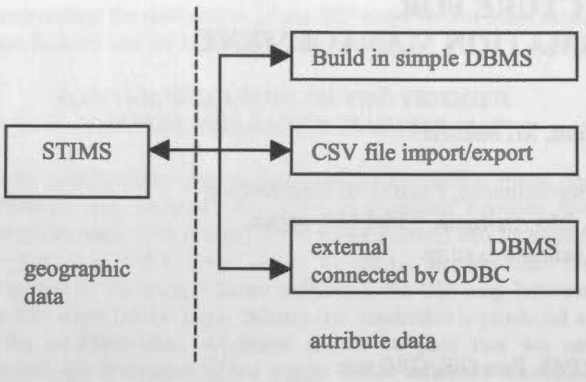


Fig.1 STIMS and DBMS for attribute data

3, several spatial operations to restore topological information is described. In section 4, temporal data handling method and its data structure is described. GIS refers several dataset, including spatial entity file, attribute data file, several definitions and legends. Such kinds of data formats are described in section 5. Section 6 concludes the paper.

2. ORGANIZATION OF STIMS

Entities which managed in GIS are composed by geometric data and their attribute data. Mainly, the shape and the position of geographic entities are managed in STIMS. To handle attribute data, STIMS furnishes three methods as shown in Fig.1. One is built-in very simple DBMS, and the second one is CSV interface. The last one is ODBC interface; this one is the most useful and common interface.

To combine between geometric object and its attribute, unique `key' is assigned for both data. Usually, a unique numerical identifier (ID) is assigned for corresponding object on both systems. STIMS also assigns such ID for geometrically isolated entities and line entities. In conventional GIS, to manage regional data, TIGER or DIME structure is adopted. On these methods, the information denoting left and right region IDs is assigned for each borderline. While region is stable, this method works very well. However, when occasionally occurring changes have to be concerned, this method makes data management very complicated. Each line must have several IDs describing the history of changes.

To simplify the data management, STIMS separates attribute management from region shape. A point data located inside of a region is assigned for a region as a spatio-temporal key. The point has the location (X and Y axis position) and time data denoting the life span of the object. The shape of a region, in other word the topology of the borderline of the region, is restored when a region is specified for visualization or spatial analysis. While the restoration process, time information assigned for each line segment is referred. Using only valid line segments at the specified time, the region shape is restored.

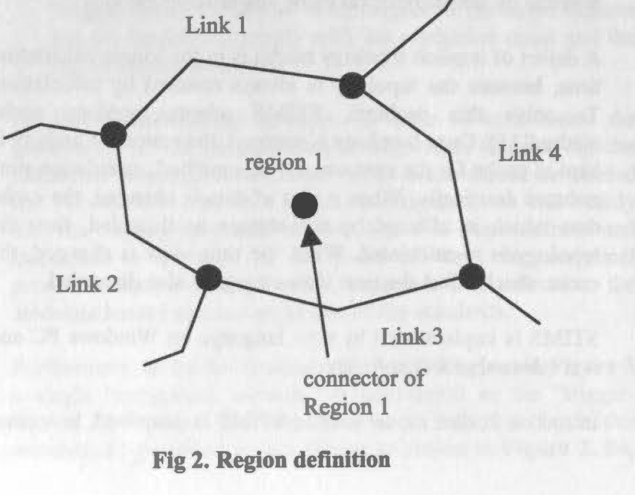


Fig 2. Region definition

Figure 3 shows a part of a network, such as road network. In this figure, three roads, L1, L2 and L3 are connected each other at node N. In the network, roads correspond to links and the intersections correspond to nodes. The explicit topology model, which has been used usual GIS, deals with the topology of the network explicitly using pointers or relation tables.

The implicit topology model, on the other hand, describe topological information by definition and method functions. This method restores the topological relation through spatial operations. Namely, other links connected to the specified link are found by a spatial retrieval specifying the end point of the link. The spatial retrieval returns the links of which one endpoint meets at the same position with the specified end point. In this paper, this operation is called as 'node expansion'.

To perform such an operation at high speed, a special data structure, so called spatial data structure, is necessary. Several kinds of spatial structures, including R-tree[6], R*-tree[7] and GBD-tree[8], have been proposed. In this paper, GBD-tree, which is developed one of the authors, is used to manage spatial data. The algorithms described in this paper can be widely applicable to other types of spatial data structures.

In the implicit topology model, the attributes and the geometrical shape of the entities are managed separately. When correspondence among them is necessary, they are combined by spatial retrieval. Figure 2 shows a line object (e.g. road), and the object has attributes (e.g. name of the road, width of the road, number of traffic lanes). Usually, when a GIS displays the entities on CRT, all attributes data are not always necessary. Thus, it is sufficient that the attributes can be retrieved when they are requested. Then, if the shape and the attributes are separated, the time required to read the entities from a hard disk can be shortened. For the reason mentioned above, it is preferable that the shape and the attributes are managed separately even on explicit topology based GIS.

The shape and the attributes are related to each other by their position. Lines and regions have extent in two-dimensional space. Then, to determine their positions, a unique point, or representative point of the object is necessary. We determine the point as the center of the minimum-bounding rectangle (MBR) of the object. The attributes are located at the

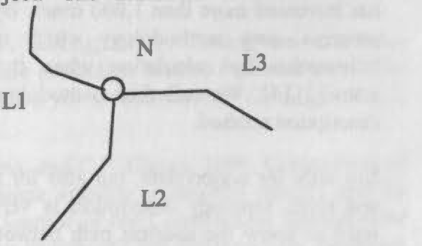


Fig.3 Network node and links

representative point. Hereafter, we call the representative point the 'connecter', in the sense that it connects the shape and the attributes.

To generate the relationship between the shape and the attributes, two inverse directional operations are necessary: (1) find the attributes (connector) from the shape, and (2) find the shape from the connector. The details of these operations are described in section 3.

Regions are expressed by a set of borderlines, as shown in Figure 2. Each closed area divided by the borderlines is a region. In the implicit topology model, all borderlines are also not mutually explicitly related. Hence, the shape of a region must be also restored by using node expansion.

The attributes of a region are related to the geometrical shape by the connector. In the case of region, the connector is placed somewhere inside the region. Because the shape of region is varied, it is not easy to determine a rule for placing the connector. For this reason, the position of the connector is restricted only to somewhere inside the region, as shown in Figure 3. To determine the attributes of a region, the following steps of operations are necessary: (1) restore the shape of the region, (2) find the position of the connector by searching inside the region, then (3) read the attributes related to the connector.

Summarizing the operations used in the implicit topology GIS, the followings basic generations are mainly necessary:

(1) node expansion
(2) generation of correspondence between the shape and the connector of line objects
(3) generation of correspondence between the shape and the connector of region objects

By combining these basic operations, more complicated geographical operations, such as the shortest path finding, are realized.

3. OPERATIONS TO RESTORE TOPOLOGY

3.1 Spatial operation for restoring topology

Figure 3 showed a simple example of topology. In explicit topology description, relation tables or pointers describe the connective relationships among L1, L2, L3 and N, allowing us to easily recognize the links that are connected to each other by referring to them. On the other hand, a GIS based on implicit topology description must restore the relation by spatial retrieval.

In STIMS, GBD-tree is used for this purpose. The GBD-tree is a multi-dimensional data indexing structure. Figure 4 shows the structure. As shown in this figure, each node has M slots. Each slot in non-leaf nodes has a pointer to the child node, its minimum-bounding rectangle (MBR) and the center point of the MBR for indexing. The MBR is the tightest rectangle bounding all entities contained in offspring nodes. A slot in the leaf nodes has pointer to the entity, MBR of the entity, and the layer of the entity. Spatial retrieval can be done efficiently by referring to the MBR.

As mentioned in Section 2, node expansion is a basic operation for restoring topology. This operation can be performed by searching the links of which an endpoint meets the position of the node. Following shows the procedure of searching the links connected to node N on GBD-tree.

(Step.1) Let T be the root node.

(Step.2) If T is not a leaf node, each offspring node E is checked whether the MBR of E includes the location of node N. For all including nodes, search the links connected to node N on the sub tree whose root node is E.

(Step.3) If T is a leaf node, all entities included the node are checked whether an end point of it meets the position of node N. If so, the entity is connected with node N.

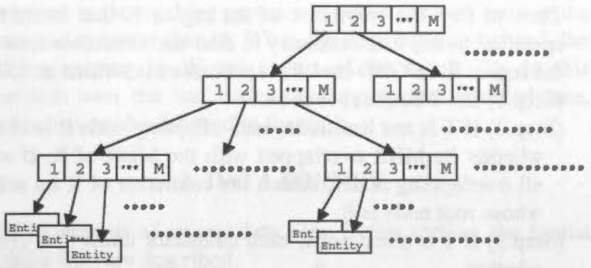


Fig.4 Hierarchical structure of GBD-

3.2 Searching connector of a specified line

The connector of a line object is found by searching the connector object corresponding with the center of the MBR of line object on GBD-tree, because the connector is located at the center of the MBR. In the following sentence, we describe the procedure to find the connector of link L on GBD-tree, when link L is specified.

(Step.1) Let T be the root node.

(Step.2) If T is not leaf node, each offspring node E is checked whether the MBR of E includes the center of the MBR of L. If so, for all including nodes, search the connector of L on sub-tree whose root node is E.

(Step.3) If T is a leaf node, check all connectors under the node whether its location accords to the center of the MBR of L.

3.3 Searching connector of a specified region

To find the connector of the region that includes the specified point, it is necessary to restore the region, and to find the connector located in the region.

In the implicit topology model, a region consists of a boundary and a connector (the representative point). The boundary consists of a collection of line segments (links). The connector is placed inside the region, and is used to connect the attribute information of the region.

Region restoration is a procedure to restore the shape of a region. This procedure is necessary in several operations including to restore the shape to display and to know the attribute of the region. This procedure is written as follows.

(Step 1) Find one of the boundary line segments of the specified region. This can be done by retrieving the rightmost boundary line segment on the horizontal line stretched from the specified point P, as shown in Figure 5. Let the found line segment S. This operation can be executed efficiently on GBD-tree.

(Step 2) Follow the line segments forming the specified region counter-clockwise from S to return back to S. At the end of S, the next segment can be found as follows. In Figure 6, let link L be a boundary line segment of which either end is N. Expand the node to find all line segment meeting at N. Then, determine the line segment located counter-clockwise neighboring to line L. Repeat the same operation at the end of the opposite side end of the segment until return back to link L. After the operation, all line segments consisting of the region are found.

Figure 5 summarizes the operations.

Then to find the connector of the region R that includes the specified point, it is necessary to find the connector located in the region R on GBD-tree. This procedure is written as follows.

(Step.1) Let T be the root node.

(Step.2) If T is not leaf node, each offspring node E is checked whether its MBR overlapped with the MBR of R. If so, for all overlapping nodes, search the connector of R on sub-tree whose root node is E.

(Step.3) If T is a leaf node, each connector under T is checked whether it is located in R.

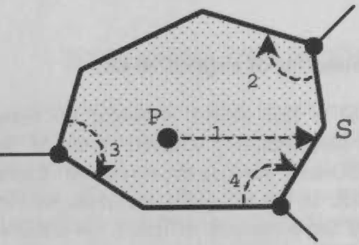


Fig.5 Region restoration

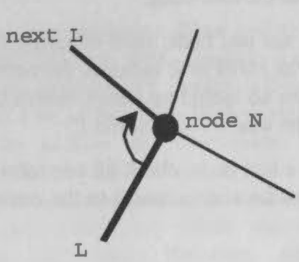


Fig.6 Node expansion

3.4 Searching line entity corresponding to a specified connector

When some attributes are specified to display the line segments meeting the conditions, the connector that satisfies the condition are retrieved. The shape of the line is searched by spatial retrieval in the way the center of the line locally corresponds to the connector.

This retrieval can be performed on GBD-tree as follows.

(Step.1) Let T be the root node.

(Step.2) If T is not leaf node, each offspring node E is checked whether its MBR include the location of the connector. If so, for all including nodes, search the line on sub tree whose root node is E.

(Step.3) If T is a leaf, each entity under the node is checked whether the center of its MBR according to the location of the specified connector. If so, it is the line object according to the specified connector.

3.5. Searching region corresponding to a specified connector

The connector of a region is located in the region, so when a connector is specified to restore the region, the procedure can be done by the same procedure described in 3.3.

4. SPATIOTEMPORAL RETRIEVALS

In GIS which deals with discrete temporal events, the types of retrievals are categorized to the following:

(1) Spatial retrieval over the present data
(2) Spatial retrieval over at a specified time data
(3) Detection of the difference between two specified times.

The first type of retrieval is not temporal retrieval. Conventional GISes usually execute this type retrieval. The proposed data structure can also execute the retrieval, just not using the extension for temporal data, as described in [10]. There is no retrieval time loss by the extension. In the following two subsections, the other two types of retrievals are described.

4.1 Expansion of GBD-tree for temporal data

Basic idea how to manage spatio-temporal data is described in [10]. The method uses geographic differential script file (GDSF) to record past data. Then, this paper describes about abstract feature of the data structure. The largest difference between the original GBD-tree and the structure for spatio-temporal data is that the latter has a priority queue attached to every leaf node to store past data.

GBD-tree uses different mechanisms when to register data and to retrieve data. The characteristic is very important when to insert temporal data. The leaf node in which the temporal data is to be inserted is determined depending on the center position of the data. On the other hand, the node is not determined uniquely on R-tree and its successors because the data are divided by only the shape of the MBR at the time that the division is invoked.

The priority of the queue is ordered by the time-print from new to old. To distinguish the MBR of temporal data from that of present data, let the MBR of the present data be denoted as MBRp, and the MBR of the temporal data be denoted as MBRt.

Retrieval from the queue is also required in order to restore the old timed data. In order to avoid omission of temporal data during spatial retrieval, each node of the GBD tree has the MBR of temporal data. When restoring the map of a specific date, spatial retrieval is executed, referencing the MBR of temporal data attached to each node. If the MBR of temporal data overlaps the specified retrieval area and if the node is a leaf node, then the queue storing temporal data is reviewed and applied to the command that is controlling the current data set.

4.2 Spatial retrieval of specified time data

The most typical spatial retrieval is range retrieval. Other spatial retrieval (for example, to find nearest neighbor) can also be executed by a combination of range retrieval. The following describes how to do range retrieval on the proposed data structure.

First, retrieval of the present data is done by the usual method of searching the GBD tree. Specifically, the overlap between the specified retrieval area and the MBRp on each node is inspected. If these two are overlapping, descend the tree and repeat the same check for all the child nodes. If the node is a leaf node, select the entities actually included in the specified range and add the entities to the result set. In this searching process, there is no overhead time caused by adding temporal information to the tree structure.

When a particular time (old date) is specified outside of the retrieval range, the priority queues attached to the leaf nodes are inspected. Range retrieval of a specified time is executed by the following steps, let the specified time be T and the specified range be R.
(1) empty the result set S.
(2) check whether MBRp or MBRt overlaps with R
(3) if they are overlapping and the node is not a leaf node, descend the tree and repeat the check for all child nodes.
(4) if they are overlapping and the node is a leaf node, copy the data in the leaf node to a working space W.
(5) apply the GDSF commands attached to the node whose time print is not older than T.
This operation is executed on the data in the working space W. Add the result of W to S.

Fig.7 summarizes the steps of temporal retrieval. Every leaf node consists of two parts, the present data and the historical data queue. After having reached a leaf node, the present data part is copied to a working space W. Then the GDSF commands are applied until specified time. The contents in W are modified to the states of the specified time. Finally, spatial retrieval is performed on W.

The most important points of the method are in step (5). The first point is that the range retrieval with specified time is executed on a working space W, and there are no effects on the spatial data stored in the GBD tree. The second is that the influence of the operations by the GDSF commands is restricted within each leaf node. For example, a command to insert an entity has no effect on the other leaf nodes. When there is an insertion command on a priority queue of a leaf node, execution of this command has no effect on the results of the operation on the other nodes. In addition, if there is a deletion command on a priority queue of a leaf node, the object must exist on the same leaf node, because the GBD tree distributes entities to individual leaf nodes according to the center points. Also, deletion commands have no effect on other leaf nodes.

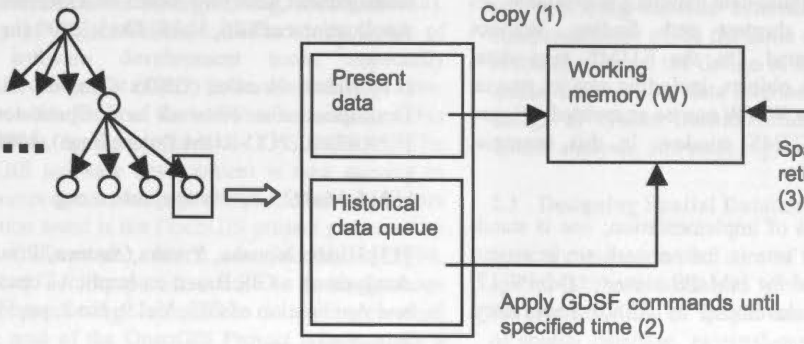


Fig.7 Steps of spatio-temporal retrieval

4.3 Detection of the difference between two specified times

The comparison of two specified points in time is a frequently requested operation in spatiotemporal GIS. The proposed data structure can execute this type of operation easily.

Let two specified points in time be t1 and t2 (t1 < t2: t1 is older than t2). This type of retrieval, the step (1) in Fig.7 is not necessary, because the state of the present has no effect on the result. Fist, skip the historical data queue until the time print is newer than t2. Then, execute the GDSF commands stored in the queue in the working memory W while the time print of the command is newer than t1. When the steps have completed, the entities existing in W are a part of the result. To do this operation over the leaf nodes overlapping the specified area, the change can be detected in the area.

5. DATA FORMAT

STIMS consists of several data files. In this section, the format of these files are described.

The *system definition file* is the most important file which describes all files used in STIMS and defines important parameters.

Spatial index file is for high-speed spatial retrieval, which is organized by GBD-tree. The contents of *vector file* and *connector file* are managed by this index.

Legend definition file defines several attributes of depicted line and regions, including the kind of hatching, kind of line, thickness and color.

Theme definition file define several kinds of theme used in several kinds of spatial retrieval. A theme is defined by a ordered set of line kinds. For example, city shape is defined the region divided by shoreline, country border, prefectural border, city border, and sheet border. The theme definition file defines such kinds of line for each theme.

Differential script file is for synchronization with external GIS. This file is similar with transaction file, recording addition and deletion of geographic entities.

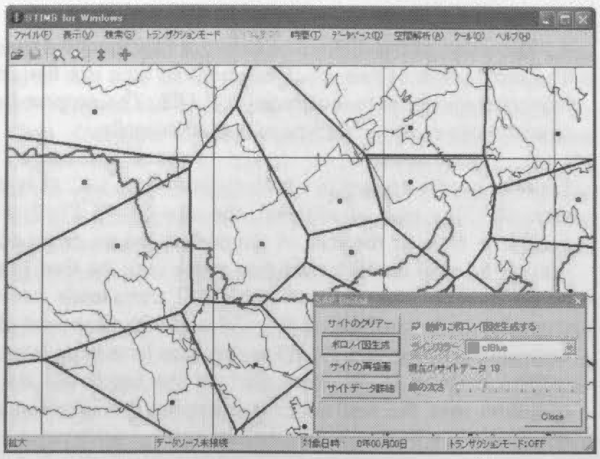


Fig.8 Voronoi diagram on STIMS

Current version of STIMS has simple DBMS to manage attribute information. The *attribute database file* showing in the figure is this file. For more complicated query, ordinary database managenet system, for example SQL server, MS ACCESS, or Postgre SQL, are used. To connect these attribute databases, STIMS provides two interfaces. The first one is *attribute database interface file*, and the other one is communication using ODBC. Suppose an order to color each city area to distinguish the population density according to a legend, the send to external attribute DBMS. Then the DBMS replay the result by an attribute database interface file.

6. STIMS IMPLEMENTATION

STIMS is now implemented on PC by VC++. At present, basic spatio-temporal data and attribute data handling is available. As spatial analysis functions, shortest path finding, Voronoi diagram are now implemented. On the STIMS map data, several kinds of multi-media objects, including image, movie, Word document, and URL for WWW can be embedded. Figure 8 shows an example of STIMS window. In this example, Voronoi diagram is displayed.

We are developing two kinds of implementation; one is stand-alone version, and the other one is for network environment. The latter one is implemented for JaMaPS viewer. JaMaPS[12] is an internet GIS viewer developed by KDDI Laboratory, Japan.

The implicit topology description model using in STIMS require much calculation time than conventional GIS. To solve this problem, we have developed topology cache method, and have implemented in STIMS already. By this topology cache, STIMS can work very fast.

REFERENCES

[1] Tor Bernhardsen, "Geographic Information Systems – An Introduction (second edition)", John Wiley & Sons, 1999

[2] Shigeru Kakumoto, Michinori Hatayama, Hiroyuki Kameda, Tokihiko Taniguchi, "Development of Disaster Management Spatial Information System (DiMSIS)", Proc. Of IEAS'97, 1997

[3] Shingo Otomo, Yutaka Ohsawa, "A Study on a No-topological Geographic Information System", Proc. 8-th Functional Graphics Information System Symposium, pp.27-32, 1997 (in Japanese)

[4] N. J. Nilsson, "Problem-solving Methods in Artificial Intelligence", McGraw-Hill, 1971 [3] Yutaka Ohsawa, Masao Sakauchi, "A New Type Data Structure with Homogeneous Nodes Suitable for a Very Large Spatial Database", Proc. of 6th International Conference on Data Engineering, pp.296-303, 1990

[5] Alfred V. Aho, Jeffrey D. Ullman, "Foundations of COmputer Science", p.483, Computer Science Press, 1992

[6] A.Guttman: R-trees: A Dynamic Index Structure for Spatial Searching", Proc. ACM SIGMOD Int. Conf. On Management of Data, pp.47-57, 1984

[7] N.Beckmann, H.P.Kriegel, R.Schneider, B.Seeger, "The R*-tree: An Efficent and Robust Access Method for Points and Rectangles", Proc. of the 1990 SIGMOD, pp.322-331, 1990

[8] Yutaka Ohsawa, Masao Sakauchi, "A New Type Data Structure with Homogeneous Nodes Suitable for a Very Large Spatial Database", Proc. of 6th International Conference on Data Engineering, pp.296-303, 1990

[9] Yutaka Ohsawa, Masao Sakauchi, "A Proposal of Multi-Dimensional Data Structure with Two Kinds of Auxiliary Data for Retrievals and Managements", IEICE Trans., J74-D-1, 8, pp.467-475, 1991 (in Japanese)

[10] Yutaka Ohsawa, Kim Kyungwol, "A Spatio-temporal Data Management Structure Based on Different Script", Theory and Applications of GIS, Vol.8, No.1, 2000 (in Japanese)

[11] Hideki Nonaka, Yutaka Ohsawa, "An Implicit Topology Description for Network and Spatio-temporal GIS", Proc. ISPRS2000, (CD-ROM Proceedings), 2000

[12] JaMaPS: http://www.jamaps.org

[13] Hideki Nonaka, Yutaka Ohsawa,"Processing Time Analysis on a GIS Based on Implicit Topology Model", Theory and Application of GIS, Vol.9, No.2, pp.17-24, 2001 (in Japanese)

MAPMANAGER: THE DESIGN OF A COM-BASED GIS COMPONENTWARE

Xiaojun Tan[a], Fuling Bian

Wuhan University, School of Remote Sensing and Information Engineering, Wuhan, 430079,China
[a] txj72@263.net

Commission IV, WG IV/1

KEY WORDS: GIS, COM/DCOM, Spatial Database, Spatial ADT, Structured Storage, Spatial Access Method, Spatial Query, Visualization

ABSTRACT:

GIS software development is now moving in the direction of components. The design of a two-dimensional GIS componentware – MapManager, which is based on the COM/DCOM technology, is introduced here. In the process of GIS software design, we should mainly consider the following aspects: spatial database design, visualization, data interface, spatial query and analysis, editing and mapping. At last, we discuss the shortcomings of MapManager and identify the need for future research in the integration with DBMS.

1. INTRODUCTION

GIS software and its applications have made great progress since 1960s. Its development could be technically divided into three stages (Fang, *et al.,* 2001b). Because of the complexity of spatial data and the relationships among spatial entities, there were great limitations in the design of early GIS software that was developed in a relatively closed environment. This is the main reason why GIS software was kept far away from the mainstream of information technology (Fang, *et al.,* 2001a). With the maturity of Object-Oriented methodology, a lot of Object-Oriented software development tools, especially component-based development (CBD) tools, are available now. It has already been a tendency of the software technology. This also deeply influences the development of GIS software. The third generation GIS software development is now moving in the direction of components. An important milestone indicates the componentization trend is the OpenGIS project proposed by OGC (Open GIS Consortium), which was founded in 1994. OpenGIS is defined as transparent access to heterogeneous geodata and geoprocessing resources in a networked environment. The goal of the OpenGIS Project is to provide a comprehensive suite of open interface specifications that enable developers to write interoperating components that provide these capabilities (OGC, 1996). The componentization of the major GIS software began in the middle of 1990s, and was almost finished by the end of twenty centuries (Fang, *et al.,* 2001b). Compared with the traditional GIS software, GIS componentware has the following advantages: efficient and seamless system integration, specific programming language unnecessary, popular, and cheap (Song and Zhong, 1998). This paper introduces the design of MapManager – a small two-dimensional GIS componentware.

2. DESIGNING MAPMANAGER

There are two important standards for CBD: Microsoft COM (Component Object Model) /DCOM (Distributed Component Object Model), and OMG (Object Management Group) CORBA (Common Object Request Broker Architecture). COM/DCOM takes a lead position in the market and becomes the industrial standard gradually. The ActiveX controls, which are based on the COM/DCOM technology, are the components used most widely in visual programming nowadays. MapManager, a GIS ActiveX control with 41 OLE automation objects and 627 interfaces (including properties, methods and events), is based on COM/DOCM. The design of these objects and interfaces is very important to the functionality of the software. Besides conforming to componentware specifications and providing standard external interfaces, the design of GIS componentware needs the same consideration as traditional GIS software such as: the design of spatial database; visualization of spatial information; data interface with external system; and the design of system functionalities (including editing, querying, spatial analysis, and mapping).

2.1 Designing Spatial Database

Spatial database is the core of GIS software. The design of a spatial database takes into account the following four aspects: the representation of spatial data types, internal-memory model of spatial database, external-memory model as well as spatial access method.

Early GIS software used the file system to store the geometric data structures. There has been suggestion and request for the use of database management systems since the early 1980s. It is now being enabled by industry's support of specialized spatial versions of database systems (e.g., Spatial Data Blades and the Spatial Data Option) or middleware (e.g., the Spatial Data Engine). To store the geometric data structures, most commercial relational databases provide long fields (also called binary large objects or memo fields) that serve as simple containers. One of the columns in the relation is declared to have variable length. The geometric representation is then stored in such a long field in a way that only the application programs can interpret it, while the database system itself usually cannot decode the representation. It is, therefore, impossible to formulate or process SQL queries against that column. To implement geographic concepts in a high-level, compact, and reusable way, the use of spatial data types has become the standard method. Although early attempts were made to rely exclusively on those data types offered by standard programming languages and database systems, it has become common practice to identify spatial data types and to link them with their related operations. Object-oriented methods, with encapsulation and hiding of implementations, have favored this approach. Abstract data types (ADTs) based on object-oriented methods provide a robust way to implement complex data types. The basic idea is to encapsulate the implementation of a data type in such a way that one can communicate with instances of the data type only through a set of well-defined operators. The internal implementation of the data type and its operators are hidden to any external users. They have no way to review or modify those interior features. Object-oriented concepts can easily be adapted to the implementation of spatial ADTs and operators (Egenhofer, *et al.,* 1999). The OGC Technical Committee has released an implementation specification named "OpenGIS Simple Features Specification for OLE/COM Revision 1.1", which has a detailed definition on the geometric class hierarchy (see Figure. 1). The definitions of spatial ADTs in MapManager are based on a subset of those geometric classes, which covers all except GeometryCollection and all its subclasses.

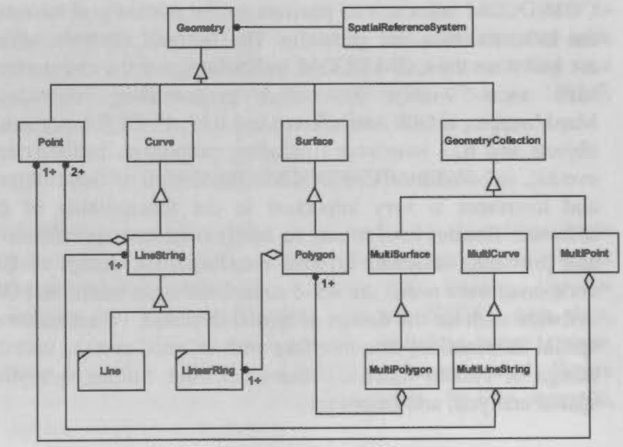


Figure 1. Geometry Class Hierarchy (from OGC 1999a)

The internal-memory model of spatial database refers to how three kinds of information – spatial metrical information, spatial topological information as well as non-spatial information – are arranged in the physical primary memory. All the features in the primary memory are stored in a global object named "Database". There are four spatial data sets, named node data set, arc data set, line data set and polygon data set respectively, and one geo-feature data set in the "Database" object. The spatial information – both metrical and topological – is stored in the four spatial data sets. The node data set stores the node's ID, coordinates, and IDs for all arcs containing this node in anti-clockwise order. The arc data set stores the arc's ID, coordinates list for all points on this arc, and IDs for all lines and polygons containing this arc in anti-clockwise order. The line data set stores the line's ID, IDs for all arcs contained in this line, and IDs for all geo-features containing this line. The polygon data set stores the polygon's ID, IDs for all arcs on the boundaries of this polygon, and IDs for all geo-features containing this polygon.

In the geo-feature data set, semantic information is defined for all geo-features (e.g. road, river, building etc.). There are four kinds of geo-features in MapManager: point feature, line feature, polygon feature and annotation feature. The coordinates of point feature and annotation feature are stored in the geo-feature data set, while those of line feature and surface feature are stored in the above four spatial data sets.

The external-memory model of spatial database refers to how the data is stored persistently in the secondary memory. As stated above, it is a trend to introduce commercial database management system into spatial data management. However, it may be better to use file system for small GIS software like MapManager. Fortunately, the structured storage technology provided by COM/DCOM is a good solution for persistent storage for MapManager. Structured storage is sometimes called "a file system within a file" because it can treat a single file as if it is capable of storing directories and files. A file created using the structured storage service contains one or more *storage*, roughly equivalent to directories, and each storage can contain zero or more *stream*, roughly equivalent to files. A storage can also contain any number of *substorages* (Eddon and Eddon, 2000). The structured storage of MapManager is shown in Figure 2.

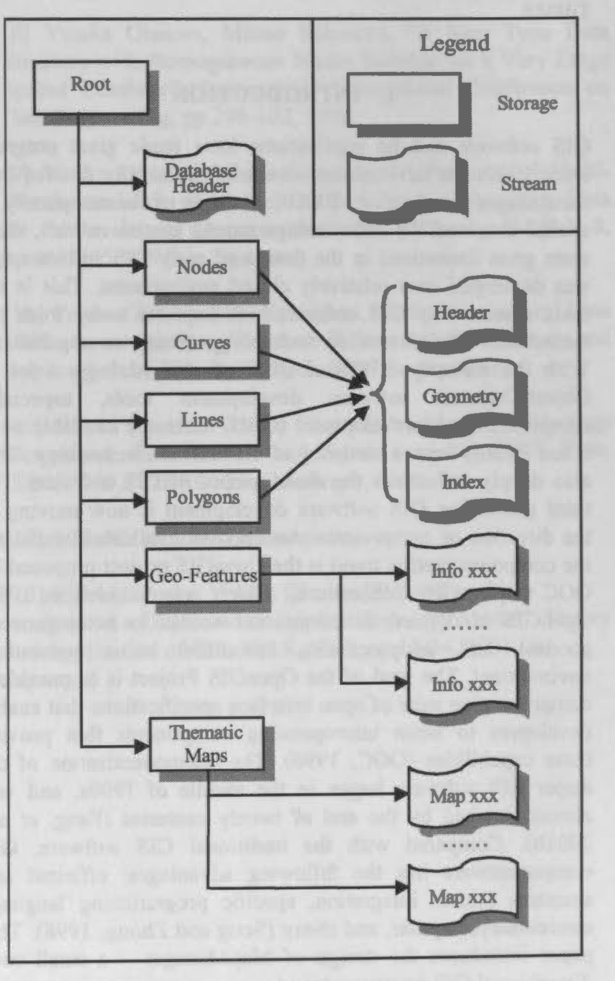


Figure 2. Structured Storage of MapManager

In order to deal with massive data in large application, a feasible spatial access method (SAM), which manages the data exchange between internal and external memory, is needed. The foundation of SAM is spatial index technology. So far there are a lot of spatial index structures such as quadtree, K-D-tree, K-D-B-tree, R-tree, R+-tree, etc. (Samet, 1990; Samet, Aref, 1994; Worboys, 1995; Garcia-Molina, *et al.,* 2000). A common approach to search spatial objects consists of a two-step process: (1) choosing an approximation (e.g., a simpler shape, such as Minimum Bounding Rectangle) that can be indexed and serves as a fast filter and (2) using the original geometry to assert the retrieval condition only for the initially retrieved objects to eliminate false hits. An index may only administer the MBR of each object, together with a pointer to the description of the object's database entry (the object ID). With this design, the index only produces a set of candidate solutions. This step is therefore termed the filter step. For each element of that candidate set we have to decide whether the MBR is sufficient to decide that the actual object must indeed satisfy the search predicate. In those cases, the object can be added directly to the query result. However, there are often cases where the MBR does not prove to be sufficient. In a refinement step we then have to retrieve the exact shape information from secondary memory and test it against the predicate. If the predicate evaluates to true, the object is added to the query result as well, otherwise we have a false drop (Egenhofer 1999). The spatial index structure of MapManager is square grid, which can manage hundreds of megabyte data stably and efficiently in practice with the help of spatial object's MBR. However, its performance is degraded quickly when the size of database grows.

2.2 Visualization

The visualization of spatial information includes the following four aspects: map projection and coordinate system, management of thematic maps and layers, symbolization, and visualization of statistical data. Like the other GIS software, MapManager supports all kinds of common map projections and coordinate systems. The visualization of statistical data could be carried out with the help of other commercial componentware. Here we mainly discuss the management of thematic maps and layers as well as symbolization.

In MapManager, data in spatial database is represented by the means of thematic maps and map layers. There are several thematic maps in a spatial database, and the arrangement of these thematic maps is similar to an atlas. Each thematic map contains several map layers, each of which has a property named "Symbol" and is linked to a kind of geo-feature in the geo-feature data set. When a map is displayed, all the features contained in the map layer are drawn with the proper symbol. All the map layers in a thematic map are organized as a tree-like hierarchy. There are no limitation on the height of the tree and the number of the sub-nodes.

There are four kinds of symbols, point symbol, line symbol, polygon symbol, and annotation symbol, according to four kinds of geo-features. A special program, Symbol Designer, has been written to help designing symbols (see Figure. 3). When a geo-feature is symbolized, the coordinate lists of that feature are passed to the symbol and the symbol itself will finish the drawing procedure. We put forward the concept of "exceptional symbol" to deal with the situation that features of the same type require different symbols. Each feature with an "ExceptionalSymbol" property will be treated differently by the drawing program. The "exceptional symbol" could avoid the conflict between the commonness and individuality of symbols. However, the drawing performance will be degraded if there are too much exceptional symbols.

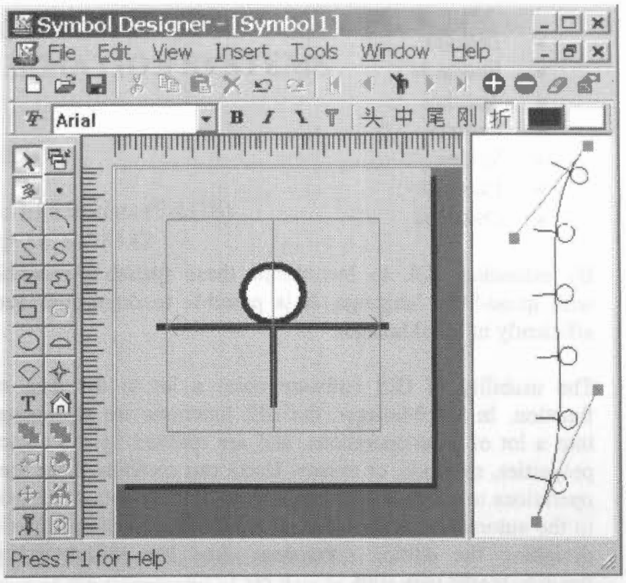

Symbol Designer - [Symbol1]
File Edit View Insert Tools Window Help
Arial
B I
头 中 尾 刚 折
T
Press F1 for Help

Figure 3. Symbol Designer Program

2.3 Data Interface

In terms of data input, MapManager doesn't offer the functionality for digitizing as well as image scanning. However, it provides consistent data interface to import spatial data from outside. This process is reversible – users can also export data to the other system. Currently, MapManager provides four vector data interfaces: Autodesk DXF file, ESRI E00 File, ESRI shape file, and Mapinfo Interchange File.

2.4 System Functions

Here we will mainly discuss three aspects of system functions: spatial query, editing and mapping.

For many years, the database market has been dominated by the Structured Query Language (SQL). There has been a long discussion in the literature as to whether SQL is suitable for querying spatial databases. It was recognized early on that relational algebra and SQL alone are not able to provide this kind of support. In some sense, however, with the success of SQL the discussion about its appropriateness has become a moot point. The question is not whether SQL should be used — SQL is and in the foreseeable future will be used to query spatial databases as well. The question is rather which kind of extensions are desirable to optimize user friendliness and performance of the resulting spatial data management system (Egenhofer, *et al.,* 1999). Various extensions to SQL have been proposed to deal with spatial data (Egenhofer, 1992). The ANSI Committee on SQL3 are developing an integrated version of such spatial extensions called SQL/Multimedia (SQL/MM), which is a suite of standards that specify type libraries using SQL's object-oriented facilities. A "quasi-SQL" language, which is an extension to SQL, is used in MapManager to implement the spatial query function. It is based on the definition of spatial relationship predicates. In OGC's implementation specification, there are five named predicates based on the dimensionally extended nine-intersection model (DE-9IM): Disjoint, Touches, Crosses, Within and Overlaps. For user convenience, eight spatial relationships between spatial objects are given as following (OGC, 1999b):

- Equals

- Disjoint
- Intersects
- Touches
- Crosses
- Within
- Contains
- Overlaps

By extending SQL to incorporate these spatial relationships with quasi-SQL language, it is possible to do spatial query efficiently in MapManager.

The usability of GIS software owes a lot to the data edit function. In MapManager, the edit functions are decomposed into a lot of meta-operations, and are realized by the objects' properties, methods, or events. Users can combine these meta-operations to achieve different effects. We pay lots of attention to the automatic maintenance of topological relationship when designing the editing subsystem. And the object-snapping function works very well.

MapManager has powerful mapping capability. There are a whole set of layout element objects that satisfy a lot of applications. The layout could be output to either hard copy devices or standard output formats such as Encapsulated PostScript file.

3. CONCLUSIONS

Some professional applications have been developed on MapManager version 1.0, which is proven to be an applicable GIS component. However, there are many shortcomings in the software, and some are listed as following:

- No supporting for complex collection spatial objects
- No concurrency control, because without the help of DBMS
- SAM not efficient enough
- Drawing performance degraded by the hierarchy of symbol object

There are lots of works to do to overcome all these shortcomings. First of all, what we want to research next is to provide accessibility to spatial data in large commercial DBMS.

REFERENCES

Eddon, G., Eddon, H., 2000. *Inside COM+ Base Services.* Microsoft Press, Redmond, Washington.

Egenhofer, M. J., 1992. Why not SQL! *Int. J. Geographical Information Systems,* 6(2), pp. 71-86.

Egenhofer, M. J., Glasgow, J., Günther, O., Herring, J. R., Peuquet, D. J., 1999. Progress in computational methods for representing geographical concepts. *Int. J. Geographical Information Science,* 13(8), pp. 775-796.

Fang, Y., Tian, G. L., Shi, Z. Z., Zhou, C. H., 2001a. Modert Itand 4th GIS software. *Journal of Image and Graphics,* 6A(9), pp. 824-829.

Fang, Y., Zhou, C. H., Jing, G. F., Lu, F., Luo, J. C., 2001b. Reasearch of 4th Generation GIS software. *Journal of Image and Graphics,* 6A(9), pp. 817-823.

Garcia-Molina, H., Ullman, J. D., Widom, J., 2000. *Database System Implementation,* Prentice Hall, Inc.

OGC, 1996. The OpenGIS™ Guide. http://www.opengis.org/techn/guide/guide1.htm

OGC, 1999a. OpenGIS® Simple Features Specification for OLE/COM. http://www.opengis.org/techno/specs/99-050.pdf.

OGC, 1999b. OpenGIS® Simple Features Specification for SQL. http://www.opengis.org/techno/specs/99-049.pdf.

OGC, 1999c. The OpenGIS™ Abstract Specification – 99-AS-RFP009. http://www.opengis.org/techno/abstract/99-AS-RFP009.pdf.zip.

Samet, H., 1990. *The Design and Analysis of Spatial Data Structures,* Addison-Wesley.

Samet, H., Aref, W. G., 1994. Spatial data model and query processing. *Modern Database Systems: The Object Model, Interoperability, and Beyond.* Addison Wesley/ACM Press, Reading, MA, pp.338-360.

Song, G. F., Zhong, E. S., 1998. Research and Development of Components Geographic Information Systems. Journal of Image and Graphics, 3(4).

Worboys, M.F., 1995. *GIS: A Computing Perspective,* Taylor & Francis Ltd, London

A MULTI-LEVEL APPROACH FOR 3D MODELLING IN GEOGRAPHICAL INFORMATION SYSTEMS

Fabien RAMOS

EADS Systems & Defence Electronics (S&DE)
6 rue Dewoitine, BP 14, F-78142
Velizy-Villacoublay Cedex (France)
Fabien.ramos@sysde.eads.net

KEY WORDS: Topology, Modeling, *GIS*, Network analysis, Data coherence.

ABSTRACT:

The paper focuses on the problem of 3D data modeling within Geographical Information Systems (GIS) .
The approach presented here brings a solution to manage the diversity and the complexity of 3D data in a Geographical Information System. Our approach is not to use one unique model but 2 levels of modeling in order to describe the whole 3D scene.
The first level uses a 3D topology composed of nodes, edges, and faces. A volume is defined by its shell. This model insures object's consistency. Each object is composed of constructive structure, e.g. a house is a set of three constructive structure which are the roof, the wall and the floor. Each constructive structure is made of structural primitives. Several components of one object or several objects could share the same topological primitive.
The second level uses a network topology composed of nodes and arc. It could describe physical networks as well as abstract networks. Road network is an example of physical networks in which road sections are classed as arc and crossroad or buildings are classed as node. Abstract networks allow creating a link between objects, which are possibly physically independent ones from the other. Buildings sharing the same Intranet network could illustrate those "abstract networks". This level is very useful to support queries on classical networks or to materialize dependencies between objects. Computation of optimal path on graph is also related with this level of structuring.

RESUME :

Le propos de cet article est d'aborder les problèmes de modélisation des données 3D au sein d'un système d'information géographique (SIG).
Nous proposons ici une solution pour manipuler des données 3D complexes et diversifiées dans un système d'information géographique. Notre approche n'est pas d'utiliser un modèle unique de données pour décrire l'ensemble des informations d'une scène 3D, nous préférons une modélisation suivant 2 niveaux descriptifs.
Le premier niveau décrit une topologie 3D de composition s'appuyant sur les primitives nœud, arête, face et volume. Un volume est défini par son enveloppe externe. Ce modèle assure une cohérence 3D des objets. Chaque objet est composé de structures constitutives, une maison se décompose par exemple en toit, murs et sol. Chacune de ces structures constitutives se décompose elle même en primitives structurelle (nœud, arête, face, volume).
Le deuxième niveau de description s'appuie sur une topologie de réseau classique fondée sur des primitives de type arc et de type nœud. Ce niveau topologique peut décrire aussi bien des réseaux physiques réels que des réseaux abstraits. Le réseau routier est un réseau physique alors qu'un réseau intranet reliant différents bâtiments d'une entreprise sera qualifié d'abstrait. Ce niveau de modélisation est très utile pour répondre à des requêtes classiques sur réseaux (de type calcul d'itinéraire par exemple) et permet de matérialiser des dépendances entre objets.

1. INTRODUCTION

Huge communities of geographical and spatial data users as geologists, militaries, town planners or communication and utility managers are interested in a GIS being able to handle the third dimension. Lot of commercial solutions that can be found on the GIS market don't actually deal with a true 3D description of objects. Most of the time they are limited to a simple 3D extrusion of 2D outlines, which could be sufficient to provide a general idea of the 3D aspect of the scene. Unfortunately, because they use a unique z altitude for each couple of (x,y) coordinate, such models have important lacks in term of interaction and manipulation of 3D data

In order to fill this gap, several research teams have been involved in that issue and some interesting 3D topological models have therefore been developed for a few years. Among the different solutions proposed, we can find two main tendencies:
Some models, generally inspired by CAD products, are essentially based on geometrical description of the object. Simple topological characteristics could possibly be added.

Other models give more weight to topological aspects. Those models are often issued from GIS communities and they suggest a 3D extension of classical 2D topological models. They

introduce new topological primitives, new topological relationship and they manage to extend 2D classical model to the third dimension.
Each of those models owns some advantages as well as some disadvantages. Geometrical models offer faster time in computing but they are not able to bring efficient solutions to solve more specific problems like network analysis or data coherence issues.
In this paper we focus on the topological issues and we present a conceptual model fitted to the needs of our multidimensional geographical information system prototype.

2. TOPOLOGY IN GIS

2.1 Definitions

Topology is the mathematical study of objects properties which are preserved through deformations, twistings, and stretchings.

There is also a formal definition for a topology defined in terms of set operations. A set *X* along with a collection *T* of subsets of it is said to be a topology if the subsets in *T* obey the following properties:

1. The (trivial) subsets *X* and the empty set ϕ are in *T*.
2. Whenever sets *A* and *B* are in *T*, then so is $A \cap B$.
3. Whenever two or more sets are in *T*, then so is their union

Although this definition of topology is rigorous, it remains abstracted and quite distant from geomatic's people concern.

Moreover the term "topology" used in geomatics does not refer to the theories surrounding this definition, but is rather located within the framework of graph theory.

A graph is a binary relation in a set. If vertex are the elements of this set and edge the couples of vertex in relation, one can say that a graph is also the data of a couple (V, E) where V is the set of vertex and E the set of edge, formed by couples of vertex (Langlois,1994).

In other terms, a Graph G=(V,E) consists of a set of vertices, V, and a set of edges, E. Each edge e member of E is a pair, e=(u,v), where u and v are member of V.

But most of the time in geographic information the term topology tend to group together all the relations between objects as well as topological relations, order relations, or even directional relations.

2.2 Interest of 3D Topology

Whether it is through 2D GIS or 3D GIS, the use of topology is interesting on several points where it comes to complete the information conveyed by a purely geometrical description.

Owing to the fact that some geometrical calculations on inclusions, adjacencies, boundaries or network analysis are expensive in term of computing resources, the International Organization for Standardization (ISO) recommend to lean on the combinatorial structures of the topological complexes and to convert geometrical calculations in combinatorial algorithmics.

Insuring a better coherence of data is a second advantageous contribution of the topology. This coherence limits the errors of spatial analysis and avoids some display aberration connected to geometrical incoherencies.

(Zeitouni, 1995) explains that the first interest of topology lies in its spatial semantic arguing that the way objects are laying out in a given space constitutes a natural model more easily understandable by users.

Sometimes topology is presented as being able to limit the size of data bases. It is true that topology allows avoiding redundancies led by duplicated geometrical primitives, but a detailed topological description can also be very voluminous.

Unfortunately topological information storage is not exempt from every constraints. Beyond the possible problems of data volumes to be stored, the major inconvenience of a topological description is linked to integration and data update issues. The availability of 3D data is a real problem today from which it is difficult to cast off the framework of a three-dimensional geographic information system.

2.3 3D Topological Models

Although previous published works seem to assert that a boundary description suits well with 3D topological problems, two lineages of models distinguish themselves:

- A first classic approach decomposes the topological space in 4 fundamental types of primitive which are node, edge, face and volume.
- A second approach inspired by combinatorial topological cards (DCEL or winged-edge structure) introduces the concept of primitive associating edge with face.

We shall present there only the main lines of these models, knowing that they can be variously declined according to the objectives to reach.

2.3.1 Classical Modeling

This rather simple approach of topology is adopted by numerous authors of whom Martien Molenaar (Molenaar, 1990). It consists in associating the four basic primitives that are node, edge, faces and volumes with construction and inclusion relations. The basic skeleton of this conceptual model is summarized in Figure 1.

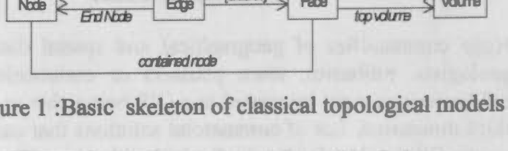


Figure 1 :Basic skeleton of classical topological models

Numerous variations were proposed from this skeleton.

(Molenaar, 1990) introduces the notion of directed arc between edge and nodes. However the model is limited to the case of planar faces, implying that topology depends on geometry.

ISO organization take its inspiration with this model but recommends a total independence between the topological model and the geometrical model.

A software package developed under the DARPA image understanding program defines an interesting topological concept which is the k-chain, where k is the number of dimension of the element linked by the chain. An 0-chain is a sequence of points, a 1-chain is a sequence of connected edge, a 2-chain is a sequence of connected face.

Kevin Trott resumes partially this concept of k-chain in his proposition of 3D extension for VPF format. He names "ring" such sequence of edge and "shell" such sequence of face (Trott, 1999). Contrary to the software Target Jr the notion of volume is present in Trott's model (Target Jr describes only boundaries of volume).

2.3.2 3D Extensions of Topological Map

With regard to a simple model dealing with only composition relations, the model of 2D topological map brings the notion of edge cycles around nodes. This notion of edge cycles could be found in numerous 2D models as DCEL structure (David, 1991) and winged-edge data structure.

In the extension of this concept to the third dimension Kevin Shaw (Shaw, 1998) and Arnaud De La Losa (De La Losa, 1999) suggest articulating the 3D objects description around edge by introducing a new primitive named EFace in (Shaw, 1998) and (Arc, Face) couple in (De La Losa, 1999).

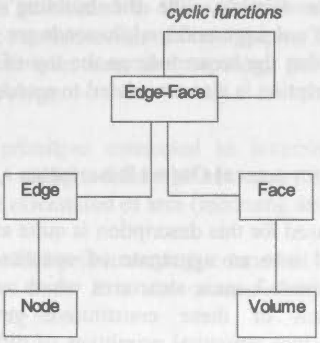


Figure 2 : 3D Topological map extension model

Figure 2 clearly shows that face primitive and edge primitive are not directly connected any more. Connections are made by operators or functions applied to this Edge-Face new primitive.

The introduction of this new primitive is the keystone of the 2 models, but they differ from one another in many aspects examined in 2.3.1.

2.3.3 Other Topological Models

Both approaches we have described represent main trend in topological modeling. However some more marginal attempts, that are not without interest, should be mentioned.

2.3.3.1 The Face Adjacency Hypergraph

(Floriani, 1988) considers as topological information only the relations of adjacency. Furthermore the face adjacency hypergraph makes a distinction according to the dimension of the adjacency (adjacency according to a point or according to an edge) (Figure 3).

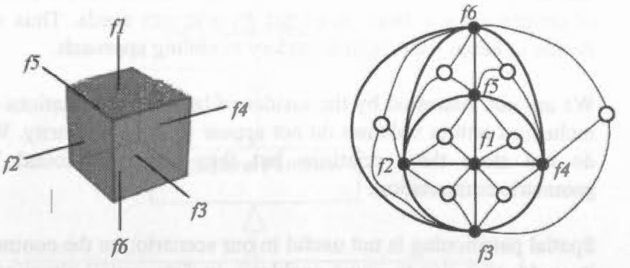


Figure 3 : Face adjacency hypergraph

2.3.3.2 Horizontal and Vertical Specification

(Zeitouni, 1995) introduces the notion of horizontality and verticality to take into account the specificity of the geographic data. The MNT is connected with 3D objects by the relation "is posed on".

3. A 3D GIS PROPOSITION

Within the framework of a preliminary study on modeling and exploitation of multidimensional geographic database, we have thought about a conceptual data model which is more convenient for our GIS objectives, needs and constraints.

3.1 Our Objectives

In order to determine the adequate conceptual data model, it is important to define well the objectives to achieve.

One of our main objectives is to be able to model data of various dimensions, that is to say punctual, linear, planar and voluminous data. Furthermore, the selected model should be able to describe any 3D object shapes.

We do not want to limit ourselves to the 2,5D as it is often the case in commercial 3D GIS extensions. It should be possible to model objects with one or more z value for a given couple of (x, y) coordinate. Many objects need this specification to be modeled, e.g. trees, bridge, archway, or simple buildings with door and windows modeled.

The algorithms of intervisibility, optimal path or network analysis should be able to be applied without difficulties on our data.

The prototype realized from this model should be able to perform algorithmic computation in "real time". The general ergonomics should not suffer from slow computations.

Data should be interactively editable without loosing their coherencies.

It should be possible to import 3D data without any trouble.

3.2 Needs, Constraints and Technical Choices

In the universe we try to describe, edge are very unusually shared by more than two faces. So cyclic relation, i.e. the notion of couple (edge - face), does not fit with our needs. Thus we decide to adopt a classical boundary modeling approach.

We are not interested by the insides of building, so relations of inclusions within volumes do not appear to us as a priority. We do not store these relations but they could be found by geometric computation.

Spatial partitioning is not useful in our scenario; on the contrary it could give rise to some problems in few spatial situations. That's why we have chosen to model volumes by their external shelves instead of modeling them in the strict sense of a volume entity.

Modeling complex surfaces, like NURB, is not an immediate necessity and is even problematic in term of performance for intervisibility computation. Therefore, our faces will be planar (and our edge rectilinear).

We describe 3D objects by their faces but edges have no utility within the framework of our GIS. We prefer to describe each face by an ordered list of nodes instead of an ordered succession of edges. We know that this is a limitation for the creation of faces with holes, but these last ones can be easily decomposed into 2 hemi-faces.

In order to keep inside and outside information, faces are modeled with their orientation. Edges are also directed.

A geographic object should be able to be associated with one or more topological description. On one hand it should be possible to describe its 3D intrinsic structure (a building will be described by wall, roofs and grounds which composes it) and on the other hand it should be possible to associated an object to one or more network topological primitives (e.g. a building can be associated to a node in a road network).

An object should be able to be decomposed into a set of elementary structures. A building is decomposed into 3 elementary structures: walls, roof and ground. These three elementary structures are themselves constituted of topological primitives which are nodes, edges, faces and volumes.

Network modeling is very different from structure object modeling. According to this important point, we have separated network modeling and structural object modeling in two different data conceptual model.

3.3 Our Conceptual Data Model

3.3.1 General Presentation

The 3D model retained for our 3D GIS prototype takes into account various points presented in previous paragraphs.

The UML Diagram of the Figure 5 summarizes our conceptual data model. Some of the relations between classes are not explicitly named in order to avoid displaying an unreadable figure.

On one hand geographic data is described in its 3D structure by a model based essentially on composition relations and on the other hand the same geographic data could be associated with a network in a network primitive form.

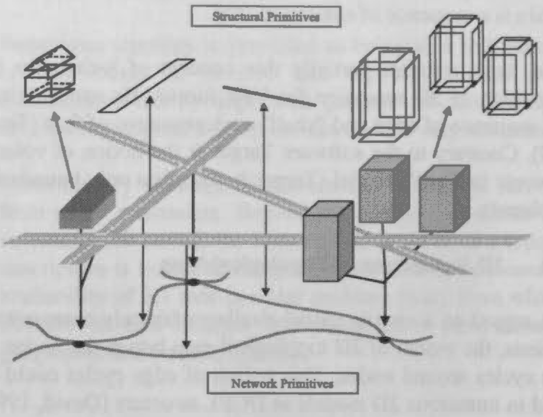


Figure 4 : Two modeling levels

On figure 4 you can notice that the structure of buildings consists of a set of face, the structure of the main roads consists of faces too but the structure of minor roads consists of edges.

Each of these objects or group of objects can be associated with network primitives. On Figure 4 we see that the house is associated to a node, like the building's block which is associated to a unique node, while roads are associated to arcs. Let us note that the crossroads on the top of the figure without any 3D description is also associated to a node.

3.3.2 3D Structural Object Description

The model used for this description is quite simple. Each object is composed into an aggregate of constitutive structures. A building possess 3 main structures which are roof, walls and ground. Each of these constitutive structures is itself decomposed into structural primitives of different dimensions. The basic primitive is the vertex, geometry is supported by vertices, i.e. each vertex is associated to a coordinate triplet. Edges and faces are described by their vertices and possess an orientation. Volumes are described by the faces which surround them. The notion of hole or cavity is not directly expressed on this model, they should be described by joining several primitives. By composition of these primitives any shape can be described whatever its spatial dimension is.
This model satisfies and guaranteed our constraints of coherence on 3D objects.

Note that there is not direct relations between faces and edges which compose their boundaries. The advantage of this conceptual choice is to limit the number of topological relations and thus to limit the complexity of the model. On the other hand our modeling is not the best one for a step by step navigation among topological primitives.

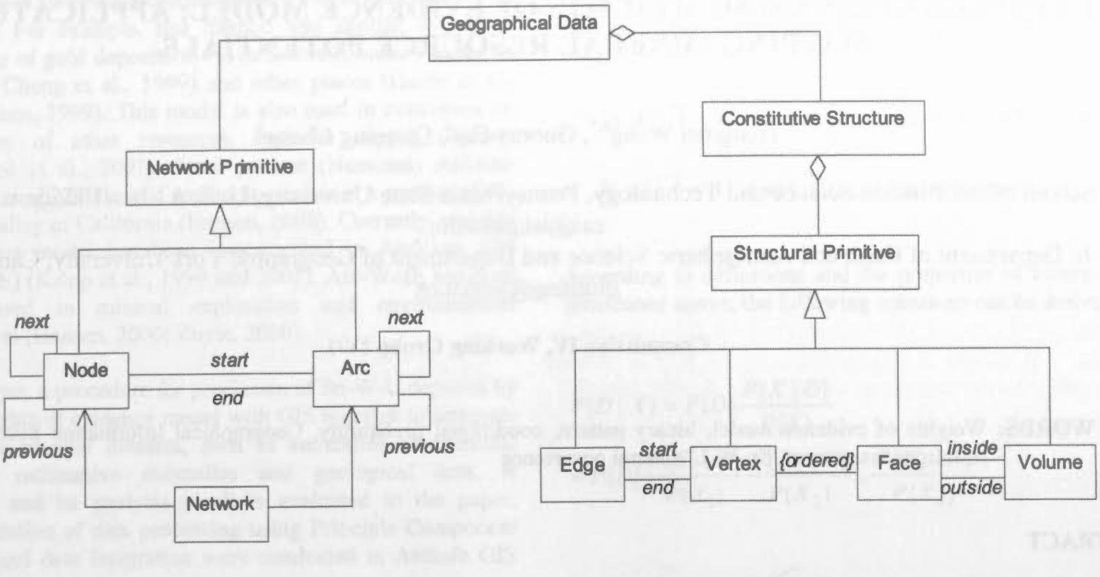


Figure 5 : UML diagram of the conceptual model

3.3.3 Network Topology

Network topology does not lean on the same primitives that structural topology. Primitives used here are those that one finds classically in 2D GIS modeling, i.e. node and arc. Our network modeling is a multi-value planar topological graph. No geometry is associated to these primitives, we do not propose graphic representation of these graphs. There is a full independence between structural topology previously seen and network topology.

A node is a primitive connected to incoming arcs and outgoing arcs. A node is linked to next and previous nodes according to the orientation of arcs (incoming and outgoing).

An arc is a directed primitive. It is connected with its start node and its end node. It is also connected with preceding arcs and following arcs.

The objective of this modeling is to be able to handle as well as possible all network analysis and queries. So it is well adapted to map out a route on a road network, to simulate pollution on a hydrological network or to look for electrical sub-networks.

4. CONCLUSION

The first characteristic of the data model which we have just presented is to completely dissociate 3D object modeling and interconnections between those objects. Our structural model of objects is certainly not the most complete, but it is perfectly advisable for our needs because it allows a minimal but sufficient topological description and its simplicity makes it rather sober and light to avoid being limited by long data access delay. Furthermore, more a data model is simple, easier are import, update and maintenance operations.
At last, networks are classically modeled by a multi-value planar graph. This structure has demonstrated its efficiency in network analysis for a long time.

REFERENCE

David, B., 1991. Modélisation, représentation et gestion de l'information géographique, une approche en relationnel étendu. Thesis, University of Paris VI.

De La Losa, A., Cervelle, B., 1999. 3D Topological Modeling and Visualisation for 3D GIS. Computers & Graphics, 23(4), pp. 469-478.

Floriani, L., Falcidieno, B., 1988. A Hierarchical Boundary Model for Solid Object RepresentationACM Transactions on Graphics, 7(1), pp. 42-60.

Langlois, P., 1994. Formalisation des concepts topologiques en géomatique. Revue internationale de géomatique, Vol. 4, pp 181-205.

Molenaar, M., 1990. A Formal Data Structure for three dimensional vector maps. In: 4th international symposium on spatial data handling, Zurich, Switzerland, Vol. 2, pp. 830-843.

Shaw, K., Abdelguerfi, M., and Ladner, R., 1998. VPF+: A VPF Extension Providing 3D Modeling and 3D Topology. In: Image '98 Conference, Scottsdale, Arizona, pp. KB1-KB8.

Trott, K. and Greasley, I., 1999. A 3D spatial data model for terrain reasoning. In: 4th International Conference on GeoComputation, Mary Washington College in Fredericksburg, VA-USA.

Zeitouni, K. and De Cambray, B., 1995. Topological Modelling for 3D GIS. In: 4th International Conference on Computer in Urban Planing and Urban Managment, Melbourne, Australia.

DATA INTEGRATION USING WEIGHTS OF EVIDENCE MODEL: APPLICATIONS IN MAPPING MINERAL RESOURCE POTENTIALS

Hongmei Wang[a,*], Guoray Cai[a], Qiuming Cheng[b]

a: School of Information Science and Technology, Pennsylvania State University, U. S. A - hzw102@psu.edu and cai@ist.psu.edu;

b: Department of Earth and Atmospheric Science and Department of Geography, York University, Canada - qiuming@yorku.ca

Commission IV, Working Group IV/1

KEY WORDS: Weights of evidence model, binary pattern, conditional probability, Geographical Information System, Mineral resources assessment, Sn-W-U mineral occurrence

ABSTRACT

Weights of evidence model is used to predict occurrence of an event with known evidences in a study area where training data are available to estimate the relative importance of each evidence by statistical methods. It provides a quantitative method for integrating multiple sources of evidences. It avoids subjective choice of evidences and subjective estimation of weights for evidences with comparing with other methods, such as Fuzzy logic method. This paper discusses our implementation of Weights of Evidence model as a data integration method in a GIS environment, and demonstrates working of the model using an application in the prediction of Sn-W-U mineral occurrence in the South Mountain Batholith in Nova Scotia, Canada. After introduction of the principal of weights of evidence model, the case study case is described, including data collection, data processing, and data integration with use of weights of evidence model. Finally, a Sn-W-U potential map with posterior probability values in the study area is generated and shown in this paper. Based on analysis of weights of five evidential maps and the final prediction result, it is concluded that the prediction result is successful and the model is useful in the study.

1. INTRODUCTION

The purpose of this paper is to demonstrate working of Weights of evidence model with using an application in the prediction of Sn-W-U mineral occurrence in the South Mountain Batholith in Nova Scotia, Canada. The research done here has two objectives: 1) to study the model for prediction of Sn-W-U deposits in the study area and 2) to evaluate analysis result of application of this model.

A Geographical Information System (GIS) is a "computer-based system which integrates the data input, data storage and management, data manipulation and analysis, and data output for both spatial and attribute data to support decision-making activities" (Malczewski, 1999). After over 40 years of development, GIS have been applied to serve important roles in many fields, such as environment monitoring, resources management, applications in commerce and business field, and different utilities. The ultimate purpose of GIS is to make evaluations or predictions with different specific data integration models to combine spatial and attribute data from various sources to provide support for decision-makers.

According to Bonham-Carter (1994), the data integration models in GIS are divided into two categories, data-driven and knowledge-driven models based on different methods for estimation of weights of different evidential maps. In data-driven models, the weights are calculated by using statistical methods and data of evidences in a training area to estimate the spatial relationships between the evidential maps and the final response maps. Data-driven models include Logistic Regression, Weights of Evidence, Neural Network, and so on, and the weights in those models are calculated from training data. While, the weights are estimated based on experts' opinions in knowledge-driven models. The knowledge-driven models include Fuzzy Logic, Dempster-Shafer Belief Theory and their weights are given with experts' opinions.

Weights of evidence model is used to predict a hypothesis about occurrence of an event based on combining known evidence in a study area where sufficient data are available to estimate the relative importance of each evidence by statistical methods. In the case of mineral resources assessment, the evidence consists of a set of evidential datasets (maps), and the model is used to predict the hypothesis about the occurrence of a given type of deposit in a study area. The weights are estimated from the measured association between known mineral occurrences and the values on the evidential maps. Based on combination of the evidential maps selected, the final result is extracted as a mineral potential map with a single index representing probability of occurrence of the given type of mineral deposit.

The model was originally developed for prediction of the probability that a new patient would be diagnosed to have a disease based on presence or absence of a set of symptoms (Kemp et al. 1999). Currently, it is still used in the medical fields (Zuyle. 2000; Johnson et al., 2001). The model has been applied in mineral resources assessment or geological survey with GIS together since the late 1980s (Agterberg, 1989,

---

[*] Corresponding author

Agterberg, et al. 1993; Bonham-carter, 1991;Bonham-carter, et al. 1989). For example, this method was applied to predict occurrence of gold deposits in Nova Scotia (Bonham-carter, et al. 1988; Cheng et al., 1999) and other places (Harris et al., 2000; Raines, 1999). This model is also used in evaluation on distribution of other resources, such as copper deposits (Tangestani et al., 2001), fossil packrat (Neotoma) middens (Mensing et al., 2000), and archaeology sites in a portion of the Central Valley of California (Hansen, 2000). Currently, weights of evidence model has been implemented in ArcView GIS (Arc-WofE) (Kemp et al., 1999 and 2001). Arc-WofE has been widely used in mineral exploration and environmental assessments (Hansen, 2000; Zuyle, 2000).

In this paper, a procedure for prediction of Sn-W-U deposits by using weights of evidence model with GIS together to integrate different evidential datasets, such as anomalies of chemical elements, radioactive anomalies and geological data, is described and its analysis result is evaluated in the paper. Implementation of data processing using Principle Component Analysis and data integration were conducted in ArcInfo GIS software.

2. WEIGHTS OF EVIDENCE MODEL

Weights of evidence model uses a probability framework based on set theory (Bonham-Carter, 1994). One of the important concepts used in this approach is the idea of prior and posterior probability. In the situation of mineral resource assessment, assume that a hypothesis (H) about the probability of mineral deposits of a given type D of in a study area is related to a set of evidence X ($X = \{X_1, X_2, ..., X_n\}$, and $X_1, X_2, ..., X_n$ are conditionally independent with each other).

The prior probability ($P\{D\}$) is the probability of occurrence of mineral deposit of type D without consideration of any known evidence information. The posterior probability ($P\{D \mid X\}$), also named conditional probability, is the probability of occurrence of mineral deposit of type D given the condition of X existing in the study area.

Another concept used here is odd ($O\{D\}$), which is defined as a ratio of the probability that an event will occur, such as $P\{D\}$, to the probability that it will not occur, such as $P\{\overline{D}\}$. It can be expressed as following Equation:

$$O\{D\} = \frac{P\{D\}}{P\{\overline{D}\}} \tag{2.1}$$

As to the hypothesis H on D with given X, its odd is:

$$O\{D \mid X\} = O\{D\} \frac{P\{D \mid X\}}{P\{\overline{D} \mid X\}} \tag{2.2}$$

In ordinary weights of evidence model, each evidence X_i in X need be divided into two subsets which can be expressed in terms of set theory: A (presence or true) and $\overline{A}$ (absence or false). The two subsets have following properties:

$$1)\ A \cup \overline{A} = X_i\ (A \text{ and } \overline{A} \text{ make up } X_i),\ i = 1,2,...,n \tag{2.3}$$

$$2)\ A \cap \overline{A} = 0\ (A \text{ and } \overline{A} \text{ are mutually exclusive.} \tag{2.4}$$

According to definitions and the properties of binary patterns introduced above, the following equations can be derived:

$$\begin{aligned} P\{D \mid X\} &= P\{D\} \frac{P\{X \mid D\}}{P\{X\}} \\ &= P\{D\} \frac{P\{X_1 \mid D\}}{P\{X_1\}} \frac{P\{X_2 \mid D\}}{P\{X_2\}} ... \frac{P\{X_n \mid D\}}{P\{X_n\}} \end{aligned} \tag{2.5}$$

$$\begin{aligned} O\{D \mid X\} &= O\{D\} \frac{P\{X \mid D\}}{P\{X \mid \overline{D}\}} \\ &= O\{D\} \frac{P\{X_1 \mid D\}}{P\{X_1 \mid \overline{D}\}} \frac{P\{X_2 \mid D\}}{P\{X_2 \mid \overline{D}\}} ... \frac{P\{X_n \mid D\}}{P\{X_n \mid \overline{D}\}} \end{aligned} \tag{2.6}$$

$$\begin{aligned} Log\{O\{D \mid X\}\} &= Log\{O\{D\}\} + Log \frac{P\{X \mid D\}}{P\{X \mid \overline{D}\}} \\ &= Log\{O\{D\}\} + Log\{\frac{P\{X_1 \mid D\}}{P\{X_1 \mid \overline{D}\}}\} + Log\{\frac{P\{X_2 \mid D\}}{P\{X_2 \mid \overline{D}\}}\} \\ &+ ... + Log\{\frac{P\{X_n \mid D\}}{P\{X_n \mid \overline{D}\}}\} \end{aligned} \tag{2.7}$$

If X_i is present, $\frac{P\{X_i \mid D\}}{P\{X_i \mid \overline{D}\}}$ is called sufficient ratio (LS). Otherwise, it is called necessity ratio (LN). Correspondingly, the natural logarithm of LS is the positive weight of evidence (W^+) and the natural logarithm of LN is the negative weight of evidence (W^-). For convenience, let W_0 denote the natural logarithm of $O\{D\}$. Thus, the Equation (2.6) and (2.7) can be written as:

$$O\{D \mid X\} = O\{D\} LS_1 (or LN_1) LS_2 (or LN_2) ..... LS_n (or LN_n) \tag{2.8}$$

$$\begin{aligned} Log\{O\{D \mid X\}\} &= W_0 + W^+{}_1 (or W_1^-) + W^+{}_2 (or W_2^-) \\ &+ ... + W^+{}_n (or W_n^-) = \sum_0^n W_i \end{aligned} \tag{2.9}$$

Then, the posterior probability $P\{D \mid X\}$ can be converted from the equation above as follows:

$$P\{D \mid X\} = \frac{\exp(Logit\{O\{D \mid X\}\})}{1+\exp(Logit\{O\{D \mid X\}\})} \qquad (2.10)$$

The values of $P\{D \mid X\}$ calculated with using Equation (2.7) are identical to those calculated directly using Equation (2.5). The advantage of using weights (2.7 or 2.9 and 2.10), instead of directly using the conditional probability expressions (2.5 or 2.8), is that the weights are easier to be interpreted than the probability factors.

In most applications of ordinary weights of evidence model, the contrast C is defined as Equation (2.11) and is used to select the cutoff to divide continuous variables into binary patterns.

$$C = W_i^+ - W_i^-, i = 1,2,\ldots,n \qquad (2.11)$$

3. PREDICTION OF SN-W-U MINERAL OCCURRENCES IN SOUTH MOUNTAIN BATHOLITH

Mineral resources assessment involves mapping the favourability of a type of mineral resources considered in a study area and ranking all sites in the study area according to their favourability. The mineral resources assessment is a decision-making process based on spatial information available to the study area. Generally, the assessment process is carried out with aid of GIS. This process can be summarized as three steps: database building, data processing and data integration modeling (Bonham-Carter, 1994). In this section, these three steps are described after introduction of characteristics of Sn-W-U deposits in the study area.

In this study, database building and evidential maps construction are carried out with aid of ArcInfo in terms of the geological and geochemical characteristics of Sn-W-U deposits in the study area. The evidential maps are integrated with weights of evidence model and the final potential map is generated and shown with posterior probability values of Sn-W-U mineral occurrences in the study area.

3.1 Characteristics of Sn-W-U deposits in the study area

Mineral deposits associated with granitoid rocks are divided into different categories based on their geological and geochemical characteristics. Sn-W-U deposits in the study area are categorized as "granophile-type", which often involves Sn, W and U mineralization (Strong, 1981). This type of deposits is genetically associated with quartz-rich leucocratic granitoids of later phase intrusions of a large granitoid bathlith. The hydrothermal alteration resulting in enrichment of Sn, W, and U elements in the deposits results in significant concentration of elements of Flourine (F), Lithium (Li), Niobium (Nb), Rubidium (Rb), Thorium (Th) and so on. Additionally, concentration of radiometric chemical elements, such as Th, U, and K, is associated with differential intrusion phases of the granitoid suite. Therefore, it is indirectly related to occurrence of Sn-W-U deposits.

The South Mountain Batholith in Nova Scotia is a granitoid batholith with area about 8000 km^2. The batholith consists of 13 plutons which can be divided into 2 formation stages, earlier stage and later stage (Xu, 2001). The later stage consists of monzogranite, leucomonzogranite and leucogranite, which are related to formation of Sn-W-U deposits in this study area. The petrographic and geochemical studies indicate that the South Mountain Batholith was formed by a continuous series from least evolved biotite granodiorite to most evolved leucogranite with a general incremental trend in trace elements Rb, U, Li, F, Sn, and W. Many economic geology studies indicate that lots of extensive and intensive hydrothermal fluid activities during the late- and post- magmatic period caused elements of Rb, U, Li, F, Sn, and W etc. to be further enriched and deposited in some geological sites spatially associated with small intrusive bodies of the later stage (O'Reilly et al. 1992; and Sangster, 1990) Numerous small deposits and mineral occurrences have been found in or around the South Mountain Batholith (Figure 1).

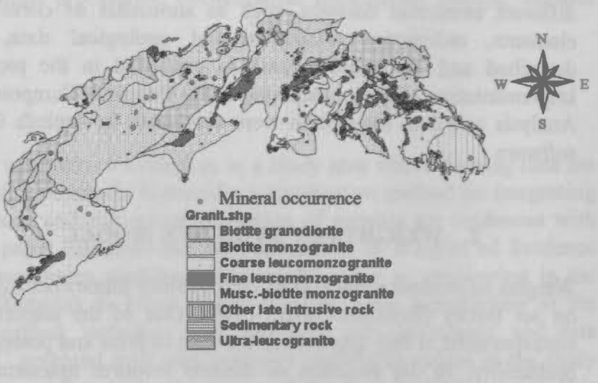


Figure 1. Geological Map of Sn-W-U mineral occurrence in the South Mountain Batholith in Nova Scotia, Canada

3.2 Datasets collected for the study

Data sets used in the current paper are mineral deposits and occurrence map, a geological map, two airborne radiometric anomaly maps, and 5 lake sediment geochemical anomaly maps (Xu, 2001).

The mineral occurrence map shows known Sn-W-U mineral deposits and occurrences in the study area. The geological map shows two granitoid phases in the study area. Later granitoid phases are favorite to occurrence of Sn-W-U mineral resources.

The two airborne radiometric anomaly maps are derived from the gamma-ray spectrometer datasets including images of eTh, eU and K, collected by Geological Survey of Canada from 1976 to 1983. According to previous research related to the study area, anomalies of the ratios eU/K and eU/eTh are associated with differentiation of granitoid phases (Xu, 2001), and they are also highly correlated with the evolved granitoid rock units, pervasive antometasomatic alteration, and granophile element mineralization (O'reilly et al. 1988).

The 5 lake sediment geochemical anomaly maps are derived from geochemical data of center-lake sediments. The anomaly maps include only anomalies of F, Li, Nb, Rb, and Th without W and Sn because the former 5 elements with relative high data quality are more suitable for spatial analysis than anomalies of W and Sn. Previous studies found that the relationship among the graniphile elements W, Sn, F, Li, Rb, Nb and radioactive element Th may reflect magmatic differentiation and hydrothermal solution activities (O'reilly et al. 1988).

Those data are used to delineate areas with significant late stage magmatic activity associated with Sn-W-U mineralization and alteration zones probably relative to Sn-W-U deposits.

3.3 Data Processing

According to the requirements on weights of evidence model, each evidential map can be binary pattern and conditionally independent each other in order to be integrated with weights of evidence model. Therefore, the collected data need to be preprocessed prior being integrated with Weights of evidence model.

It is found that most the 5 geochemical anomaly maps of lake sediments are correlated each other as shown in the correlation matrix of them (Table 1). The correlation coefficient between the two radiometric anomaly (eU/K and eU/eTh) maps is 0.24. To some extent, the two radiometric anomaly maps are correlated each other. Principle Component Analysis (PCA) method is respectively used to process the five geochemical anomaly maps and the two radiometric anomaly maps in order to get independent evidences. The first two principle components and the two principle components respectively generated from PCA of the five geochemical anomaly maps and the two radiometric anomaly maps are extracted as four evidential maps.

C	Li	F	Rb	Th	Nb
Li	1.00000	0.58701	0.57749	0.58250	0.22037
F	0.58701	1.00000	0.55676	0.57463	0.15430
Rb	0.57749	0.55676	1.00000	0.57285	0.14198
Th	0.58250	0.57463	0.57285	1.00000	0.13081
Nb	0.22037	0.15430	0.14198	0.13081	1.00000

Table 1 Correlation (C) matrix of the five geochemical anomaly maps

Suitable cutoffs for five evidential maps including a granitoid phase map and four principle component maps need to be set up in order to convert the five maps into binary patterns. After calculation of weights and studentized version of contrast C (Student (c)), which is equal to a ratio between contrast of a value and standard deviation of contrast of the value, of each value in each map, a value with the highest student(c) value is respectively used as cutoffs for each of the five evidential maps and convert them into binary patterns. For example, the value with the highest student(c) in the first geochemical anomaly component map is 12 (Figure 2), then favourite values in this map greater than or equal to 12 are converted into 1 and other values are converted into 0.

Additionally, distance to a contact between favorite granitoid phase, and the other phases is another useful evidence spatially related to Sn-W-U mineralization based on geologists' knowledge. A map representing the distance is generated by using "buffer" function in ArcInfo. Similarly, the distance map is converted in to binary in terms of a value with the highest student(c) value.

Pairwise test of conditional independence is carried out on the 6 binary evidential maps with use of Arc-WofE (Kemp et al. 1999). The results of pairwise Chi-square tests (significant at 95% level of probability, for which tabled $\chi^2=3.8$) are shown in Table 2. Table 2 shows that the two binary maps, representing the granitoid phases and the distance, are not conditionally independent each other. Therefore, the granitoid phase map is removed and the distance map is left based on their relative importance to prediction of Sn-W-U mineral occurrence.

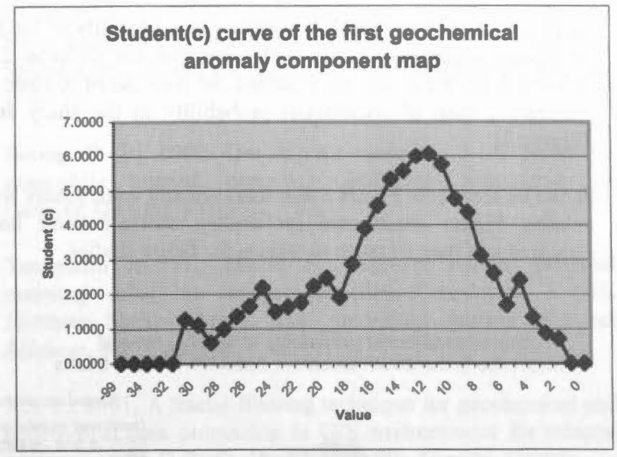


Figure 2. Student(c) curve of the first geochemical anomaly component map

Evidence	Radpc1	granbinw	ctactnw	chmpc2	chmpc1
radpc2	1.16	1.76	2.30	0.39	3.34
radpc1		0.95	1.74	0.16	1.33
granbinw			36.90	1.00	2.28
ctactnw				1.27	3.64
chmpc2					1.54

Where,

- granbinw is the binary map representing granitoid phases;
- ctactnw is the binary map representing the distance to the contact;
- chmpc1 is the binary map representing the first principle component from the 5 geochemical anomaly maps;
- chmpc2 is the binary map representing the second principle component from the 5 geochemical anomaly maps;
- radpc1 is the binary map representing the first principle component from the 2 radiometric anomaly maps;
- radpc2 is the binary map representing the second principle component from the 2 radiometric anomaly maps.

Table 2 Table of Chi-squared Statistics (significant at 95% level of probability, for which tebaled $\chi^2=3.8$)

3.4 Data integration with weights of evidence

The Sn-W-U occurrence map and the final 5 binary evidential maps including the distance map and four principle component maps are applied to calculate conditional probability of Sn-W-U mineral occurrence in the study area with the formulas 2.1, 2.9 and 2.10. It can be easily carried out with use of ArcInfo to generate a map of conditional probability in the study area (Figure 3).

It can be seen from Figure 3 that there are still some places with relative higher conditional probability which haven't been explored and may be taken as targets for future studies.

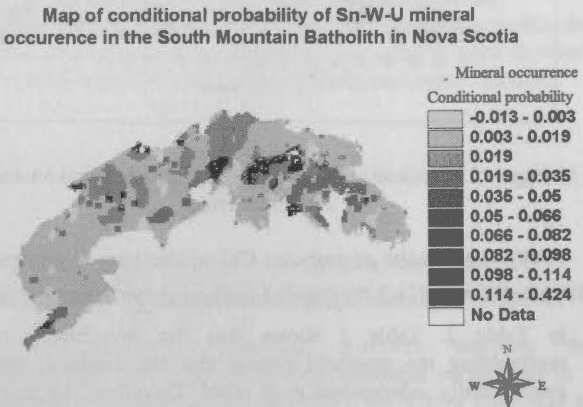


Figure 3. Map of conditional probability of Sn-W-U mineral occurrence in the South Mountain Batholith in Nova Scotia, Canada

4. DISCUSSIONS AND CONCLUSIONS

Generally, in case of mineral resources, evaluation of a prediction result is based on predictive ability of the prediction result. The best validation method is to see if the prediction result may be identical with newly discovered deposits. However, this is also the most difficult way to be executed and it takes much time and much money to test the result by this way.

Some practical evaluation methods are suggested by Raines (1999) as follows: 1) to test if the weights make sense; 2) to test if the known deposits are in the areas with high favourability; and 3) to compare the prediction results with the ones from other methods and to test if they agree with each other. In this paper, the first two methods are used to evaluate the prediction result.

It can be obviously seen from Figure 3 that most mineral occurrences are within areas with relative high probability. Table 3 summaries the weights, contrast, and student(c) results calculated for weights of evidence model. All of the five evidential maps has relative high contrast (greater than .5) and student(c) values (greater than 1), and they are conditionally independent each other at significant 95% level (Table 2). So, they are eligible to be applied in weights of evidence model.

The evidential map representing distance to the contact is the best evidence in terms of contrast and student(c) values. This evidence has highest contrast and student(c) values in the all 5 evidential maps. The three evidential maps, respectively representing the distance to the contact and the first two principle components of the 5 geochemical anomaly maps, have relative higher positive weights and student(c) values, therefore, their present patterns are useful to recognize sites where Sn-W-U mineral resources may occur. The two evidential maps, representing the two principle components of the two radiometric anomaly maps, have negative weights lower than those of the other three maps and positive weights close to zero, therefore, their present patterns are useful to recognize sites where Sn-W-U mineral resources may not occur.

Evidence	Class_field	W-	W+	Contrast	Student(c)
ctactnw	Value	-0.5526	1.2130	1.7656	9.2102
chmpc1	Value	-0.3829	0.7054	1.0883	6.0686
chmpc2	Value	-0.0203	1.7085	1.7288	2.9183
radpc2	Value	-0.4939	0.0856	0.5795	2.0312
radpc1	Value	-0.7474	0.0587	0.8061	1.9154

Table 3 Summary of evidential maps defined by weights of evidence model

Based on results of these two evaluation methods, it can be concluded that the prediction result with use of weights of evidence is successful and weights of evidence model is useful to predict Sn-W-U occurrence in the study area.

Weights of evidence model has both advantages and disadvantages. For example, this method is data-driven and can avoid subjective choice of evidences and subjective estimation of weights for evidences. In the other hand, with use of binary patterns (as a common case), some useful information in the continuous evidences may be lost after data conversion. Cheng and Agterberg (1999) developed Fuzzy weights of evidence method, based on a fuzzy membership function and weights of evidence method, in order to minimize the lost information from data conversion.

REFERENCE

Agterberg, F. P., 1989. Computer programs for mineral exploration. *Science*, 245, pp. 76 – 81.

Agterberg, F. P., Bonham-Carter, G. F., Cheng Q., and Wright, D. F., 1993. Weights of evidence modeling and weighted logistic regression for mineral potential mapping. In: Davis, J. C., and Herzfeld, U. C., eds., *Computers in Geology*. Oxford University Press, New York, pp. 13 - 32.

Bonham-Carter, G. F., 1991, Integration of geoscience data using GIS. In Maguire, D. J., Goodchild, M. F. and Rhind, D. W., eds., *Geographical Information Systems*, Vol. 2: Applications, Longmans Scientific and Technical, pp. 171 – 184.

Bonham-Carter, G. F., 1994. *Geographic Information System for Geosciences: Modelling with GIS.* Pergamon Press, Oxford.

Bonham-carter, G. F., Agterberg, F. P. and Wright, D. F., 1988. Integration of geological datasets for gold exploration in Nova Scotia. *Photogrammetric Engineering and Remote Sensing*, 54(11), pp. 1585 – 1592.

Bonham-Carter, G. F., Agterberg, F. P. and Wright, D. F., 1989. Weights of evidence modeling: a new approach to mapping mineral potential. In: Agterberg, F. P. and Bonham-carter, G. F., eds, *Statistical Applications in the Earth Sciences*, Geological Survey of Canada, Paper 89-9, pp. 171 - 183.

Cheng, Q. and Agterberg, F. P., 1999. Fuzzy weights of evidence method and its application in mineral potential mapping. *Natural Resources Research*, 8(1), pp. 27–35.

Harris J. R., Wilkinson L., Grunsky E. C., 2000. Effective use and interpretation of lithogeochemical data in regional mineral exploration programs: application of Geographic Information Systems (GIS) technology. *Ore geology reviews*, 16(3-4), pp. 107-143.

Hansen, D. T., 2000. Describing GIS Applications: Spatial Statistics and the Weights of Evidence Extension in the Analysis of the Distribution of Archaeology Sites in the Landscape. In: *Proceedings of the Twentieth Annual ESRI User Conference,* San Diego, United States. http://www.esri.com/library/userconf/proc00/professional/papers/PAP174/P174.htm (Accessed on Feb. 25, 2002)

Johnson C. J. D., Pynsent P. B., Grimer R. J., 2001. Clinical features of soft tissue sarcomas. *Annals of the royal college of surgeons of England*, 83(3), pp. 203 - 205.

Kemp, L.D., Bonham-Carter, G.F. and Raines, G.L., 1999. Arc-WofE: Arcview extension for weights of evidence mapping. Ottawa, Canada. http://ntserv.gis.nrcan.gc.ca/wofe (Accessed on Feb. 25,2002).

Kemp, L.D., Bonham-Carter, G.F., Raines, G.L. and Looney, C.G., 2001. Arc-SDM: Arcview extension for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural network analysis. Ottawa, Canada. http://ntserv.gis.nrcan.gc.ca/sdm/ (Accessed on Feb. 25, 2002).

Mensing S. A., Elston R. G., Raines G. L., Tausch R. J., Nowak C. L., 2000. A GIS model to predict the location of fossil packrat (Neotoma) middens in central Nevada. *Western North American Naturalist*, 60(2), pp. 111-120.

Malczewski, J., 1999. *GIS and Multicriteria Decision Analysis*. John Wiley & Sons, New York.

O'Reilly G. A., Corey, M. C. and Ford, K. L., 1988. The role of airborne gamma-ray spectrometry in bedrock mapping and mineral exploration: case studies from granitic rocks within the Meguma zone, Nova Scotia. *Maritime Sediments and Atlantic Geology*, 24(1), pp. 47-60.

O'Reilly, G. A. and Kontak, D. J., 1992. The East Kemptville tin deposit and the southwest Nova Scotia tin domain. In: *Mines and Mineral Branch*, Report of Activities 1991, D. R. MacDonald and Y. Brown eds., 92-1, pp. 65-74.

Raines, G. L., 1999, Evaluation of weights of evidence to predict epithermal-gold deposits in the Great Basin of the Western United States. *Natural Resources Research*, 8(4), pp. 257 – 276.

Sangster, A. L., 1990. Metallogeny of the Meguma Terrain, Nova Scotia. In: *Mineral Deposit Studies in Nova Scotia*, A. L. Sangster ed., Geological Survey of Canada, Paper 90-8, 1, pp. 115 - 152.

Strong, D. F., 1981. Ore deposit models – 5. A model for granophile mineral deposits. *Geological Association of Canada*, 8(4), pp. 155-162.

Tangestani M. H., Moore F. Porphyry copper potential mapping using the weights-of-evidence model in a GIS, Northern Shahr-e-Babak, Iran. *Australian Journal of Earth Sciences*, 48(5), pp. 695 – 702.

Xu, Y., 2001, A fractal filtering technique for geochemical and geophysical data processing in GIS environments for mineral exploration. Ph.D thesis, York University, Toronto, Canada.

Zuyle. P. V., 2000, Using Weights of Evidence (WofE) in a GIS for the Prediction of Tuberculosis Cases: Application of a New ArcView Extension. In: *Proceedings of the Twentieth Annual ESRI User Conference,* San Diego, United States. http://campus.esri.com/campus/library/bibliography/RecordDetail.cfm?ID=8958&CFID=363225&CFTOKEN=65560048(Accessed on Feb. 25, 2002)

FUZZY LOGIC INTEGRATION FOR LANDSLIDE HAZARD MAPPING USING SPATIAL DATA FROM BOEUN, KOREA

Kwang-Hoon Chi[a], No-Wook Park[a,*], Chang-Jo Chung[b]

a National Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources,
Daejon 305-350, Korea - nwpark@kigam.re.kr
b Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario K1A 0E8, Canada
- chung@gsc.nrcan.gc.ca

KEY WORDS: Fuzzy logic, Data Integration, Validation, Future Landslides

ABSTRACT:

Since the early 1990s, the mathematical models have been developed and applied to landslide hazard mapping using GIS. Among various models, this paper discusses the effectiveness of fuzzy set theory for landslide hazard mapping. In this study, we collected several data sets related to landslide occurrences in Boeun, Korea, and then digitally represented as the fuzzy membership functions. To integrate them, fuzzy inference networks by using a variety of different fuzzy operators, especially combination of fuzzy OR and fuzzy γ operator, are designed, and experiments are carried out. Owing to the cross-validation based on the spatial portioning of the landslide distribution, we could quantitatively compare with various fuzzy inference networks designed for the influence of choice of γ value. The results show that the fuzzy set theory can integrate effectively various spatial data for landslide hazard mapping, and it is expected that some suggestions in this study are helpful to further real applications including integration, and interpretation stages in order to obtain a decision-supporting layer.

1. INTRODUCTION

Landslides cause extensive damage to property and occasionally result in loss of life throughout most of country. So it is necessary to delineate the area that will be likely to be affected by the future landslides. For landslide susceptibility analysis, a unified and general framework has been proposed and termed Favourability Function by Chung and Fabbri (1993). FF models can be based on probability, evidential reasoning or fuzzy set theory, depending on the quantitative relationships between input causal factors and the past landslides. These approaches with their own mathematical backgrounds have provided powerful schemes for decision-supporting information, through several case studies (Van Westen, 1993; Chung and Fabbri, 1998, 1999; Carrara *et al*., 1998; Jibson *et al*., 1998; Lee and Min, 2001). Conventional probabilistic approaches implicitly assume that most of the information on which decision-making is based is probabilistic in nature, and that precise probability judgements can be formulated for each hypothesis of the problem concerned. On the other hand, in terms of soft computing, uncertainty may have different nature and should be modelled in different frameworks, and a hard decision should be drawn only towards the end of the processing (Binaghi *et al*., 1998). Especially, fuzzy set theory can provide us with a natural method of quantitatively processing multiple data sets and many scientists have applied the fuzzy set theory to their studies and proved that this theory is very useful to reflect natural phenomena or irregular behaviros (Zadeh, 1965; An *et al*., 1991; Chung and Fabbri, 1993; Zimmermann, 1996).

In this paper, we apply and investigate the fuzzy logic information representation and integration for landslide hazard mapping. First, we construct the input causal factors related to landslide occurrences, and then assignment of fuzzy membership functions is followed. To integrate fuzzy membership functions, we construct "fuzzy inference network" by using various fuzzy operators. As an essential part for landslide hazard mapping, in order to validate the significance of the prediction results, we exemplify whether and to what extent a prediction can be extended, in space, to neighbouring areas with similar geology. A case study from Boeun, Korea is carried out to illustrate above schemes.

2. STUDY AREA AND DATA SETS

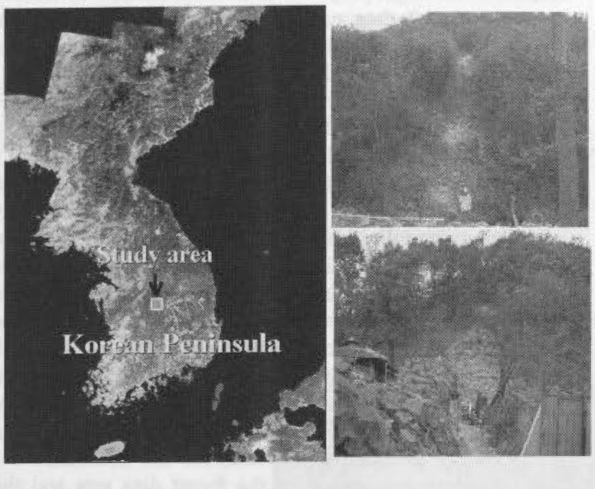
The Boeun area in Korea, which had much landslide damage following heavy rain in 1998, was selected as the test area (Figure. 1). A two-day intensive rainfall between August 11 and 12, 1998 had induced many landslides in the study area. Landslides usually induced due to rainfall, local downpour, earthquakes and volcanic activities. Landslides triggered by heavy rainfall are the most common throughout Korea. Landslides are usually categorized into falls, topples, slides, spreads, and flows. Shallow landslides occur in material defined as engineering soils: unconsolidated, inorganic mineral, residual, or transported material (colluviums or alluviums) including rock fragments (Varnes, 1978). In the study area, the landslides were mainly debris flows that occurred during 3–4 hours of high intensity rainfall, or shortly after (Kim *et al*., 2000).

The input data for a test consist of several layers of map information (Table 1). The slope and aspect were calculated from the 1: 5,000 scale DEM. As for the soil data sets, the texture, topography, drainage, material, and thickness of soil were acquired from 1:25,000 scale soil maps. As for the forest data sets, the type, diameter, age, and density of timber were acquired from 1:25,000 scale forest maps. The lithology map was obtained from 1:50,000 scale geological map. After pre-processing, all data sets were built on a cell-based database, and the whole study area consists of 1,879 X 1,444 pixels (= 2,713,276 pixels), covering approximately 68km^2. Each pixel corresponds to a 5m by 5m area on the ground.

---

* Corresponding author.

The aerial photographs taken in 1996 and 1999 were used to detect landslide locations, and the locations were verified by fieldwork. In total, 375 debris flows type landslides were mapped. The target pattern, with the entire landslide bodies, consists of two separate and distinct sub-areas, the scarp area and the deposit area. The geomorphologic characteristics of these two sub-areas are distinctly different. In this study, the topographically highest 20% of the scars of the landslides are considered as trigger areas.



(a) (b)

Figure 1. (a) The location map of study area, (b) photos showing one of landslide scars occurred in this area

Data types	
Topographic data sets	Slope
	Aspect
Forest data sets	Forest type
	Forest diameter
	Forest age
	Forest density
Soil data sets	Soil texture
	Soil topography
	Soil drainage
	Soil material
	Soil thickness
Lithology	
Landslides location map	

Table 1. Data sets used in this study

3. FUZZY INFORMATION REPRESENTATION

3.1 Fuzzy membership function

The fuzzy set theory was introduced by Zadeh (1965), which facilitates analysis of non-discrete natural processes or phenomena as mathematical formulae (Zimmermann, 1996).

If X=$\{x\}$ denotes a universe of the attribute values, the fuzzy set A in the X is the set of ordered pairs

$$A = \{x,\ \mu_A(x)\},\ x \in X \tag{1}$$

$\mu_A(x)$ is known as grade of membership of x in the A. Usually, $\mu_A(x)$ is an integer or a floating number in the range [0,1] with 1 representing full membership and 0 non-membership. The grade of membership reflects a kind of ordering that is not based on probability but on admitted possibility. The value of $\mu_A(x)$ for the attribute value x in A can be interpreted as the degree of compatibility of the predicate associated with set A and attribute value x.

Fuzzy membership functions closely associated with semantic analysis can be determined either normatively or empirically. The derivation of membership functions is crucial in fuzzy information processing and the lack of simple and generally acceptable methods to build membership functions may cause it less favourably with other information processing methods. Despite the lack of scientific foundation for membership functions, many fuzzy systems have demonstrated satisfactory performance when compared with two-valued logic system composed of crisp set theory (Zimmermann, 1996). The normative approach is commonly used for deriving membership functions for linguistic values because impreciseness inherent to these values is subjective. However, this approach was basically designed for engineering applications. So it is judged not suitable for geoscientific applications such as mineral potential mapping and landslide hazard mapping. For integration of multiple geological data sets, An *et al.*(1991) assigned the fuzzy membership functions using empirical procedure based on expert's opinion and the mineral deposit model. Most of studies for mineral potential mapping, assignment of fuzzy membership functions are based on expert's opinion. Meanwhile, for landslide hazard mapping, Chung and Fabbri (2001) assigned the fuzzy membership functions using the relationships between input causal factors and known past landslides.

3.2 Assignment of fuzzy membership function

In this study, assignment of fuzzy membership functions to each data layer followed Chung and Fabbri (2001)'s approach. Our target proposition is "a pixel p in the study area will be affected by future debris flow type landslides".

First, we investigated the relationships between input causal factors and past landslides. For this, the likelihood ratio function of each map, which can highlight the difference between areas affected by past landslides and areas not affected by past landslides, was calculated and compared with each other. In slope map, the steeper the slope, the greater the landslide possibility. Most landslides occurred between 15° and 35°. The slope angle is an essential component of landslide susceptibility. In general, it is expected that low slope angles have a low possibility of landslides due to lower shear stresses associated with low gradients. Steep natural slopes, however, may not be susceptible to shallow landslides (Lee and Min, 2001). In aspect map, the landslide occurrence possibility value was similar at all directions. In forest maps, the possibility of landslide occurrence is higher in larch and artificial Chestnut trees, very small diameters, younger timber, and loose density forest. These results are related to location of forest and amount of roots. In soil maps, the possibility of landslide occurrence is higher in well-drained soil, red-yellow podzolic soils and lithosols, acidic rocks residuum, mountainous areas, thick soils. These results are related to increase of unit weight and shear stress of soil due to pore-water increase. In lithology map, most landslides had occurred in biotite granite areas.

The definition of membership functions for input causal factors was performed by using above relationships between landslides and input causal factors based on the likelihood ratio functions and slightly modified by considering expert's opinion.

4. DATA INTEGRATION

4.1 Fuzzy operators

Each data layer of target information denoted from fuzzy theory can now be integrated by using fuzzy operators. When two membership functions $\mu_A(x)$ and $\mu_B(x)$ are combined, Some of the useful fuzzy set operators are as follows (An *et al.*, 1991; Chung and Fabbri, 2001):

1. Fuzzy OR

$$\mu_{OR}(x) = \text{MAX}\,[\,\mu_A(x),\ \mu_B(x)\,] \qquad (2)$$

2. Fuzzy AND

$$\mu_{AND}(x) = \text{MIN}\,[\,\mu_A(x),\ \mu_B(x)\,] \qquad (3)$$

3. Fuzzy Algebraic Sum

$$\mu_{SUM}(x) = 1 - \prod_{i=1}^{2} \mu_i(x) \qquad (4)$$

4. Fuzzy Algebraic Product

$$\mu_{PRODUCT}(x) = \prod_{i=1}^{2} \mu_i(x) \qquad (5)$$

5. Fuzzy γ operator

$$\mu_\gamma(x) = [\mu_{SUM}(x)]^{\gamma} \times [\mu_{PRODUCT}(x)]^{1-\gamma} \qquad (6)$$

When the fuzzy OR and AND operators are used, only one of the contributing fuzzy sets has an effect on the resultant value. The fuzzy algebraic sum and algebraic product operators make the resultant set larger than, or equal to the maximum value and smaller than, or equal to the minimum value among all fuzzy sets, respectively.

Meanwhile, the resultant set that is combined by the fuzzy γ operator has the value between that of the fuzzy algebraic product operator and that of the fuzzy algebraic sum operator. The determination of optimum value is closely associated with degree of compensation between the two extreme confidence levels. In cases of $\gamma = 1$ (full compensation) or $\gamma = 0$ (no compensation), these operators with different values are equivalent to algebraic sum operator or production sum operator, respectively. Therefore, the choice of can produce the resultant value that can ensure a flexible compromise (Figure 2).

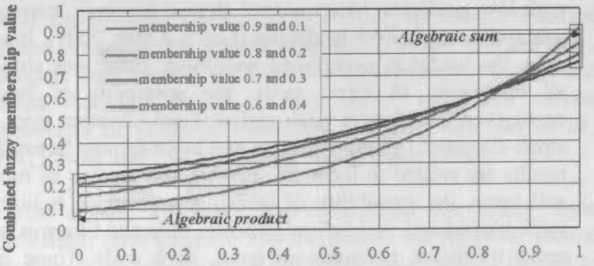


Figure 2. A graphical description of the resultant fuzzy membership value obtained by combining two fuzzy membership functions using fuzzy γ operator

4.2 Design of fuzzy inference network

Though several authors have recently used fuzzy logic approaches for integration of multiple data sets, selecting an optimum fuzzy operator has always been a difficult task. It has become apparent that fuzzy operators depend very much on the types of spatial data to be integrated (Choi *et al.*, 2000). In geoscientific applications of spatial data integration, spatial data have, in most cases, varying degree of information content with respect to the target proposition. In these cases, it is necessary to combine spatial data using several different fuzzy operators separately or a combination of selected operators depending on the characteristics of each data layer (Moon, 1998).

In this study, instead of using one operator, we constructed "fuzzy inference network" by using a variety of different fuzzy operators. We combined the fuzzy membership functions using the intermediate fuzzy information representation and various fuzzy operators (Figure 3). The intermediate fuzzy information representation is divided into three parts; topographic data sets, forest data sets, and soil data sets. The fuzzy OR operator was used in order to combine the topographic data sets including slope and aspect. The relationship between slop and aspect is not fully known and one may be considered as a primary information and the other is a secondary information. So we combined these maps using OR operators. When fuzzy OR operator is used, only one of the contributing fuzzy sets has an effect on the resultant set. While, for the intermediate fuzzy information representation of the forest data sets and the soil data sets, γ operator was used to integrate them. Forest data sets and soil data sets have the typical characteristics wit respect to landslide occurrences. So certain classes of each map have a positive potential for landslide occurrences. However, the relationships among forest data sets and soil data sets are very complicated, so we would not expect that certain type of data have higher possibility than others. To consider this, fuzzy γ operator, which can make it possible that all the contributing fuzzy sets have an effect on the resultant set, was used. Finally, to integrate three intermediate fuzzy information and lithology data, γ operator is experimented.

We divided several fuzzy inference networks into 4 classes, depending on the choice of γ value. Class 1 is one that all high γ values were used for fuzzy intermediate representation and integration. In class 2, high γ values were used for fuzzy intermediate representation of forest and soil data, and low or middle γ values were tested for final integration. Class 3 is the opposite to class 2, that is, low or middle γ values, and high γ values were tested for fuzzy intermediate representation and integration, respectively. In class 4, all low γ values were used.

Through above procedures, we prepared some prediction maps. To visualize the prediction maps, we used rank order statistics. We first computed the score for each pixel and then sort all scores by increasing order to determine the ranks of the scores. The pixel that has the smallest score (the smallest prediction value) has rank one, and the pixel that has the maximum score has the maximum rank. Then the ranks are normalized so that the maximum value is 1 or 100%, and the normalized values are termed the favourability indices or simply indices. The pixel with the index 100% had the largest score of the prediction function. If the pixels have index, 99.5%, it means that the ranks of their function scores are within the top 0.5% (99.5% - 100%)

in the study area. These indices over the study area constitute a landslide susceptibility map.

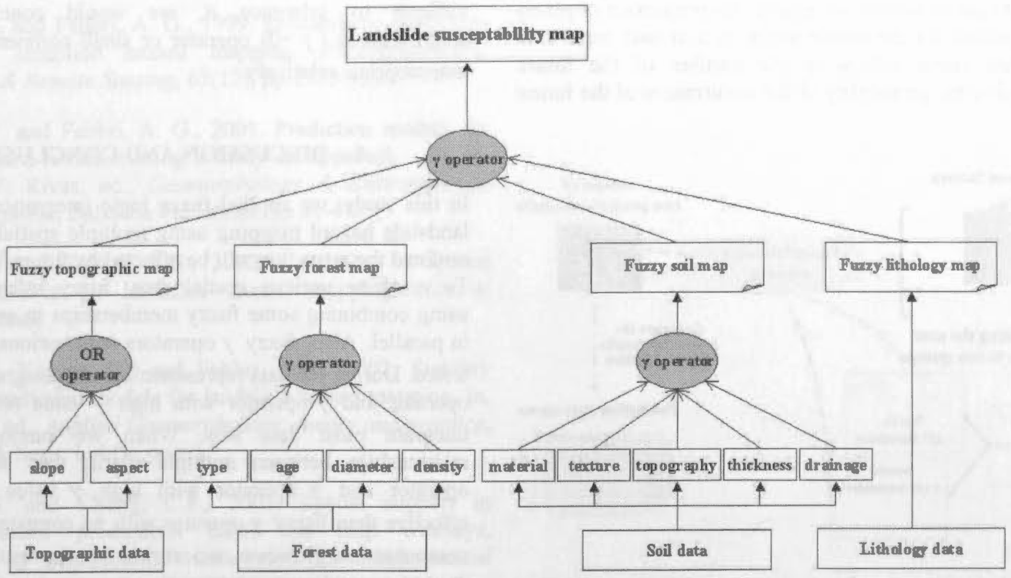

Landslide susceptability map
γ operator
Fuzzy topographic map
Fuzzy forest map
Fuzzy soil map
Fuzzy lithology map
OR operator
γ operator
γ operator
slope
aspect
type
age
diameter
density
material
texture
topography
thickness
drainage
Topographic data
Forest data
Soil data
Lithology data

Figure 3. Fuzzy inference network designed in this study

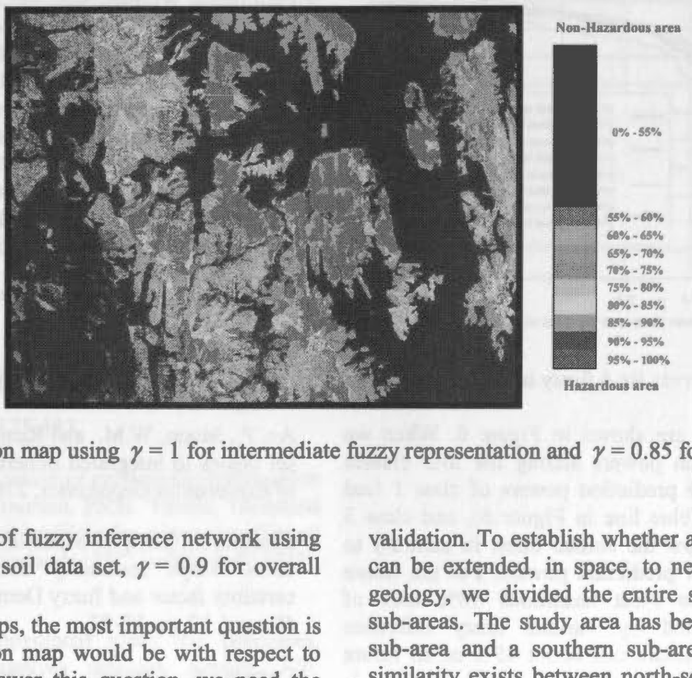

Non-Hazardous area
0% - 55%
55% - 60%
60% - 65%
65% - 70%
70% - 75%
75% - 80%
80% - 85%
85% - 90%
90% - 95%
95% - 100%
Hazardous area

Figure 4. Prediction map using $\gamma = 1$ for intermediate fuzzy representation and $\gamma = 0.85$ for final integration

Figure 4 shows one example of fuzzy inference network using $\gamma = 1$ for forest data set and soil data set, $\gamma = 0.9$ for overall combination.

After we get the prediction maps, the most important question is "how successful this prediction map would be with respect to the future landslides". To answer this question, we need the information interpretable with respect to the future event. It leads to the next essential step of validation.

5. VALIDATION OF PREDICTION RESULT

The critical strategy in prediction models is the task of validating the prediction results, so that the prediction results can provide meaningful interpretation with respect to the future landslides (Fabbri and Chung, 2001; Chung and Fabbri, 2002). To carry out the validation, we must restrict the use of all the data of the past landslides in the study area. By partitioning the data, one subset is used for obtaining a prediction map; the other subset is compared with the prediction results for validation. To establish whether and to what extent a prediction can be extended, in space, to neighbouring areas with similar geology, we divided the entire study area into two separated sub-areas. The study area has been subdivided into a northern sub-area and a southern sub-area. This was because greater similarity exists between north-south than east-west sub-areas. We selected one of two sub-areas to construct a prediction model and the other to validate the prediction. Through this validation procedure, we can assess the prediction powers of various fuzzy inference networks, and compare with them quantitatively.

The space-partition technique used in this study consisted of the following steps (Figure 5):

- The 237 scarps distributed in the north sub-area were used to compute the south sub-area fuzzy inference networks.
- Similarly, the 138 scarps in the south sub-area were used to compute the north sub-area fuzzy inference networks.
- Then we assembled them into a mosaic of the two representations.

In order to validate a mosaic prediction map, we computed the prediction rate curve, which can explain the proportion of pixels correctly classified for the whole scarps in a mosaic map. This prediction rate curve relates to the number of the future landslides and to the probability of the occurrences of the future landslides.

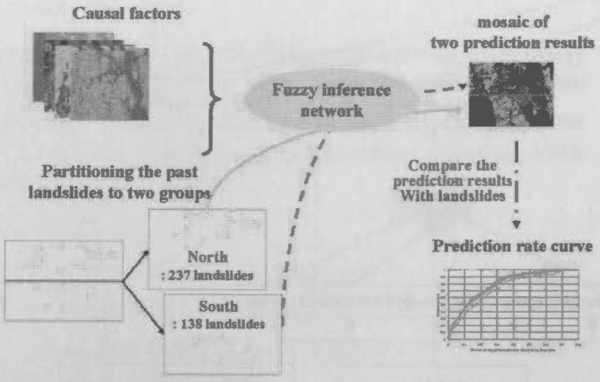

Causal factors
Fuzzy inference network
mosaic of two prediction results
Partitioning the past landslides to two groups
North : 237 landslides
South : 138 landslides
Compare the prediction results With landslides
Prediction rate curve

Figure 5. A graphical representation of cross-validation approach in this study

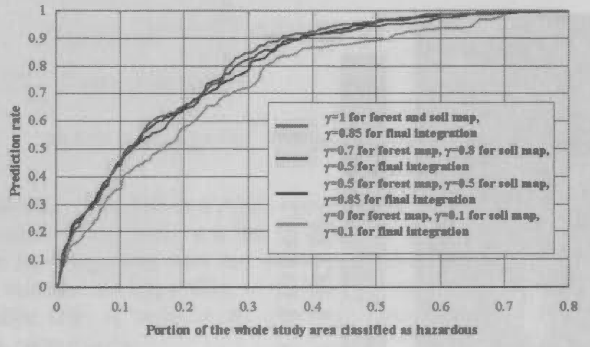


Figure 6. Prediction rate curves for 4 fuzzy inference networks

The prediction rate curves are shown in Figure 6. When we compare with the prediction powers among the four classes mentioned in section 4, the prediction powers of class 1 (red line in Figure 6), class 2 (blue line in Figure 6), and class 3 (black line in Figure 6) show the similar ones. In contrary to those, class 4 shows poorer prediction powers. For the future landslides, if we take the most hazardous 10% area of prediction images generated by various fuzzy inference networks, then we may estimate that about 45% of all future landslides will be located in the delineated area of class 1, class2, and class 3. The prediction rate curves are shown in Figure 6. When we compare with the prediction powers among the four classes mentioned in section 4, the prediction powers of class 1 (red line in Figure 6), class 2 (blue line in Figure 6), and class 3 (black line in Figure 6) show the similar ones. In contrary to those, class 4 shows poorer prediction powers. For the future landslides, if we take the most hazardous 10% area of prediction images generated by various fuzzy inference networks, then we may estimate that about 45% of all future landslides will be located in the delineated area of class 1, class2, and class 3. However, in case of class 4, about 35% of all future landslides will be located. In our study area, if we choose the relatively higher γ value for integration, we may say that the results are effective and the fine difference of γ value does not affect the final prediction results. When the relationship between data sets is not fully known and it is difficult to inference it, we would conclude that no compensation (γ=0) operator or small compensation may be inappropriate relatively.

6. DISCUSSION AND CONCLUSION

In this study, we applied fuzzy logic integration appraoch for landslide hazard mapping using multiple spatial data sets, and outlined the areas that will be affected by future landslides.
To combine various spatial data, fuzzy inference networks using combining some fuzzy memberships in series and others in parallel. Also, fuzzy γ operators with various γ values were tested. During the data representation and integration, fuzzy OR operator and γ operator with high γ value could effectively integrate most data sets. When we cannot be sure of relationships between multiple spatial data sets, fuzzy OR operator and γ operator with high γ value can be more effective than fuzzy γ operator with no compensation or small compensation. However, we remind that the results in this study are not general ones, so extensive experiments should be made in several study areas to strengthen the situation here identified.
To assess quantitatively the prediction powers of various fuzzy inference networks, cross-validation approach was also performed. With the help of cross-validation approach, we can evaluate the prediction results quantitatively, and compare with models. Without this kind of the cross-validation technique, prediction maps cannot be evaluated.
For the future works, several aspects still need to be considered. For any prediction models to generate reasonably "good or significant" results, the prediction result should be robust and stable (Chung *et al*., 2002). For this, we are currently evaluating the stability analysis using matching rate function. In addition, we will try to involve the fuzziness of boundaries in categorical maps such as forest, soil, and lithology maps, in data representation stage.

REFERENCES

An, P., Moon, W.M. and Rencz, A., 1991. Application of fuzzy set theory to integrated mineral exploration. *Canadian Journal of Exploration Geophysics*, 27(1), pp.1-11.

Binaghi, E., Luzi, L., Madella, P., Pergalani, F. and A. Rampini, 1998. Slope instability zonation: a comparison between certainty factor and fuzzy Dempster-Shafer approaches. *Natural Hazard*, 17, pp.77-97.

Carrara, A., Guzzetti, F., Cardinali, M. and Reichenbach, P., 1999. Use of GIS technology in the prediction and monitoring of landslide hazard. *Natural Hazards*, 20, pp.117-135.

Choi, S.W., Moon, W.M. and Choi, S.G., 2000. Fuzzy logic fusion of W-Mo exploration data from Seobyeog-ri, Korea. *Geosciences Journal*, 4(2), pp.43-52.

Chung, C. F. and Fabbri, A. G., 1993. The representation of geoscience information for data integration. *Nonrenewable Resources*, 2(2), pp.122-139.

Chung, C.F. and Fabbri, A. G., 1998. Three Bayesian prediction models for landslide hazard. In, A. Bucciantti, ed., *Proceedings of International Association for Mathematical Geology Annual*

Meeting (IAMG'98), Ischia, Italy, October 3-7, 1998, pp.204-211.

Chung, C. F. and Fabbri, A. G., 1999. Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Engineering & Remote Sensing*, 65(12), pp.1389-1399.

Chung, C. F. and Fabbri, A. G., 2001. Prediction models for landslide hazard zonation using a fuzzy set approach. . In, M. Marchetti, V. Rivas, ed., *Geomorphology & Environmental Impact Assessment*, Balkema Publishers, pp.31-47.

Chung, C.F. and Fabbri, A.G., 2002. Validation of spatial prediction models for landslide hazard mapping. *Natural Hazard* (in press).

Chung, C.F., Kojima, H., and Fabbri, A.G., 2002. Stability analysis of prediction models for landslide hazard mapping. In, R.J. Allison, ed., *Applied Geomorphology: theory and practice*, John Wiley & Sons, Ltd (in press).

Fabbri, A.G. and Chung, C.F., 2001. Spatial support in landslide hazard predictions based on map overlays, *Proceedings of International Association for Mathematical Geology Annual Meeting (IAMG'2001)*, Cancun, Mexico, September 10-12, 2001, CDROM.

Jibson, R.W., Harp, E.L., and Michael, J.A., 1998. A method to produce probabilistic seismic landslide hazard maps. In, A. Bucciantti, ed., *Proceedings of International Association for Mathematical Geology Annual Meeting (IAMG'98)*, Ischia, Italy, October 3-7, 1998, pp.211-217.

Kim, K.S., Kim, W.Y., Chae, B.G. and Cho, Y.C., 2000. Engineering geologic characteristics of landslide induced rainfall –Boeun, Chungcheong Buk-do-. The Korean Journal of Engineering Geology, 10(2), pp.163-174.

Lee, S. and Min, K.D., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. *Environmental Geology*, 40(9), pp.1095-1113.

Moon, W.M., 1998. Integration and fusion of geological exploration data: a theoretical review of fuzzy logic approach. *Geosciences Journal*, 2(4), pp.175-183.

Van Westen, C.J., 1993, *Application of geographic information systems to landslide hazard zonation*, Ph.D. Thesis, Technical university of Delft, International Institute for Aerospace Survey and Earth Sciences, The Netherlands, ITC publication 15(1), 245p.

Varnes, D.J., 1978. Slope movement types and processes, landslides analysis and control. *Special Report 176. Transportation Research Board*, Washington, DC pp.11-80.

Zadeh, L.A., 1965. Fuzzy sets. *IEEE Information and Control*, 8(3), pp.338-353.

Zimmermann, H.J., 1996. *Fuzzy set theory and its applications*. Kluwer Academic Publisher, 435p.

THE DEVELOPMENT OF A SPATIAL-TEMPORAL ANALYSIS SYSTEM PROTOTYPE FOR RICE INVENTORY DATA IN TAIWAN

Jung-Hong Hong[a] Zeal Su[b]

[a] Associate Professor, [b] Ph. D. Student, Dept. of Surveying Engineering, National Cheng-Kung University
Tainan, Taiwan, Republic of China junghong@mail.ncku.edu.tw zeal@www.sv.ncku.edu.tw

KEY WORDS: spatial-temporal analysis, rice inventory, arable lands, cropping pattern, historical change

ABSTRACT:

Rice is one of the major agricultural crops in Taiwan. The Council of Agriculture (COA) has spent millions of dollars on the creation and research of rice inventory data in the last twenty years. Although a huge volume of data archive is created, unfortunately most of the analysis has been only concentrating on overall statistical numbers. We argued in this paper that the rice inventory data actually provides both spatial and temporal perspectives toward the rice crop status, and an appropriate system design based on the spatial-temporal nature of rice inventory data can help COA to have a better control of rice production, as well as the adjustment of related agricultural policy. The proposed system prototype not only successfully fulfill spatial (single arable land or region) and temporal (single or multiple seasons) querying requirement, it also enables COA to analyze cropping patterns for the selected area. The latter is valuable to the understanding of real world phenomena and has potential to serve as reference information for such applications as nature hazard assessment and the supervised classification of remote sensing images. The major contribution of this research is spatial-temporal analysis, which were either time consuming or even impossible to complete by human operators in the past, can now be completed readily with computers. Though the efficiency of the system prototype still needs further improvement, nonetheless, the preliminary result clearly indicates the value and potential by adding spatial-temporal consideration to management and analysis of rice inventory database.

1. INTRODUCTION

In Taiwan, rice is one of the major agricultural crops and plays an indispensable role to human daily lives. To accurately control a reasonable market price, the Council of Agriculture (COA) has been monitoring rice production for more than 20 years. The result not only provides the assessment of the total rice production in a single cropping season, so that the governments can promptly and appropriately adjust production policy, it is also an important source for determining the amount of agricultural compensation for individual farmer. Though a huge amount of rice inventory data is created and accumulated, the analysis, however, has been only concentrating on the numbers on overall production statistics without taking its spatial and temporal perspective into consideration. After becoming an official member of World Trade Organization (WTO) in 2002, the local market shall be facing a tough challenge from import agricultural products. A more efficient and effective system for the management and analysis of rice inventory data at a reasonably low cost is therefore desperately necessary.

The core of the current COA approach is based on human operators' photo interpretation and manual editing. Numerous approach were proposed to improve the accuracy and efficiency of the interpretation and classification process (Tseng, 1998; Hsiao, 2001), but none of them practically dealing with issues related to the spatial-temporal management and analysis of the rice inventory data. Any query of rice crop status regarding multiple seasons either demands a huge amount of operators' works, or is even impossible to complete. We believe a spatial-temporal management and analysis system prototype is capable of at least bringing the following impacts. First, a spatial-temporal system can quickly response to users' spatial and temporal constraints and provide historical changes and spatial distribution analyzed result in map formats. The analyzed area can be as small as an arable land, or as large as the nation coverage, in comparison with the current vague statistical numbers only based on political units. Second, the land laws enforce rigorous regulations on the use or even the ownership transfer of arable lands, the land use and associated crops for most agricultural lands may follow a fixed pattern. The spatial-temporal analysis during a period of multiple seasons helps to identify arable lands with fixed patterns, e.g., arable lands constantly used for rice crops or not used anymore. This is very helpful for locating major area for rice production or arable lands that are used least frequently. Finally, the spatial-temporal system enables COA to derive knowledge about rice crops. Such information may increase the accuracy of the damage assessment if natural hazards should strike. We believe there is tons of useful information hidden in the abundant rice inventory database, only if we can have a powerful mechanism to find them. The proposed spatial-temporal system prototype would try to provide an answer to this argument.

Based on the above discussion, the key functions in the proposed system prototype include:

- query cropping status of any arable land in any cropping season,
- analyze historical cropping status of any arable land in a selected time period,
- provide analysis and production statistics of the cropping status for a selected area, and
- analyze spatial and temporal cropping patterns for a selected region

The remaining of this paper is organized as follows. Section 2 analyzes the basic property of rice inventory data, followed by sections 3 discussing the basic principles of our proposed approach. Section 4 evaluates the result by testing our system prototype and analyzes if the proposed system fulfills users requirement. Finally section 5 concludes our major work and discusses possible future research.

2. 2. DATA ANALYSIS

2.1 The Creation of Rice Inventory Data

The total area used for rice cultivation is approximately 188588 hectares in the first cropping season of the year 2001 (COA, 2001). Since most arable lands in Taiwan allow two crops in a year, two different data sets, respectively describing the status of two cropping seasons, are created every year. To fulfill COA requirements, the rice inventory data must be both complete and accurate. The completeness requirement ensures that all the arable lands are investigated, while the accuracy requirement ensures 100% of the result to be correct. The current operational procedure has gone through several major changes with the introduction of new remote sensing technologies. Although there are some impressive research on data acquisition and interpretation, unfortunately none of the proposed approach can fully satisfy the above two requirements. This is because either the cost is too high, or the accuracy is not good enough due to the typical limited sizes of arable lands in Taiwan. Consequently the current procedure still has to heavily rely on human operators' photo interpretation. Operators' loading becomes even more when operators have to further manually mark the actual boundary used for rice crops. The rice inventory data is stored in PC ArcInfo format based on the coverage of 1/5000 national topographic map systems. This framework is a grid tessellation with every map covering an area of a 1.5 x 1.5 minute latitude and longitude difference. It is likely an arable land located near the border of the map be divided into two or even more polygons in different maps. A rice inventory file is then a polygon-type geographic file that stores both the spatial and attribute components of arable lands located in its specific map coverage. Other than its original 8 digit map number, the filename of each rice inventory file is expanded to include additional 3 digits, the first two refer to the year and the last one refers to the season. For example, a file '94192038852' represents a rice inventory result of map 94192038 in the 2nd season in the year of 1996 (the calendar year 85 in Taiwan). Operators interpret whether an arable land is used for rice crop based on aerial photos, then record the result with 0 (not used), 1 (rice crops) and 2 (other crops). The combination of the first and the last cases represent arable lands not used for rice crops. Figure 1 illustrates the result of a rice inventory file by using crop status as the primary controlling factor.

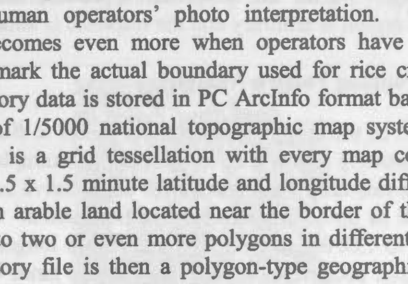


Figure 1. The illustration of rice inventory data

2.2 Basic Characteristics of Rice Inventory Data

The design of the proposed system prototype largely depends on the spatial-temporal property of rice inventory data. We'll discuss this issue from the perspectives of spatial/attribute, temporal and their integration.

□1□Spatial perspective

An arable land is recorded as a polygon and stored in the corresponding map coverage. Since only lands related to agricultural use are of interests, non-arable lands are first excluded. The interpretation result of each arable land is recorded in its associated attribute table, which include altogether 13 attributes, for example, area, land number, crop status, etc. This set of data can be directly transformed into ArcView Shape format.

□2□Temporal perspective

There have been some major discussions regarding the temporal modeling of geographic data (Barrera et al., 1990; Langran, 1993; Kemp and Kowalczyk, 1994). Guptill (1995) devoted a chapter discussing geographic data quality from temporal perspectives. The following discussion tries to clarify the inter-organization of rice inventory data from a temporal perspective. First, a rice inventory data represents rice crop status in 'world time' (Wachowicz and Healey, 1994), meaning its content is based on the time the data is collected. Second, a rice inventory file represents a time-slicing scenario of the real world at the chosen time. It is therefore dangerous to exaggerate any judgments other than what is stated, e.g., the situation between two crop seasons. Third, the temporality of rice inventory data for a crop season can be represented as a 'time-point', although conceptually 'interval-based' (Allen, 1983) approaches seem to be an intuitive choice. However, since the data of each season is created independently, there are only two possible temporal relationships existing in our system: disjoint and equal. The 'equal' relationship means selected rice inventory files are in the same season (note their actual world time may be different, but the difference is not important here), while the 'disjoint' relationship means selected files are corresponding to different seasons. This scenario is different from the 'meet' relationship existing in cadastre data (directly change from one state to another) so that event-based approaches (Peuquet and Duan, 1995) do not seem to be necessary here. Under such circumstance, a combination of year and season (a time point) would be sufficient to temporally distinguish two files.

□3□Spatial-temporal perspective

An arable land with explicitly stored spatial and temporal components will be defined as a 'temporal-arable' in the remaining of this paper. A rice inventory file is therefore a collection of 'temporal-arable' inside the corresponding map coverage in the same crop season. Since the crop status may be quite different from one crop season to another, the total number of temporal-arable and their respective boundaries in the same regions may be different as well. Based on this analysis, we introduced a five-level data management hierarchy (Figure 2). From the bottom to the top, the five levels respectively refer to a temporal-arable, a rice inventory file, a set of rice inventory file for local areas (e.g. political unit), a single-season rice inventory database and the historical rice inventory database (multiple seasons). The first level contains the collections of accumulated rice inventory data created in different seasons. As noted earlier, a disjoint temporal relationship exists between two rice inventory databases of different seasons (Level 2). From a spatial perspective, Level 2 to level 5 represent a spatial subdivision hierarchy, the relationships among them can be described by a 'part-of' spatial relationship. In our system, users have to first specify temporal constraints (what seasons are of interests), then issue spatial constraints to further narrow down the files to process.

When only parts of the arable land is used, a temporal-arable in a season may correspond to two or even more temporal-arable in another season. Sometimes, even the boundaries of the same arable land may be different in two seasons. Therefore, any spatial-temporal analysis must take both spatial and attribute (land number) component into consideration.

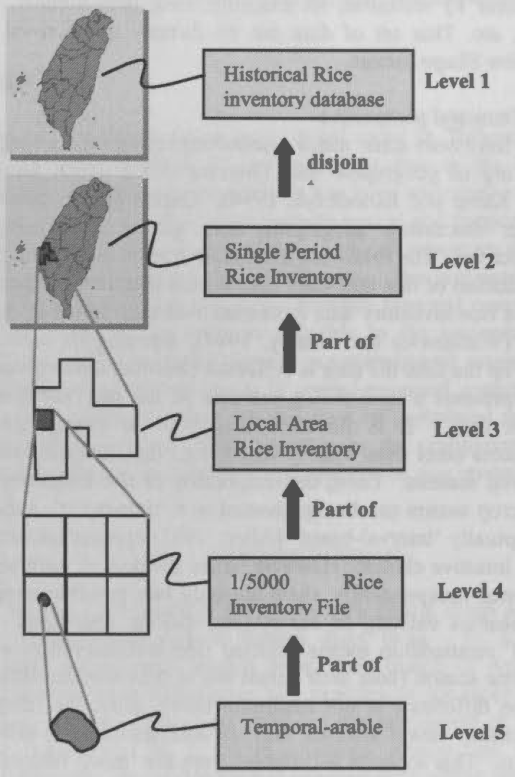

Historical Rice inventory database
Level 1
disjoin
Single Period Rice Inventory
Level 2
Part of
Local Area Rice Inventory
Level 3
Part of
1/5000 Rice Inventory File
Level 4
Part of
Temporal-arable
Level 5

Figure 2. The five-level spatial-temporal hierarchy of the rice inventory database

3. SPATIAL-TEMPORAL ANALYSIS PROTOTYPE

3.1 Query regarding single season

With the introduced five-level hierarchy, querying the rice crop status in a particular season is fairly simple. As long as users can explicitly express temporal (what season) and spatial (location or land number) constraints, we can navigate through the hierarchy to locate appropriate rice inventory file (Level 4) in which the queried temporal-arable is stored. After the temporal constraint is determined, the remaining procedures are only spatial-related. Even for a query regarding a selected region (or even queries across a number of map coverage), related files can still be readily determined by a geometric intersection test between the boundary of selected region and the map coverage framework (Level 4). Since all of the selected rice inventory files share the same attribute schema, such query as the total area of rice crops or the percentage of area lands used can all be easily calculated with statistical tools.

3.2 Query regarding multiple season

Queries regarding multiple seasons (historical query) are only meaningful when analyzing the crop status of the same region in different seasons. Users have to choose both a level to analyze (Level 2 to Level 5) and seasons of interests (Level 1). For example, when querying the historical change of an arable land, the analysis will be based on the comparison on Level 1 and Level 5. We introduced the familiar 'layer' concept here, with every layer referring to crop status of the selected region in a particular season. To query the historical change for the selected region, the system only needs to find the corresponding temporal-arable (either spatial or attribute) within all the selected layers. When the boundaries of temporal-arable in the two seasons are the same, only the comparison on land number is necessary. On the other hand, when the temporal-arable configurations of the two seasons are different (sometimes even the land number changes), an additional geometric intersection test must be included. Because the geometric test is often time-consuming, the system is default to execute only when the areas of the corresponding temporal-arable are not the same. The querying result is a list of rice crop status (with both spatial and attribute description) for the queried arable land in time dimension. This concept can be further extended for region queries related to a number of arable lands. The historical change of total area of rice crop can be easily calculated and illustrated (e.g., bar chart). In the past, the analysis basis in COA statistical report is limited to pre-selected political boundary. With the proposed approach, the selection of queried region is arbitrary and the response is almost instant.

3.3 Cropping Pattern Analysis

The goal of the cropping pattern analysis is to find arable lands that constantly follow a fixed pattern, for example, arable lands used for rice crop for a number of consecutive seasons may be regarded as the major area for rice production. One important pattern to COA is to find arable lands no longer used for rice production because more and more arable lands are losing due to the booming economics. Without the pattern analysis mechanism, COA can only control crop status in a single season. Our goal is to identify arable lands with fix pattern so that COA can have a better control of how land use change and appropriately adjust its policy. Cropping pattern analysis is again based on the layer concept. While querying lands possibly no longer used for rice production, we can first issue a 'no crop' constraint on the crop status of the selected temporal-arable and store the queried result into a temporary layer. After creating a temporary layer for each selected season, the geometric intersection of temporal-arable in these layers then represents arable not used for rice crop in the selected seasons. Figure 4 illustrates this scenario.

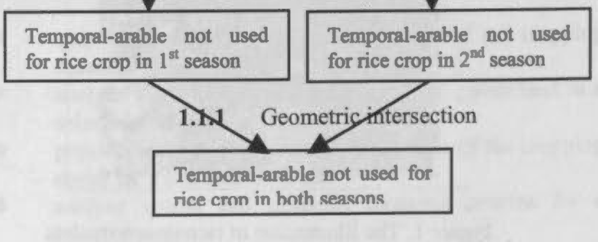

1st season temporal-arable
2nd season temporal-arable
SQL
SQL
Temporal-arable not used for rice crop in 1st season
Temporal-arable not used for rice crop in 2nd season
1.1.1 Geometric intersection
Temporal-arable not used for rice crop in both seasons

Figure 3. Basic procedure for pattern analysis

If the only concern is whether the arable lands are used for rice production or not, the addition of a season will double the number of possible combinations (either used or not used), represented as follows:

Number of patterns = 2^N

N denotes the number of seasons being analyzed. Obviously the number of possible patterns quickly increases with the increase of the number of seasons. Even with fast computers, such a dramatic increases is still not preferred. Although theoretically all possible patterns can be analyzed, nevertheless, only four patterns are practically of interests in this paper:

Arable land used for rice crops in every season (two-crop pattern)
Arable land only used for rice crops in the 1st season every year (single-crop pattern)
Arable land only used for rice crop in the 2nd season every year (single-crop pattern)
Arable land not used for rice crops (no crop pattern)

Figure 5 shows the analysis of no-crop pattern for rice inventory file '94193048' in the two seasons of 1996, we can easily conclude the majority of arable lands in this region are not used for rice crop in 1996. After analyzing rice inventory data for several consecutive seasons, the result can be regarded as a strong indication to arable lands no longer used for rice production.

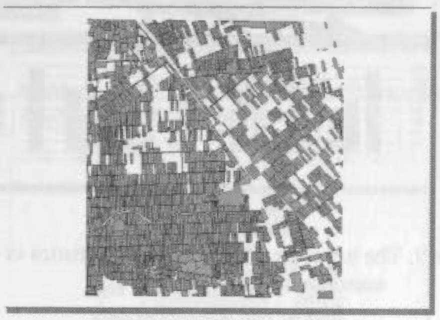


Figure 4. No-crop pattern analysis for rice inventory file '94193048' for the two seasons in 1996.

One major bottleneck in this approach is the cost of geometric intersection. Since each layer may contain thousands of polygons, the volume of calculation is sometimes still unbearable even with the introduction of the equal-area rule. For just a rice inventory file, its pattern analysis involving a number of seasons may take minutes to complete. We introduce a filter-refining strategy, based on the Minimum Bounding Rectangle concept, to reduce the number of possible intersection test. Though the efficiency did indeed improve, a more powerful strategy is still needed. We proposed a layer-difference model in (Tseng and Hong, 2001) and will do more tests in the future research.

Another property is the total area of arable lands fulfilling a certain pattern decreases with the increase of analyzed seasons. This is not surprising as the intersection of two polygons always yields a polygon that is equal to the original polygons (if the two are the same), or parts of the original polygons (the common part of the two polygons) or none at all (two polygons do not intersect). The analyzed result may dramatically change when an abnormal incident happens. For example, a drought may cause fallow in a particular season and lead to a conclusion that no arable lands fulfilling two-crop pattern in all the analyzed seasons, regardless of the fact that these arable lands may actually be used for rice crops in all the other seasons. It is certainly preferred to be able to flexibly express a query like 'find arable lands that are used for rice crops for at least 5 times in the 10 analyzed seasons'. The process is however very time-consuming because it will require altogether 638 patterns to test and may take hours or even days to complete with current system configurations.

4. TEST RESULT ANALYSIS

4.1 Test sites

We select three test sites respectively located in northern, central and southern Taiwan, each site contains either 15 or 18 maps. Most of the arable lands in northern and central test sites are two-crops pattern, while those in the southern test site are either one-crop or no-crop pattern. Under our request, COA provides ten seasons (from 1996 to 2000) of rice inventory data for the selected test sites. Some of requested files were unavailable and totally 453 files were acquired and processed. After converting from its original ArcInfo format to ArcView shapefile format, the files are organized into spatial-temporal rice inventory databases according to the five-level hierarchy discussed in section 2. The system prototype is developed in ArcView 3.2 environment with Avenue.

4.2 Single-season query

For queries regarding only a season, system will prompt users to first specify the season of interests. Users then have the choice to either use land number or create a queried region in the visual map interface. The land number constraint allows only single temporal-arable query, while the map interface constraint determines all the temporal-arable that intersect with the queried region and prompt users a list to make further selection. After selecting a temporal-arable from the list, its associated attributes are shown accordingly for users inspection. Figure 5 shows a queried region simultaneously intersecting three arable lands. Note all three of them are selected. Meanwhile, the use of visual able 'hue' further enhances the thematic display of rice inventory data (Robinson et al., 1996). Figure 6 shows the analyzed result to a query regarding the total area statistics of rice crop for arable lands located in the prefecture of Tainan. Note the file on the top-right corner in the selected season is missing, suggesting a possible crisis in the future multiple-season analysis.

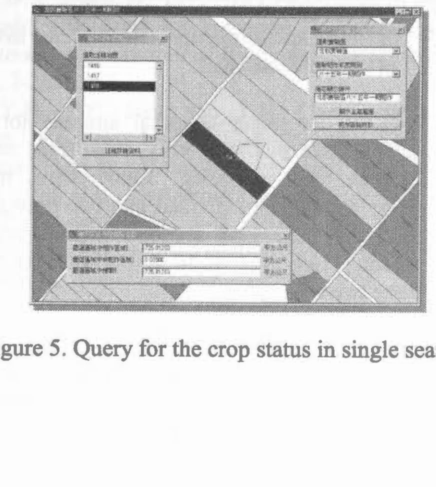


Figure 5. Query for the crop status in single season

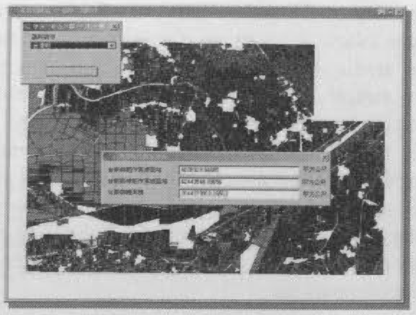


Figure 6. Rice crop statistics for selected region.

4.3 Multiple-season query

Multiple-season queries deal with the historical change of a selected arable land or region. Depending on the number of seasons of interests, the interface consists of the same number of map window, respectively displaying the rice crop status of the selected seasons. Figure 7 demonstrates the result after selecting 10 seasons of rice inventory data. Spatially speaking, these map windows are 'dynamically' linked with each other, such that whenever users zoom in/our or pan the content of a map window, the content and coverage of the other windows adjust correspondingly and automatically. This provides a great advantage that users can visualize the same area simultaneously and can create an overall impression easily. The selection of queried region can be easily completed with the index map locating at the left of the map windows. By overlaying 1/25000 scale topographic maps as reference maps, the index map provides users a convenient tool to select area of interests. After creating spatial constraints in the index map window, the rice crop status of the selected seasons automatically appears in the map windows. Note the contents of these map windows are slightly different in Figure 7, as some arable lands not categorized as arable lands were later identified as arable lands not used for rice crop. A possible reason is the change of interpretation standard from COA or misjudgment by human operators. Whatever the reason may be, this suggests a possible crisis regarding the correct analysis of the proposed spatial-temporal system.

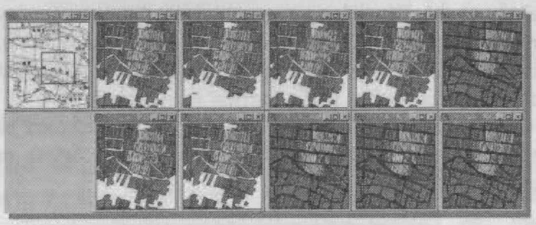
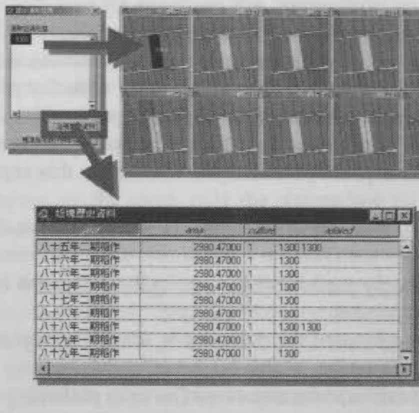


Figure 7. Map interface for the illustration of rice inventory data in multiple seasons.

Figure 8 shows the spatial-temporal analysis for a selected temporal-arable (land number 1300). Note there are two cases in the attribute table recording '1300' twice, meaning the corresponding polygons are divided into two polygons in these two seasons and their land numbers remain unchanged. A possible reason is only parts of the arable lands are used for rice crops in these two seasons.

Figure 9 shows a bar chart based on the total area of arable lands used for rice crop (vertical axis) and selected seasons (horizontal axis). After users create queried region, system automatically searches the same area coverage in other temporary layers and calculate the total area for every season. The pattern in Figure 9 appears to be either all arable lands are used for rice crop or not at all, likely a scenario for area that totally rely on irrigation water. When irrigation water supply is insufficient, farmer within the regions are forced to leave their arable lands fallow under current regulations.



八十五年二期稻作	2980.47000	1	1300 1300
八十六年一期稻作	2980.47000	1	1300
八十六年二期稻作	2980.47000	1	1300
八十七年一期稻作	2980.47000	1	1300
八十七年二期稻作	2980.47000	1	1300
八十八年一期稻作	2980.47000	1	1300
八十八年二期稻作	2980.47000	1	1300 1300
八十九年一期稻作	2980.47000	1	1300
八十九年二期稻作	2980.47000	1	1300

Figure 8. Historical query regarding temporal-arable.

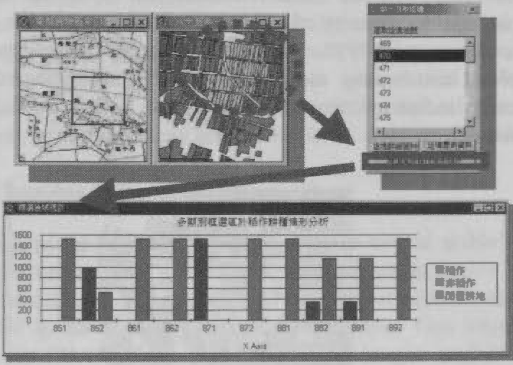


Figure 9. The bar chart of the total area statistics in different seasons.

4.4 Cropping pattern analysis

Figure 10 and Figure 11 demonstrate two cropping pattern analysis of the rice inventory file '94202038' (in central test site) from the two seasons in 1996. Figure 10 shows the arable lands used for rice crops in both seasons, while figure 11 shows arable lands not used for rice crops in both seasons. It is clear that the majority of arable lands in this region fulfill the two-crop pattern and those not used mainly locate in the northern part of the map. The combination of arable lands with two-crop and no-crop pattern is close to the total area of the arable lands in the map coverage. Further analysis on the single-crop pattern supports this conclusion because only a few arable lands fulfill the constraint. It is therefore reasonable to conclude the analyzed region is a major area for rice production. A natural hazard striking this region has a better chance to cause more damage to the rice production. On the contrary, a similar case would not bring too much damage to arable lands with 'no crop' patterns (please refer to Figure 4.).

We also conducted several tests on the crop status of the southern test site, the result indicated most of the arable lands are single-crop or no-crop pattern. Nevertheless, how many patterns there are or what pattern is found within what region are not really the major concerns of this paper, what we would really like to demonstrate is users' loading is simplified to be just a few mouse clicks for creating spatial and temporal constraints, in comparison with the possibly tremendous operator loeading in the past.

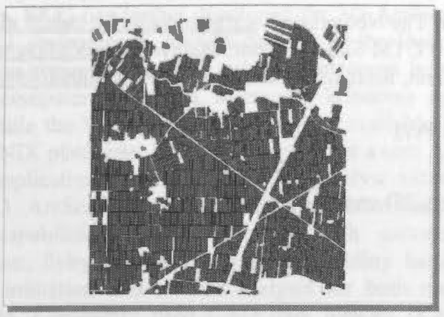


Figure 10. Arable lands with two-crop pattern for rice inventory file '94202038' in 1996.

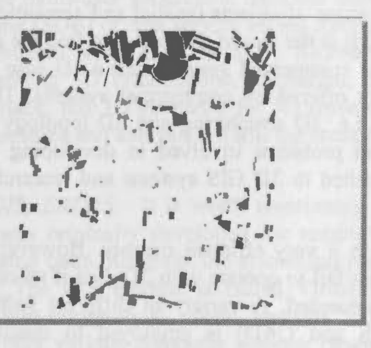


Figure 11. Arable lands with no-crop pattern for rice inventory file '94202038' in 1996.

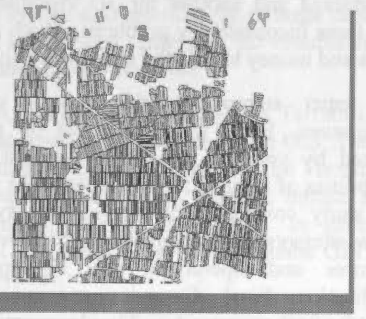


Figure 12. Arable land with two-crop pattern for rice inventory file '94202038' in the period of year 1996~1998.

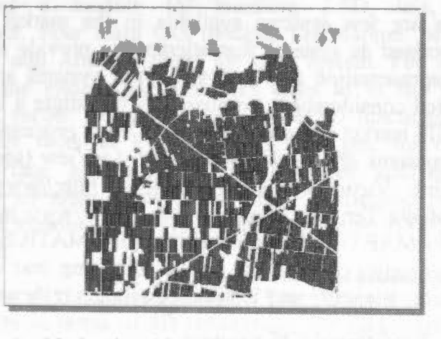


Figure 13. Arable lands with two-crop pattern for rice inventory file '94202038' in the period of year 1996~2000.

Figure 12 and Figure 13 further illustrate the analysis of the same region with two-crop pattern during the period of year 1996~1998 (6 seasons) and 1996~2000 (10 seasons). Although both visual inspection and total area statistics show a decreasing trend on the arable lands (as discussed in section 3.3), it is still safe to support the above 'major production area' conclusion because the trend does not drop dramatically with the addition of analyzed seasons. Such a quantitative analysis was almost impossible to COA in the past, but it is no longer a problem from now on.

5. CONCLUSION

In the past, COA has to rely sorely on the statistic numbers to monitor and control rice production. The proposed system prototype successfully encompasses spatial and temporal nature of rice inventory data, and is capable of providing an effective spatial-temporal analysis capability to the quickly increasing rice inventory database. The five-level hierarchy enables users to efficiently narrow down the required data with given spatial and temporal constraints. By analyzing arable lands with fixed patterns from the historical rice inventory database, COA can not only query the historical change of rice crop for any region, but also accumulate spatial-temporal knowledge about rice production, for example, a better understanding to major rice production area or area no longer used for rice production. By introducing the spatial-temporal concept, this system prototype has great potential to improve the current analysis procedure and even open a new dimensionality for the use of rice inventory data in the future.

Acknowledgement

This result of this paper is supported by the Council of Agriculture, under research grant 90 Agriculture Science-4.1.3-Food-Z1(4).

Reference

Allen, J. F., 1983. Maintaining Knowledge about Temporal Intervals, *Communications of the ACM*, pp. 832~843.

Barrera, R., Frank, A. U. & Khaled, A.-T., 1990. Temporal Relations in Geographic Information Systems, *NCGIA Technical Report* 91-4.

Guptill, S. C. , 1995, Temporal Information, in edited book titled *Elements of Spatial Data Quality*, pp. 153~167.

Hsiao, K.-H., Lau, C.-C. and Shih, T.-Y., 2000. Combined Remote Sensing and GIS Data for Rice Paddy Landuse Interpretation, *Journal of Photogrammetry and Remote Sensing*, 5(4), pp. 1-22.

Langran, G., 1993. Time in Geographic Information Systems, Taylor & Francis.

Peuquet D. J., Duan, N., 1995. An Event-Based Spatialtemporal Data Model (ESTDM) for Temporal Analysis of Geographic Data, *International Journal of Geographic Information Systems*, 9(1), pp. 7-24.

Tseng, Y.H., P.H. Hsu, and Y.H. Chen, 1998. Automatic Detecting Rice Fields by Using Multispectral Satellite Images, Land-parcel Data and Domain Knowledge, *Proceedings of the 19th Asia Conference on Remote Sensing*, pp. R-1-1~R-1-7.

Tseng, T.-Y. and Hong, J.-H., 2001, The Spatial-temporal Model for Rice Inventory Database, *Proceedings of the 20th Conference of Surveying and Mapping*, Vol. 2, p. 785-792, National Central University, Chung-Li, Taiwan.

Wachowicz, M. and Healey, R., 1994, Towards Temporality in GIS, in edited book titled *Innovations in GIS*, pp. 105~118, Taylor & Francis.

3D GIS: CURRENT STATUS AND PERSPECTIVES

Siyka Zlatanova[a], Alias Abdul Rahman[b], Morakot Pilouk[c]

[a]Delft University of Technology, Thijsseweg 11, 2629JA, Delft, The Netherlands, s.zlatanova@citg.tudelft.nl
[b]Department of Geoinformatics, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia, alias@fksg.utm.my
[c]Environmental System Research Institute (ESRI), 380 New York Street, Redlands, California, USA, mpilouk@esri.com

Commission IV, WG IV/1

KEY WORDS: spatial data, database, CAD, 3D topology, 3D visualization, 3D analysis

ABSTRACT:

Currently, variety of software is already capable of handling a wide range of spatial problems, beginning with approaches for describing spatial objects to quite complex analysis and 3D visualisation. However, increasing number of applications need more advanced tools for representing and analysing the 3D world. Among all types of systems dealing with spatial information, GIS has proven to be the most sophisticated system that operates with the largest scope of objects (spatial and semantic), relationships and provide means to analyse them. However, what is the status of the 3D GIS? It is the aim of this paper to find the answer by analysing both software available and efforts of researchers. An overview of several commercial systems and a 3D case study performed in Oracle and Microstation provides knowledge about the 3D functionality offered by commercial systems. The most significant achievements in the 3D research area concerning key issues of 3D GIS, i.e. 3D structuring and 3D topology portray the current research status. At the end, the paper addresses some of the issues and problems involved in developing such a system and recommends directions for further research. The scope of the paper is limited to 3D GIS systems and research in vector domain. Problems of subsurface applications are excluded as well.

1 INTRODUCTION

The need of 3D information is rapidly increasing. Currently, the most often quoted areas of human activities can be summarised as 3D urban planning, environmental monitoring, telecommunications, public rescue operations, landscape planning, geological and mining activities, transportation monitoring, real-estate market, hydrographical activities, utility management and military applications. Practically, the area of interest grows significantly when the 3D GIS functionality is available on the market. The role of geo-information in all kinds of business processes is getting quite transparent. Such term as "location-specific information" and "location-based services" become a part of the daily business language to denote the link between the virtual world of information (transactions, events, internet communication) and the real world of information - customers, inventory, shipping and the like. Most business transactions rely on information systems to be executed successfully as the geo-information (location-specific information) is critical for many of them (see Sonnen and Morris, 2000). Once the developments in the 3D GIS provide a compatible functionality and performance, the spatial information services will evolve into the third dimension.

Traditionally, the GIS system should be able to maintain information about spatial phenomena and provide means to analyse it and thus gain knowledge of the surrounding world. In general, consensus on the demanded functionality of GIS is achieved already years ago. The tasks or the functions of a GIS are specified as follows (see Raper and Maguire, 1992): 1) data capture, 2) data structuring, 3) data manipulation, 4) data analysis, and 5) data presentation. Indeed, 3D GIS aims at providing the same functionality as 2D GIS. Unfortunately, such 3D systems are still not available on the market. The development of 3D GIS is not an easy task. Nowadays, 2D GISs are common and widely used to handle most of the 2D GIS tasks in a very efficient manner. However, the same kind of systems fail to operate with 3D data if more advanced 3D tasks are demanded. A variety of different software (i.e. 2D GIS, DBMS and CAD) is employed to maintain the objects of interest and extract the required information. Due to deficiency of any of the system to handle 3D objects, the data are spread between several systems. For example, one system is used for data storage and another for 3D visualisation. This situation often faces inconsistency problems, which results in extra time, efforts and money to find the appropriate solution.

This paper summarises the current status of 3D GIS development. First, we concentrate on recent achievements reported by vendors. We briefly present our survey on the possibilities of some GIS's available in the market and analyse a case study completed on commercial systems. Second, we review attempts of researchers toward providing an appropriate structures and operations for 3D spatial analysis and visualisation. Final discussion recommends directions and topics for further research and implementations.

2 3D GIS IN THE MARKET

There are few systems available in the market that can be categorised as systems that attempts to provide a solution for 3D representation and analysis. Four systems are chosen for detailed consideration, because they constitute a large share of the GIS market and provide some 3D data processing functions. The systems are the 3D Analyst of ArcView (see ESRI Inc.), Imagine VirtualGIS (ERDAS Inc., http://www.erdas.com), GeoMedia Terrain (Intergraph Inc., http://www.integraph.com) and PAMAP GIS Topographer (PCIGEOMATICS, http:// www .pcigeomatics.com). Parts of the following text are based on available literature and Web-based product reviews.

2.1 Traditional GIS vendors

ArcView 3D Analys, ESRI: The 3D Analyst (3DA) is one of the modules available in ArcView GIS. ArcView is designed to provide stand alone and corporate wide (using client-server network connectivity) integration of spatial data. With 3DA one can manipulate basically 2.5D data such as surface generation, volume computation, draping raster images, terrain intervisibility from one point to another. The system works mainly with vector data. Raster files can be incorporated into 3DA, but only for improving the display of vector data. During the last three years, ESRI has further developed the 3D Analyst for the ArcGIS 8.1 environment. ArcGIS consists of the Desktop and Workstation components. The Desktop component is based on personal computer (PC) and Microsoft Windows operating system, while the Workstation component is available for both PC and UNIX platforms. ESRI also introduced a new ArcScene desktop application as part of the 3D Analyst extention to ArcGIS 8.1. ArcScene is a stand-alone application that provides all the capabilities similar to 3DA with enhanced 3D visualization, flyby, texture mapping on building facades, 3D symbols, animation and surface analysis for both raster and vector data. Commonly used CAD data formats (e.g. DGN, DXF, DWG) can directly be read and displayed in ArcScene. ArcScene can also access and display both raster and vector data stored on the multi-user geographic database using ESRI Spatial Data Engine (SDE) or data service on the Internet in the distributed environment using ESRI ArcIMS. Although major progress on improving 3D visualization, animation, and data access has been made, full 3D geometry for 3D representation, topological relationships and analysis are still the areas left to be addressed.

Imagine VirtualGIS, ERDAS: It is worth mentioning that the Imagine system was originally developed for remote sensing and image processing tasks. Recently, the system has provided a module for GIS. The GIS module is called VirtualGIS and provides some three-dimensional visual analysis tools. The system has run under various computer systems ranging from personal computers to workstations. It is a system that has an emphasis on dynamic visualisation and real-time display in the 3D display environment. Besides various and extensive 3D visualizations, the system also provides fly-through capabilities. As with 3DA this system also centres on 3D visualization with true 3D GIS functions hardly available.

GeoMedia Terrain, Integraph.Inc: GeoMedia Terrain is one of the subsystems that work under the GeoMedia GIS. The system runs under the Windows operating systems. The Terrain system performs three major terrain tasks, namely, terrain analysis, terrain model generations, and fly-through. In general, the Terrain serves as DTM module for the GeoMedia GIS without true 3D GIS capabilities.

PAMAP GIS Topograph, PCIGeomatics: It runs under Windows95/98 and NT operating systems. PAMAP GIS is a raster and vector system. Besides its 2D GIS functions, the system has a module for handling 2.5D data, called Topographer. Four main GIS module, i.e. Mapper, Modeller, Networker and Analyser form the core system. For 2D data handling, the system performs GIS tasks as in the systems mentioned earlier. Most of the so-called 3D functions in the Topographer work as by any DTM packages, for example terrain surface generation, terrain surfaces analysis (e.g. calculation of area, volume) and 3D visualisation (such as perspective viewing). This system also focuses on 3D display of terrain data.

In summary, all the systems revealed little provision of 3D GIS functionality in terms of 3D structuring, 3D manipulation and 3D analysis but most of them can handle efficiently 3D data in the 3D visualization aspect. A fully integrated 3D GIS solution has yet to be offered by general purposed GIS vendor.

2.2 The tandem DBMS &CAD

The GIS, i.e. integration of semantic, geometric data and spatial relationships, seems to be the most appropriate system ensuring a large scope of analysis and thus serving many applications and daily activities. Therefore vendors dealing with either spatial or semantic information attempt to provide some GIS functionality already for years. CAD vendors (such as Autodesk, Bentley) provide means to link semantic data to 2D, 3D geometry and organise topologically structured layers; DBMS (Oracle, Informix) introduce spatial descriptors to represent geometry data and maintain them together with the semantic data.

A logical consequence of all the attempts is the agreement on the manner for representing, accessing and disseminating spatial information, i.e. the OpenGIS specifications (see OpenGIS specifications). This agreement makes possible efforts of vendors and researches from different fields to be united and streamed to one direction, i.e. development of a functional GIS. As a result, increasing number of DBMS offer already functionality to store, retrieve and analyse spatial data. Moreover, growing number of CAD vendors develop tools to access, visualise and edit the spatial data maintained in DBMS. Among the several DBMS (Oracle, Informix, Ingres) and CAD/GIS applications (Microstation, AutoCAD, MapInfo) already embracing OpenGIS specifications, we have selected Oracle Spatial 8i and the new product of Bentley, Microstation Geographics iSpatial to investigate the 3D operations offered. GeoGraphics iSpatial establishes a connection directly to Oracle Spatial 8i.

The organisation of data within Geographics iSpatial is defined in a project hierarchical structure. *Project* refers to as the root and represents the data for the entire study area. The second level is the *category*, which groups features with a similar theme (e.g. buildings, rivers). One project can have many categories but a category may belong to only one project. *Feature* is at the third level and represents one or more objects in the real world (e.g. the bank building, the school building). A feature incorporates all the attribute and geometric data available for a particular real object. A category may have many features but a feature may belong to only one category. Feature is the basic structural unit in GeoGraphics iSpatial. To be able to distinguish between different spatial objects stored in Oracle Spatial 8i, each object has to be assigned to a feature. Furthermore, edited and newly created objects cannot be posted in the database without attributing predefined features to them. Geometry of the objects is organised in one or more *spatial layers*.

The geometry in Oracle Spatial 8i is defined by the *geometric type*. Oracle Spatial 8i supports 2D geometric types, i.e. point, line and polygon (see Oracle Spatial 8i). Lines and polygons are represented as an ordered set of points. The indication for a closed polygon is equivalence of the first and the last point. Self-intersecting lines are supported and they do not become polygons. Self-intersecting polygons are not supported. The geometric types are defined in the Oracle Spatial 8i object-relational model as objects (i.e. *mdsys.sdo_geometry)* and contain information about type, dimension, coordinate system, holes, and provide the list with the coordinates. The structure of the object is given bellow:

Name	Null?	Type
SDO_GTYPE		NUMBER
SDO_SRID		NUMBER
SDO_POINT		SDO_POINT_TYPE
SDO_ELEM_INFO		SDO_ELEM_INFO_ARRAY
SDO_ORDINATES		SDO_ORDINATE_ARRAY

Despite the 2D geometry types, 3D coordinates can be maintained. For example, the five parameters of the geometry object for a 3D polygon with four vertices *v*(X,Y,Z), i.e. *v1* (10, 10, 0), *v2* (11, 9, 1), *v3* (11, 12, 0) and *v4* (9, 11, 1) will have the following values:

SDO_GTYPE = 3003. The first 3 indicates three-dimensional object and the second 3 indicates a polygon.

SDO_SRID = NULL. The coordinate system is not specified, i.e. decoded in the coordinates.

SDO_POINT = NULL. The described type is polygon and therefore the value is NULL.

SDL_ELEM_INFO = (1, 1003, 1). The first 1 in the sequence 1,1003,1 gives details about the geometry type (i.e. a simple polygon connected by straight lines). 1003 indicates that the polygon is an exterior ring. The final 1 specifies the geometry type, i.e. polygon. Furthermore, these particular values certify that the polygon does not contain holes.

SDO_ORDINATES = (10, 10, 0, 11, 9, 1, 11, 12, 0, 9, 11, 1, 10, 10, 0).

2.2.1 Case study: the city of Vienna: In this case study, we assumed the following scenario: the user has 3D data organised only in a database (a quite common case for real world data), i.e. no file with graphical information (e.g. DGN) exists. We have experimented with a set of 1600 buildings from the city of Vienna. Planar rectangular faces constitute each building. The data are organised according to the Simplified Spatial Structure (see Zlatanova 2000) and further converted to the geometry representation of Oracle Spatial 8i. The conversion is completed with a topology-geometry procedure similar to the one described in Oosterom et al 2002.

Since the Oracle Spatial 8i geometry does not maintain a true 3D object, we represented every building as a set of faces (walls, flat roofs and foundations). The faces are stored as polygons with 3D coordinates. The data set with 1600 buildings is organised in a relational table (BODY_SDO) that originally consisted of only four columns namely (MSLINK, BODY_ID, FACE_ID and SHAPE). The column SHAPE contains the *mdsys.sdo_geo*metry object, i.e. the polygons. Thus the links between FACE_ID and SHAPE is 1:1 and the link between FACE_ID and BODY_ID is m:1. Table 1 illustrates the content of the relational table BODY_SDO before (in bold) and after connection to Geographics iSpatial. Furthermore, the spatial data is indexed with R-three index (i.e. BODY_IDX_RT$ table was created), and registered in the USER_SDO_GEOMETRY_METADATA table by giving the name of the table (i.e. BODY_SDO), the name of the column with geometry (i.e. SHAPE) and the range of the data set. Examples with SQL statements accomplishing these operations can be found in Stoter and Oosterom 2002.

Table 1: Description of BODY_SDO table. The columns in regular font are added by GeoGraphics iSpatial (see the text)

Name	Null?	Type
MSLINK	**NOT NULL**	**NUMBER(10)**
BODY_ID		**NUMBER(10)**
FACE_ID		**NUMBER(10)**
SHAPE		**MDSYS.SDO_GEOMETRY**
BODY_SDO_DFLAG		NUMBER(10)
BODY_SDO_UDL		RAW(200)
BODY_SDO_LOCK		NUMBER(10)
BODY_SDO_FID		FCODE_LIST
BODY_SDO_CREATED		DATE
BODY_SDO_REVD		DATE
BODY_SDO_RETIRED		DATE
BODY_SDO_XML		VARCHAR2(1024)
BODY_SDO_TXT		VARCHAR2(1024)
BODY_SDO_STYLE		UG_STYLE

This user-defined representation of geometry is further accessed within GeoGraphics iSpatial. Since the steps that one has to follow are not that trivial, they will be explained in the following section.

1. *Creating project, category and features*. Bearing in mind, the basic conceptual structure of GeoGraphics iSpatial we created a project (Vienna), a category (buildings) and several features (build1, build2, build3 and build4) in GeoGraphis iSpatial. This operation resulted in 12 relational tables in Oracle Spatial 8i. The names of the tables created by GeoGraphics iSpatial and us (in bold) are listed bellow:

BODY_IDX_RT$, BODY_SDO, CATEGORY, FEATURE, MAPS, MSCATALOG, UGCATEGORY, UGCOMMAND, UGFEATURE, UGJOIN_CAT, UGLAYER, UGMAP, UGMAPINDEX, UGTABLE_CAT

Among all the tables, MSCATALOG and FEATURE are of practical interest. The first table maintains reference to all the tables used in the project. The second one contains information (names, codes, unique identifiers, etc.) related to all the features created by the user.

The spatial data (BODY_SDO table) becomes visible in the Query tool (see Figure 1, Spatial Query), i.e. it is possible to query and display the entire layer. However, the settings are not sufficient to post data in the database. The table has to be linked to a spatial layer and the objects to features.

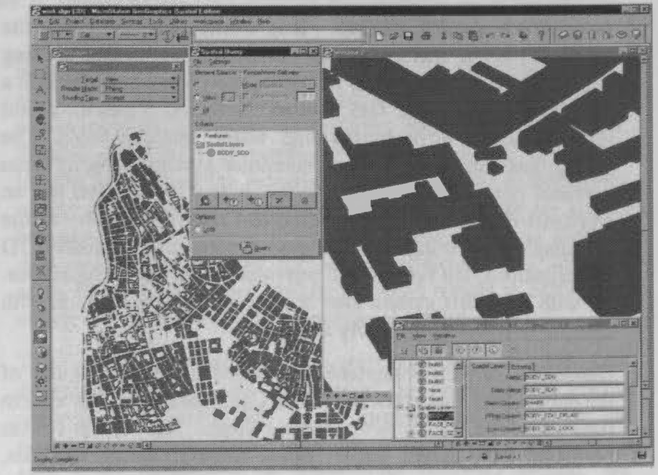


Figure 1: GeoGraphics iSpatial, query of the layer BODY_SDO

2. *Creating spatial layer*. The table with the geometry (i.e. BODY_SDO) with geometry column SHAPE has to be referred as a spatial layer in GeoGraphics iSpatial. Further, all the features that are to be associated with objects in this layer need to be assigned to the layer (again in GeoGraphics iSpatial). This operation extended our table BODY_SDO with 10 new columns (see Table 1).

3. *Linking spatial objects with features*. First, one should make sure that the table with the spatial data (i.e. BODY_SDO) is declared in the table MSCATALOG. The project tables CATALOG and FEATURE are automatically registered there by GeoGraphics iSpatial under entity numbers 1 and 2. Second, the column BODY_SDO_FID (in the BODY_SDO table) has to be populated. The column references a feature (from FEATURE) to a particular object (from BODY_SDO). The operation can be performed either in GeoGraphics iSpatial or Oracle. Last, all the values in the column BODY_SDO_LOCK (giving information about the owner of the data) have to be set 0 (i.e. belong to the owner of the table). A PL/SQL script (a high-

level language supported by Oracle) completes these two operations:

```
... FOR i in n..m LOOP
     update body_sdo set body_sdo_fid = fcode_list (fcode_item
(2,4,1,0), fcode_item (5,i,0,0)) where body_id=i;
    update body_sdo set body_sdo_lock = 0 where body_id=i;
 END LOOP; ...
```

Fcode_item (p_1, p_2, p_3, p_4) provides the link between feature and spatial object. The first of *fcode_item's* is related to the feature as it is described in the FEATURE table and the second to the spatial object from the BODY_SDO table. Parameter p_1 is the number of the two tables in the MSCATALOG (as it appears under the column ENTITY). Parameter p_2 is the number of the feature in FEATURE table (given in MSLINK column) and the identifier of the object (i.e. BODY_ID). Note that in this case, one feature (i.e. number 4) is assigned to several objects. The third parameters give indication whether the description is for feature (i.e. 1) or spatial object (i.e. 0). Cases with multiple references between object and feature are resolved by introducing a new *fcode_item* in the *fcode_list* description.

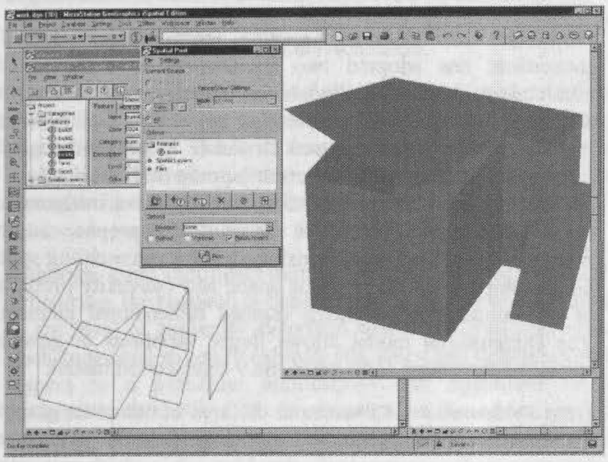


Figure 2: GeoGraphics iSpatial, editing and posting a feature

Having all the initialisations done, it became possible to query the data as they are defined in Oracle Spatial 8i. The query can be specified either per layer (see Figure 1) or per feature (see Figure 2). We have tested editing, creating new objects and posting them to the database. More examples, related to combining 2D and 3D data can be found in Stoter and Oorsterom, 2002.

2.2.2 Analysis: This case study exhibited valuable information related to 3D functionality currently offered. It has clearly showed that the operations needed to access and manipulate spatial data are still not transparent, standardised and user-friendly. The user is expected to have excellent skills in both systems, i.e. understanding the conceptual representation in GeoGraphics iSpatial and being aware of the implementation in relational tables in Oracle Spatial 8i.

Data Structuring. The concepts implemented in both systems follow closely the OpenGIS specifications, i.e. the notation of a *geographic feature*, which spatial characteristics are represented by geometric and topological primitives. Nevertheless, the implementations are still not completely application independent. The test has revealed that one significant part of the information about the geographic feature is maintained at a database level. However, the notations (table names, columns, object definitions) have very specific application-oriented (in this case Microstation) meaning. For example, if the user decides to keep the database and change the CAD package, he/she will need to create the feature-geometry link from scratch.

Moreover, the *feature* introduced in GeoGraphics iSpatial, allows the user to define arbitrary number of *feature types* and link them to geometric data. However, the further classification of features is restricted to only one level (categories), i.e. classification of categories in not supported. Two classifications levels may appear insufficient in describing 3D world objects. For example, looking at a building, at least three levels of hierarchy might be necessary, i.e. a particular room, an apartment, and a floor. Conceptually, the *layer* can be used as a container of geometry types with specific characteristics, but one-to-one correspondence between a spatial layer in GeoGraphics iSpatial and a relational table in Oracle Spatial 8i, may lead to an unnecessary partitioning of the data and complicate the consistency check.

As it was mentioned before, despite the lack of a real 3D object, description of 3D data is possible in the geometry types of Orcale spatial. The Z value is maintained together with the X,Y values, i.e. it is not an attribute. Another positive discovering is the definition of *mdsys,sdo_geometry* object in Oracle Spatial 8i 8i, which allows a straight forward extension toward describing a 3D object. Stoter and Oosterom, 2002 propose appropriate values for SDO_GTYPE, SDO_SRID, SDO_POINT, SDO_ELEM_INFO and SDO_ORDINATES D to describe 3D objects (e.g. tetrahedron, polyhedron, polyhedron with holes, etc.). The SDO_ORDINATES array is suggested to have two sections, i.e. a list of coordinates and references to the list. This approach will reduce considerably the size of the array, which is a critical consideration in maintaining 3D data.

The support of parameters to describe physical properties of 3D objects is still missing. Currently, the feature description (in FEATURE and UGFEATURE tables) permits properties of lines (e.g. colour, width, gaps width, type line) to be specified, but no properties of polygons are considered. For example, the colour of the polygon (in a rendering mode) is selected with respect to the colour of the line. 3D realistic visualisation is practically not possible due to lack of a mechanism to specify texture parameters per face.

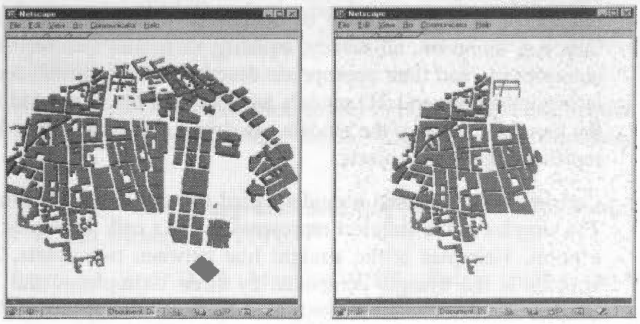


Figure 3: Oracle Spatial 8i query: left) spatial operator SDO_WITHIN_DISTANCE and right) function FOV

Data Analysis. Real possibilities to analyse 3D data in GeoGraphis iSpatial and Oracle Spatial 8i are still missing. As mentioned before, the topological primitives are not implemented yet. Tools in GeoGraphics iSpatial to create 2D topological layers or tools in Oracle Spatial 8i to perform spatial operations are provided but they operate with only X,Y coordinates. Some of the operations accept X,Y,Z values but the computations are purely 2D. Figure 3 illustrates a query performed on the same data set (table BODY_SDO) utilising the Oracle spatial 8i operator SDO_WITHIN_DISTANCE and a further extension to find a Field-of-View (FOV) for given direction and angle of view. The SQL query utilising the spatial operator is given bellow:

```
SELECT body_id, face_id, shape FROM body_sdo
WHERE sdo_within_distance (shape, mdsys.sdo_geometry
(3001,NULL, mdsys.sdo_point_type (x_Input, y_Input, z_Input),
NULL, NULL),'distance=700') = 'TRUE';
```

Data Manipulation and Visualisation. Apparently the greatest benefits of the DBMS-CAD integration are in the area of visualisation and editing of data. It is well known and frequently commented that the amount of data to be visualised in 3D increases tremendously and requires supplementary techniques (LOD, on-fly simplification, etc.) for fast rendering. Having 3D data stored in a database, the user has the possibility to extract only a limited set of data (e.g. one neighbourhood instead of one town) and thus critically reduce the time for loading. For example, the whole Vienna data set (about 19000 polygons) is loaded for about 3-4 minutes in comparison to one building that comes up for fractions of a second. Locating, editing and examining a particular object become also quick, simple and convenient. Indeed, the elements that can be edited correspond to the geometry representation in Oracle Spatial 8i. In our case, one building is aggregation of several faces but practically the accessible elements are “loose” polygons. The editing operations are restricted to the defined objects (in our case polygons and their vertices). For example, a shift of one vertex will change the vertex of the selected polygon (see Figure 2). Moreover, many of the shapes provided by MicroStation cannot be posted to the database. This is to say, spheres, cylinders, cubes and all types of extruded shapes, have to be simplified to points, lines and polygons.

3 3D GIS IN THE RESEARCH

The research in 3D GIS is intensive and covers all aspects of the collecting, storing and analysing real world phenomena. Among all, 3D analysis and the issues related (topological models, frameworks for representing spatial relationships, 3D visualisation) are mostly in the focus of investigations.

Topological model: The topological model is closely related to the representation of spatial relationships, which are the fundament of a large group of operations to be performed in GIS, e.g. inclusion, adjacency, equality, direction, intersection, connectivity, and their appropriate description and maintenance is inevitable. Several 3D models have already been reported in the literature. Each of the models has strong and weak points for representing spatial objects.

Carlson 1987 proposed a model called the simplicial complex. The simplex is the simplest representation of a cell. 0-simples is a point, 1-simples is the straight line between two points, 2-simplex is the triangle composed by three 1-simplexes and 3-simples is the tetrahedron composed by three 2-simplexes. The author uses the simplexes to denote spatial objects of node, line, surface, and volume. The model can be extended to *n*-dimensions. Molenaar 1992 presents a 3D topological model called 3D Formal Vector Data Structure (3DFDS). The model maintains nodes, arcs, edges and faces that are used to describe four types of features named points, lines, surfaces and bodies. Compare to the simplex approach, 3DFDS has less restrictions to the objects, e.g. the 2-cell (face) can have arbitrary number of 1-cells. Furthermore, some spatial relationships are explicitly stored, i.e. face-body. The model belongs to the group of Boundary representations (B-reps). Cambray 1993 proposes CAD models for 3D objects combined with DTM as a way to create 3D GIS that is a combination of Constructive Solid Geometry (CSG) and B-rep. Pigot 1995, developed a 3D topological model based on 0,1,2, 3 cell, which maintains an explicit description of relationships between cells. Work by Pilouk 1996 focussed on the use of TIN data structure and relational database for 2D and 2.5D spatial data. He proposes an integrated data model for 3D GIS (i.e. TIN and 3D FDS), which produced a practical approach to the problem. Moreover, the author develops the Tetrahedron Network (TEN) data structure that is based on simplexes. The structure assures strict consistency check, built on the generalised Euler’s Equality. De la Losa 1998 and Pfund 2001 propose object-oriented models similar to Molenaar’s one but they have include several more explicitly stored spatial relationships. For example, De la Losa maintains the relationship arc-faces as strict ordering of faces is introduced. Zlatanova 2000 discusses some aspects of the data structuring and 3D visualisation with respect to data query over the Web. The proposed data structure lacks the 1-cell in order to improve the performance of the system. Abdul-Rahman 2000 focuses on the object-oriented TIN (2D and 3D) based GIS. The conceptual and the logical model are developed based on the Molenaar’s data model.

The consensus on a 3D topological model is not achieved yet.

Formalism for detecting spatial relationships: OpenGIS consortium has adopted two frameworks to detect spatial relationships known as Egenhofer operators and Clementini operators based on the 9-intersection model (see Egenhofer and Herring, 1992, Clementini and Di Felice 1994). Although the topology is considered the most appropriate mechanism to describe spatial relationships, the study on other mathematical frameworks continues. Billen et all 2002 propose another framework (i.e. the Dimensional model) for representing spatial relationships, built up in affine space and convexity properties of the constructing elements (named dimensional elements). The Dimensional model allows larger variations in grouping spatial relationships compare to the 9-intersection model.

Data Presentation: Advances in the area of computer graphics have made visual media a major ingredient of the current interface in the communication and interaction with computers. Therefore the research related to the visualisation of real world 3D data is mostly “shifted” to the computer graphics society. Many viewers and browsers as stand-alone applications and plug-ins have been developed to quickly visualise and navigate through 3D models for a variety of applications. New algorithms and implementations are reported daily. The design criteria, however, are fast rendering techniques based on internal structures rather than utilisation of database representations. TerraExplorer (SkyLine, 2002), the current leader for visualising large 3D textured data from real world and the first software with acceptable performance, also requires re-structuring of data.

Another significant area of 3D GIS research is devoted to Web applications. The Web has already shown a great potential in improving accessibility to 2D spatial information (raster or vector maps) hosted in different computer systems over the Internet. 3D data were even not transferable over the Web until five years ago. The first attempt to disseminate and explore 3D data, i.e. VRML, appeared to be rather “heavy” for encoding real geo-data due to the lack of a successful compression concept. Despite the drawbacks, the language became a tool for research visualisation. Researchers could concentrate on data structuring and analysis and leave the rendering issues to browsers offered freely on Internet. The research on spatial query and 3D visualisation utilising VRML has resulted in a few prototype systems (see Coors and Jung 1998, Lindenbeck and Ulmer 1998, Zlatanova2000). GeoVRML (VRML extended with geo-nodes) and Geographic Modelling language (GML) are another promising opportunities for representing 3D data on the Web. Based on XML concepts, GML provides larger freedom, flexibility and operability than VRML.

4 DISCUSSION

In this paper we reported our study on current 3D GIS status considering developments reported by vendors and researchers. The major 3D progress is observed in the area of data presentation. All traditional GIS vendors provide extended tools for 3D navigation, animation and exploration. However, still many of these systems are lacking full 3D geometry for 3D representation. OpenGIS specifications seem to be adopted rapidly by DBMS&CAD&GIS developers. In this order, the understanding for GIS is changing. Instead of a monolith, desktop, individual system, GIS is becoming an integration of strong database management (ensuring data consistency and user control) and powerful editing and visualisation environment (inheriting advanced computer graphics achievements). At present, only the first step is made, i.e. the implementations focus mostly on the geometry. 2D topological representations and operations are intensively in process of implementation. The third dimension with respect to topological issues is still in the hands of the researchers.

The case study clearly showed the benefits of a standardised spatial data structuring as well as revealed the very early stage of the integration. The large number of specialised settings, the application dependent feature-geometry linkage, the limited semantic hierarchy, the spatial operators utilising only X, Y values, are some of the issues that need further improvements.

Although, quite significant number of works devoted to 3D data structuring, the research is concentrated around few basic ideas, as the level of explicitly described spatial relationships varies. Each suggested structure exhibits efficiency and deficiency with respect to a particular applications and operations to be performed. Still 3D GIS functionality to be addresses: 3D buffering, 3D shortest route, 3D inter-visibilities are some of the most appealing for research. Integration of object-oriented approaches with the 3D GIS raises research topics at a database level toward standard object descriptors and operations.

3D visualisation within 3D GIS requires appropriate means to visualise 3D spatial analysis, tools to effortlessly explore and navigate through large models in real time. Observations on the demand for 3D City models show user preferences for photo-true texturing. Trading photo-true texture brings up necessities to store parameters for mapping onto the geometry.

ACKNOWLEDGEMENTS

We express our sincere gratitude to Kees van Proojen from Bentley Systems Europe, The Netherlands for his respective contributions to the case study.

REFERENCES

Abdul-Rahman, A., 2000. The design and implementation of two and three-dimensional triangular irregular network (TIN) based GIS. *PhD thesis*, University of Glasgow, Scotland, United Kingdom, 250 p.

Billen, R., S.Zlatanova, P. Mathonet and F. Boniver, 2002, The Dimensional model: a framework to distinguish spatial relationships, In: *Proceedings of Spatial Data Handling*, 8-12 July, Ottawa, Canada (to be published)

Cambray, de B., 1993. Three-dimensional (3D) modelling in a geographical database, In: *Proceedings of AutoCarto 11*, Bethesda, Maryland, pp. 338-347

Carlson, E., 1987, Three-dimensional conceptual modelling of subsurface structures, In: *Technical Papers of ASPRS/ACSM Annual Convention*, Baltimore, Vol. 4 , pp. 188-200

Clementini, E. and Di Felice P., 1994, A comparison of methods for representing topological relationships, In: *Information Sciences (80)*, pp. 1-34

Coors, V. and V. Jung, 1998, Using VRML as an Interface to the 3D Data Warehouse, In: *Proceedings of VRML'98*, New York

De La Losa, 1998, Toward a 3D GIS: Conception of a 3D GIS with a complete topological management, In: *Proceedings of GIS PlaNET'98 Conference,* Lisbon, Portugal

Egenhofer, M. J. and J. R Herring, 1992, Categorising topological relations between regions, lines and points in Geographic databases, *Technical report 94-1, NCGIA*, University of California.

ESRI, 1997, Using ArcView 3D Analyst. *ESRI Publication*, Redlands, California, USA, 118. p.

Lindenbeck, C. and H. Ulmer, 1998, Geology meets virtual reality: VRML visualisation server applications, In: *Proceedings of WSCG'98*, Vol. III, 3-19 February, Plzen, Czech Republic, pp. 402-408

Microstation GeoGraphics iSpatial, 2002, Bentley Systems Inc., available at: http://www2.bentley.com/products/default.cfm

Molenaar, M., 1992, A topology for 3D Vector Maps, In: *ITC Journal*, 1992-1, pp. 25-33

Oosterom, P. v, J. Stoter, W. Quak and S. Zlatanova, 2002, The balance between geometry and topology, In: *Proceedings of Spatial Data Handling*, 8-12 July, Ottawa, Canada (to be published)

OpenGIS specifications, 2002, OpenGIS Consortium, Inc., available at http://www.opengis.org/

Oracle Spatial 8i, 2002, User's guide and reference, available at: http://technet.oracle.com/doc/inter.815/a67295/toc.htm

Pfund, M., 2001, Topologic data structure for a 3D GIS, In: *Proceedings of ISPRS*, Vol.34, Part 2W2, 23-25 May, Bangkok, Thailand pp. 233-237

Pigot, S., 1995, A topological model for a 3-dimensional Spatial Information System, *PhD thesis*, University of Tasmania, Australia

Pilouk, M., 1996. Integrated modelling for 3D GIS. *PhD thesis*, ITC, The Netherlands, 200 p.

Raper, J, and D. J. Maguire, 1992, Design models and functionality in GIS. *Computer & Geosciences*, Pergamon Press, Vol. 18, No. 4, pp. 387-394

SkyLine, Software Systems, Inc., available at:
http://www.skylinesoft.com/corporate/corporate_home.asp

Sonnen, D. and H. Morris, 2000, Location in CRM: Linking virtual information to the real world, IDC White Paper, 12p., available at: http://www.idc.com

Stoter, J. and P. van Oosterom, 2002, Incorporating 3D geo-objects into a 2D geo-DBMS, *Proceedings of ASPRS/ACSM*, 19-26 April, 2002, Washington, USA, (to be published)

Zlatanova, S., 2000, 3D GIS for urban development, *PhD thesis*, ITC, The Netherlands, 222 p.

AN EMPIRICAL EVALUATION OF HEDONIC REGRESSION MODELS

Xiaolu Gao[a, *], Yasushi Asami[b], and Chang Jo Chung[c]

[a] JST Domestic Research Fellow, Dept. of Urban Research, National Research Institute for Land and Infrastructure Management. Tachihara 1, Tsukuba, Ibaraki 305-0802, JAPAN - gao-x925t@nilim.go.jp.
[b] Center for Spatial Information Science, the University of Tokyo. Hongo 7-3-1, Bunkyo, Tokyo 113-8656, JAPAN - asami@csis.u-tokyo.ac.jp.
[c] Spatial Data Analysis Lab., Geological Survey of Canada, Natural Resources Canada. 601 Booth Street, Ottawa, Ontario, CANADA K1A 0E8 - chung@gsc.nrcan.gc.ca.

Commission IV, WG IV/1

KEY WORDS: Empirical evaluation, Cross-validation, Prediction, Spatial model

ABSTRACT:

Statistical criteria are poor at evaluating spatial exploratory models of hedonic regression because they are heavily dependent on uncertain assumptions concerning spatial relationships. To resolve this problem, empirical evaluation methods are proposed in this paper. A simple linear model and three spatial models for a housing and land price dataset in Tokyo are studied for illustration. The prediction power of models is emphasized. With a cross-validation technique, the housing and land price at each sample point are predicted, then, numerical and graphical criteria are supposed using these predicted values and real observed prices. The methods not only help us to select suitable models for a dataset, but also provide an alternative for the significance test of concerned spatial relationships.

1. INTRODUCTION

Recently, a variety of hedonic regression models have been proposed in addition to a simple formed model to study the spatial nature of variables. All of them proposed to make use of the spatial characteristics of variables to improve models. Since they are often exploratory models, it is critical issue as how well a model is, whether the spatial characteristics identified by a model are convincible, and how to select the most suitable model for a dataset.

A difficulty in making such judgment is the lack of appropriate evaluation techniques. This paper sets out to address this issue.

Traditionally, statistical testing criteria such as *R*-square, *t*-statistic, *F* statistic, AIC (*Akaike Information Criterion*, a tradeoff of likelihood and the number of estimated parameters), and so on, are often used to evaluate regression models. However, models are always based upon such assumptions as linear assumption, normality, or non-collinearity, thus statistical testing methods inevitably depend on these assumptions. For exploratory models, this creates problems because assumptions are open for test.

To overcome this problem, other evaluation methods have to be developed.

2. ALTERNATIVE REGRESSION MODELS

In this paper, we use a housing and land price dataset in Tokyo for illustration. It covers the transaction price, structural attributes, environmental attributes, and *x*, *y* coordinates of 190 detached housing lots. See Appendix A for details of it.

2.1 A Simple Linear Regression Model

Equation (1) shows the simplest form of hedonic regression models:

$$y = a_0 + \sum a_k x_k + \varepsilon, \qquad (1)$$

where, y is unit price, x_k for $k = 1, 2, \ldots, m$ are independent variables, ε is an error term, and a_0 and a_k are parameters to be estimated.

An ordinary linear model in this form was developed in Gao and Asami (2001). It explains 75.6% of unit price, has 16 independent variables, all of which statistically significant at the level of *F* below 0.05. See Appendix B.

Since this model has not fully considered the impact of spatial location on unit price, to study the spatial characters of dataset with location data is thought of a way to improve it.

One of the spatial relationships being explored is spatial dependency. Another is spatial heterogeneity. The estimates of regression parameters in the presence of spatial dependency have been discussed by a number of literatures, *e.g.*, Dubin (1992; 1998) and Can (1990; 1992), while, various localized modeling techniques were proposed to capture spatial heterogeneity (Casetti, 1972; Getis and Ord, 1992; Fotheringham and Brunsdon, 1999). In addition, some models are developed to detect both of them (Anselin, 1988; Can, 1992).

---

[*] Corresponding author.

In this paper, we choose three spatial regression models to explore the data.

2.2 A Spatial Dependency Model

In order to study spatial dependency effects, a *prior probability* method proposed by Switzer *et al.* (1982) is applied. The model had been used to process satellite classification maps. To evaluate the class at the center of a window using the data for that location and the prior probability estimates obtained from the nearby observations in the window area was shown to have increased classification accuracy. Similarly, the prior characteristic terms of sample lots, denoted by x_k' for $k = 1, 2, \ldots, m$, are added to the simple linear model:

$$y = a_0 + \sum a_k x_k + \sum a_k' x_k' + \varepsilon. \qquad (2)$$

For simplification, let x_k' take the value of x_k that is associated with the nearest neighbor of a sample. Appendix C gives the OLS estimates of this model, with an *R*-square of 0.801.

Although one might argue that correlations between x_k and x_k' may lead to unstable estimates, the question is secondary to the concern whether spatial dependency is present. If the absence of spatial dependency effects were demonstrated, the model itself would be improper, let alone its estimates.

2.3 A Geographically Weighted Regression (GWR) Model

To identify spatial variations in relationships, a GWR model proposed in Fotheringham, *et al.* (1998) is adopted. The model extends traditional regression framework by allowing parameters to be estimated locally so that the model in Equation (1) is rewritten as

$$y_i = a_{0i} + \sum a_{ki} x_{ki} + \varepsilon_i, \qquad (3)$$

where, a_{0i} and a_{ki} represent the values of a_0 and a_k at point *i*. In order to estimate the model, an observation is weighted in accordance with its proximity to point *i*. Let W_i be an $n \times n$ matrix whose diagonal elements w_{ij} denotes the *geographical weighting* of all observed data for point *i*, and the off-diagonal elements are zero. Data from observations close to *i* are weighted more than data from observations far away. Equation (4) gives the estimation of a_i:

$$\hat{a}_i = (X^T W_i X)^{-1} X^T W_i Y, \qquad (4)$$

where, *X* and *Y* are the matrix of explanatory variables and unit prices, respectively.

A weighting function in Equation (5) is applied:

$$w_{ij} = \exp(-\frac{d_{ij}^2}{\beta^2}). \qquad (5)$$

where, d_{ij} is the Euclidean distance between point *i* and *j*, β is a bandwidth. By minimizing (6), β is set to 1,250m.

$$\sum_i [y_i - \hat{y}_{\neq i}(\beta)]^2. \qquad (6)$$

Note that $\hat{y}_{\neq i}(\beta)$ is the estimates at *i* with samples near to *i* but not *i*.[*]

At the location of each of the 190 observations, regressions are run. The obtained localized parameter estimates exhibit a high degree of variability over space and demonstrate fairly complex spatial patterns of each variable.

The localized regressions result in high *R*-square fittings, beyond 0.95 at all sample points. Apparently, it is the outcome of localized regression, so these values are not comparable to that of the simple linear model.

2.4 A Spatial Dependency +GWR Model

In the following, we use a specification in (7) to investigate whether spatial dependency and heterogeneity effects are both present.

$$y_i = a_{0i} + \sum a_{ki} x_{ki} + \sum a_{ki}' x_{ki}' + \varepsilon_i, \qquad (7)$$

Similar to the spatial dependency model, let x_{ki}' represent the corresponding values of the nearest sample to *i*.

In addition, regressions are localized at each sample point with a GWR technique, adopting a weighting scheme of (5), and the optimal value of β being 1650m. The regression yields the estimates of parameters at each sample points.

3. AN EMPIRICAL EVALUATION OF THE MODELS

Now, let us discuss how to select a best one from these models. The weakness of "black-box" statistical evaluation method is well aware of. For instance, models supported by high *R*-square fitting or satisfying significance level are frequently doubted because of weak assumptions or irrational estimates of parameters, thus a clear conclusion can hardly be reached with these statistical criteria. This is an unavoidable result of heavy dependency on statistical assumptions. The problem is especially distinct in spatial exploratory models for the assumptions of models are to a large degree uncertain. This prompts us to turn to an empirical evaluation approach. In particular, the methods being presented in this paper focus on the prediction power of models.

In the practice of hedonic regression, we often focus more on estimating unbiased coefficients to identify the marginal prices of explanatory variables, rather than simply see the prediction powers. Nonetheless, when spatial relationships are introduced to a simple ordinary model, we instinctively expect that this can improve the fitting of the model. Thus conversely, if it does not work better at predicting prices, we may just be satisfied with the simple model and think that its estimates are robust. From this viewpoint, the improvement of fittings can be been seen as an alternative to test the significance of spatial relationships, too. (Can, 1992; Dubin, 1992 and 1998; Brunsdon, *et al.*, 1999)

---

[*] Suppose that point *i* itself is also included to estimate a_{0i} and a_{ki}. If β is very small, the weightings of all points except for *i* become negligible and the estimates will fluctuate wildly throughout space.

A cross-validation approach is used to carry out the idea, *i.e.*, to see how well the prices already being observed can be predicted with the rest samples. An application of a similar method can be seen in Bourassa *et al.* (2001), where, random 80% samples are modeled to predict the rest 20% samples, and the sum of square of error is used as a criterion to evaluate proposed model. However, we are concerned that, to estimate a hedonic regression model, it is important to keep as large a sample size as possible. Thus, we predict the price at every point with all the rest 189 samples.

3.1 Numerical Cross-validation Criteria

First, consider some numerical criteria, for instance, $\Sigma(y_i - \hat{y}_{\neq i})^2$ and $\Sigma|y_i - \hat{y}_{\neq i}|$*, where, y_i and $\hat{y}_{\neq i}$ denote the observed and the predicted unit price at point *i*, respectively.

Table 2 gives their calculations. The values of all the three spatial models are slightly lower than that of the simple linear model. However, the falloffs are too small to suggest any critically large improvements.

Model	Numerical criteria	
	$\Sigma(y_i - \hat{y}_{\neq i})^2$	$\Sigma\lvert y_i - \hat{y}_{\neq i}\rvert$
Simple linear model	1.96	15.29
Spatial dependency model	1.92	15.11
GWR model	1.81	14.60
Spatial dependency +GWR model	1.76	14.33

Table 2. A comparison of models by numerical criteria

Centering on prediction error, more numerical criteria could be raised. Seriously though, numerical criteria just show a general aspect of model. Thus, some graphical criteria are further employed in order to illustrate broader and more complicated aspects of the relationships between model and data.

3.2 The Distribution of Observation and Prediction

On a scatter plot of observed unit price (y_i) and predicted unit price ($\hat{y}_{\neq i}$), the prediction power of a model is indicated by the range occupied by the scatter points, in particular, the short breadth of this range. Obviously, all points being on 45° diagonal line indicates a perfect prediction, and higher prediction power is suggested by how slim and how close the range is to the 45° line.

Following the idea, the three spatial models are compared with the simple linear model in Fig. 1~3. The 95% confidence regions of the scatter points are plotted, respectively. As results, again, all the spatial models are no much better than the simple linear model for this dataset.

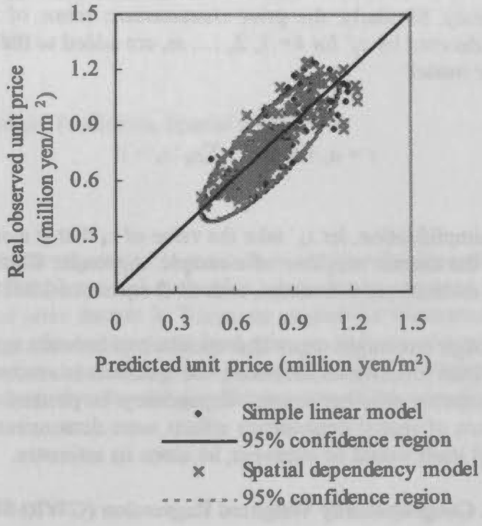


Figure 1. Simple linear model and spatial dependency model

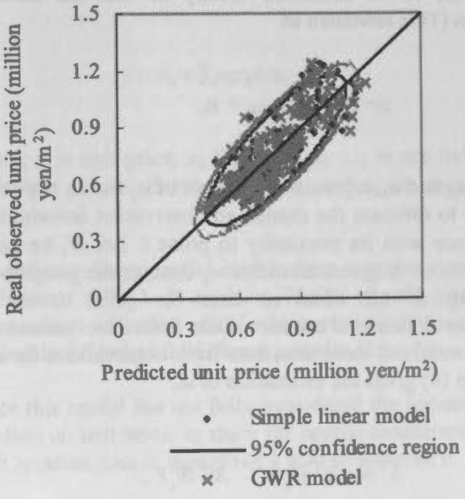


Figure 2. Simple linear model and GWR model

---

* This is deemed more robust than $\Sigma(y_i - \hat{y}_{\neq i})^2$.

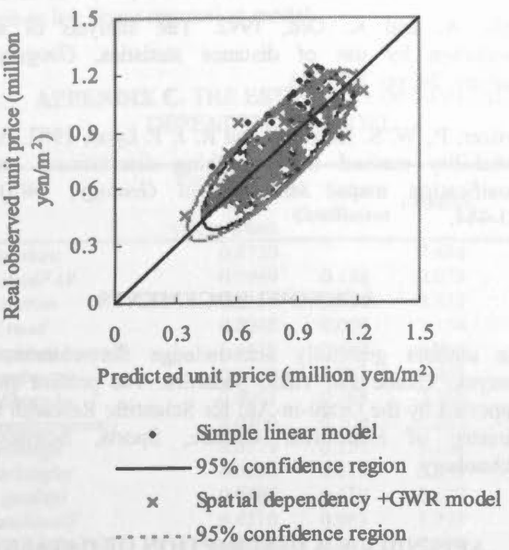


Figure 3. Simple linear model and spatial dependency +GWR model

3.3 Prediction Rate Curve

We further evaluate the models by the proportion of well-predicted samples with respect to prediction error, which is referred to as prediction rate.

To do this, let $\delta_i = |y_i - \hat{y}_{\neq i}|$ and sort δ_i. Then plot the ratio of well-predicted samples with sorted δ_i as horizontal axis. By this way, the prediction rate curve of a model is obtained.

The introduction of this criterion allows us to capture the comprehensive performance of a model. In particular, it is helpful when we select models under a given level of prediction error. Consider two models having prediction rate curves shown in Fig. 4. Highly possibly, the sums of prediction errors of them are close. At tolerance level ξ_1, 70% of samples are correctly predicted by model 1, while only 60% can be predicted by model 2. However, if tolerance ξ_2 is acceptable, more samples are correctly predicted by model 2.

From an empirical point of view, a dataset always has some poor-to-predict observations. Thus, the performance of the prediction curves within a tolerable error scope may be thought of more important.

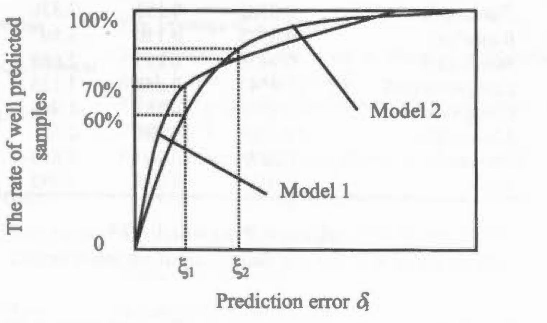


Figure 4. Evaluating models with prediction rate curves

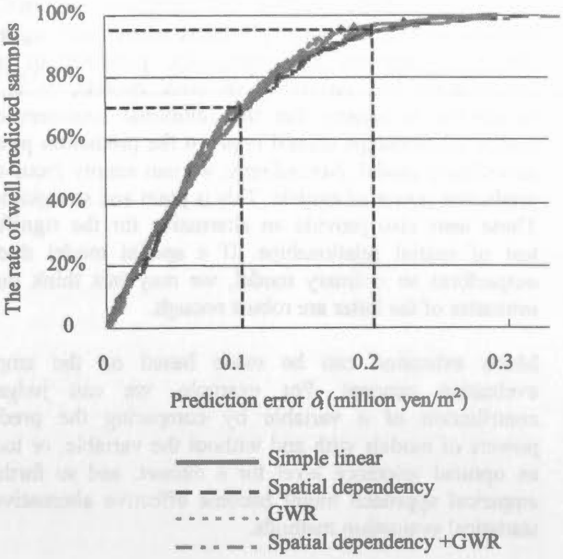


Figure 5. A comparison of prediction rate curves

Prediction rate curve actually indicates the density of the scatter points in Fig. 1~3 with respect to their distance to the 45° diagonal line. To illustrate, let us see the prediction rate curves of the simple linear model and the spatial models in Fig. 5.

The four curves twist together for most of the span. This implies that their performances are fairly alike. All of them predict the unit price of about 70% of samples if a tolerance of 0.1 million yen/m² is accepted. Within a tolerance of 0.2 million yen/m², about 95% are correctly predicted. With the prediction error being larger than 0.2 million yen/m², only about 5% of the sample can be predicted.

On the other hand, it might have been noted that the tail of the simple linear model winds up a bit earlier than the other models. This is able to explain why this model looks better than the others in Fig. 1~3. Nevertheless, we can still draw the conclusion that they are similar, allowing that the tail part is not very important.

From the above empirical tests, it seems proper to say that, for this dataset, the proposed spatial dependency model, GWR model, and the mixed model of the two are not significantly better than a simple linear model. In other words, the simple linear model is fairly satisfactory. Moreover, we should not stick to the spatial relationships revealed by the above spatial models.

The failure of these three spatial models might either result from inappreciable spatial relationships in the sample area, or result from inappropriate model specifications. In the latter case, alternative specifications might be constructed and tested with similar empirical techniques as described above.

4. CONCLUSION

In the sense that empirical criteria are not dependent on statistical assumptions, they fit for any dataset and their application is not restricted to a certain kind of models.

Nonetheless, these methods are especially outstanding to evaluate spatial exploratory models of hedonic regression, where assumptions concerning the presence of spatial relationships are uncertain. In such models, it is quite reasonable to believe that the additional consideration on spatial relationships should improve the prediction power of an ordinary model. Accordingly, we can simply focus on the prediction power of models. This is plain and straightforward. These tests also provide an alternative for the significance test of spatial relationships. If a spatial model does not outperform an ordinary model, we may just think that the estimates of the latter are robust enough.

Much extension can be made based on the empirical evaluation concept. For example, we can judge the contribution of a variable by comparing the prediction powers of models with and without the variable, or look for an optimal tolerance level for a dataset, and so forth. The empirical approach might become effective alternatives for statistical evaluation methods.

REFERENCES

Anselin, L., 1988. Lagrange multiplier test diagnostics for spatial dependence and heterogeneity. *Geographical analysis*, 20, pp. 1-17.

Bourassa, S. C., M. Hoesli, and V. S. Peng, 2001. Do housing submarkets really matter? *AsRes Sixth Annual Conference*, August 2001, Tokyo, Japan.

Brunsdon, C., A. S. Fotheringham, and M. Charlton, 1999. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 39(3), pp. 497-524.

Can, A. 1990. The measurement of neighborhood dynamics in urban house prices. *Economic Geography*, 66, pp. 254-272.

Can, A., 1992. Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, 22, pp. 453-474.

Casetti, E., 1972. Generating models by the expansion method: applications to geographic research. *Geographic Analysis*, 4, pp. 81-91.

Dubin, R. A., 1992. Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics*, 22, pp. 433-452.

Dubin R. A., 1998. Spatial autocorrelation: a primer. *Journal of Housing Economics*, 7, pp. 304-327.

Fotheringham, A. S. and C. Brunsdon, 1999. Local forms of spatial analysis. *Geographical Analysis*, 31(4), pp. 341-358.

Fotheringham, A. S., M. E. Charlton, and C. Brunsdon, 1998. Geographical weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30, pp. 1905-1927.

Gao, X. and Y. Asami, 2001. The external effects of local attributes on living environment in detached residential blocks in Tokyo. *Urban Studies*, 38(3), pp. 487-505.

Getis, A. and K. Ord, 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, pp. 189-206.

Switzer, P., W. S. Kowalik, and R. J. P. Lyon, 1982. A prior probability method for smoothing discriminant analysis classification maps. *Mathematical Geology*, 14(5), pp. 433-444.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the comments by Atsuyuki Okabe and Yukio Sadahiro. The present study is supported by the Grant-in-Aid for Scientific Research by the Ministry of Education, Culture, Sports, Science and Technology.

APPENDIX A. A DESCRIPTION OF DATASET

The database used in this paper was developed in Gao and Asami (2001). It involves 190 detached housing lots in western Tokyo. Sample area is small, about 110 hectare. Sample properties were drawn from the Oct. 1996 to Sep. 1997 issues of *Weekly Housing Information* magazine, which provides transaction price and basic information of land and houses in Japan.

In addition to the information provided by the magazine, details of sample properties, such as road, sunshine, distances to public facilities, and neighborhood environments, were obtained through site survey, GIS applications, and so on. The dataset also includes the location information of sample lots, represented by the *x* and *y* coordinates of their center point.

APPENDIX B. THE ESTIMATES OF A SIMPLE LINEAR MODEL

	Coefficient (million yen/m^2)/unit	Std. Coefficient	*t*-statistic	Sig. level
Constant	0.9115		9.165	.000
ActualFAR	0.1276	0.182	3.215	.002
T.station	-0.0157	- 0.420	- 9.612	.000
W.road	0.0209	0.137	2.855	.005
Residual.b.age/S	0.5686	0.401	6.419	.000
Landscape	-0.1726	- 0.400	-8.463	.000
T.Shinjuku	-0.0168	- 0.310	-6.596	.000
Frontage	0.0058	0.113	2.383	.018
Goodpavement	0.0420	0.114	2.798	.006
Parkinglot	0.0382	0.153	3.536	.001
B.quality1	0.0575	0.150	3.507	.001
Sunshine/S	0.9476	0.125	2.669	.008
Con.greenery/S	21.4547	0.268	3.138	.002
Con.greenery	-0.1956	-0.257	-2.968	.003
Mix-use3/S	-17.4766	-0.381	-2.438	.016
Mix-use3	0.2384	0.404	2.635	.009
Tree1	0.0335	0.085	1.992	.048

R-square: 0.756, adjusted *R*-square: 0.734.
See Appendix D for the definition of variable names.

Statistical tests showed that the coefficients were quite stable, suggesting that the model does not have serious multi-collinearity problems. In addition, the model was shown to be superior to some other simple formed models such as log linear regression model.

APPENDIX C. THE ESTIMATES OF SPATIAL DEPENDENCY MODEL

	Coefficient (million yen/m^2)/unit	Std. Coefficient	t-statistic	Sig. level
Constant	0.8720		7.464	.000
ActualFAR	0.0949	0.136	2.073	.040
T.station	-0.0117	-0.311	-3.532	.001
W.road	0.0048	0.069	1.154	.250
Residual.b.age/S	0.6610	0.467	6.910	.000
Landscape	-0.4670	-1.082	-4.453	.000
T.Shinjuku	-0.0091	-0.168	-2.363	.019
Goodpavement	0.0400	0.108	2.269	.025
Frontage	0.0779	0.151	3.102	.002
Parkinglot	0.0296	0.118	2.559	.011
B.quality1	0.0686	0.179	3.637	.000
Sunshine/S	0.6310	0.083	1.773	.078
Con.greenery/S	34.4830	0.431	4.172	.000
Con.greenery	-0.2760	-0.363	-3.925	.000
Mix-use3/S	-13.7990	-0.301	-1.879	.062
Mix-use3	0.2070	0.351	2.147	.033
Tree1	0.0209	0.053	1.122	.264
NN-actualFAR	0.0299	0.044	0.680	.497
NN-t.station	-0.0033	-0.085	-0.978	.330
NN-w.road	0.0099	0.068	1.026	.306
NN-residual.b.age/S	-0.1080	-0.074	-1.127	.262
NN-landscape	0.2780	0.650	2.629	.009
NN-t.Shinjuku	-0.0074	-0.138	-1.941	.054
NN-goodpavement	0.0242	0.065	1.338	.183
NN-frontage	0.0015	0.031	0.572	.568
NN-parkinglot	0.0080	0.032	0.725	.469
NN-b.quality1	-0.0266	-0.070	-1.433	.154
NN-sunshine/S	0.6290	0.073	1.582	.116
NN-con.greenery/S	-13.7210	-0.163	-1.435	.153
NN-con.greenery	-0.0011	-0.001	-0.014	.989
NN-mix-use3/S	-8.8210	-0.174	-0.787	.432
NN-mix-use3	0.0914	0.145	0.650	.517
NN-tree1	0.0069	0.017	0.324	.746

NN- indicates the prior characteristic terms estimated from the nearest neighbor of sample lots.
R-square: 0.801, adjusted R-square: 0.761.
See Appendix D for the definition of variable names.

APPENDIX D. THE DEFINITION OF VARIABLES

Variable name	Meaning (unit)
ActualFAR	Building floor area / lot area (ratio)
Con.greenery	Adjacent to public green space, if true, 1, otherwise 0
Frontage	The frontage of a lot (m)
Goodpavement	The pavement of the road in front of a lot being good, if true, 1, otherwise 0
Landscape	Within designated landscape areas, if true, 1, otherwise 0
Mix-use3	Intensive mixed land use, if true, 1, otherwise, 0
Parkinglot	The count of parking lots (number)
Residual.b.age	Residual building age (year)
S	The size of a lot (m^2)
Sunshine	Sunshine duration (hour)
Tree1	Greenery in the district being good, if true, 1, otherwise 0
T.Shinjuku	Time distance to central city area (minute)
T.station	Time distance to the nearest station (minute)
W.road	The width of the road in front of a lot (m)

GIS DATA MAINTENANCE AND MANAGEMENT WITH SPATIOTEMPORAL MODEL

W. Lu [a], *, T. Doihara [a]

[a] Research Institute, Asia Air Survey Co. Ltd., Tamura 8-10, Atsugi, Kanagawa, Japan - (luwei, ta.doihara)@ ajiko.co.jp

Commission IV, WG IV/1

KEY WORDS: Spatiotemporal GIS, Data Maintenance, Dynamic Change, KIWI+ Format

ABSTRACT:

Recently, GIS related research and applications have been making rapid progress. Computers with faster CPU and larger memory enable practical applications of time consuming and memory intensive algorithms, both for spatial analysis and visualization. Storage devices become cheaper and have larger space to store data of larger volume and more varieties including image, sound and video. For GIS data capture, novel technologies such as GPS, laser range finder, synthesize aperture radar are getting more and more popular, which bring us terrain data of higher precision at lower cost. Yet, our living environment is also changing rapidly, which makes it a task of high cost and complexity to maintain and manage geometric data such as cartographic or utility information. This paper proposes a system for efficient GIS data maintenance and management. It makes use of a spatiotemporal data model for describing various types of dynamic changes of GIS data, manipulating data in multidimensional space including time. The experiments performed on the initial implementation of the proposed system show its ability of efficient and accurate maintenance of dynamic GIS data.

1. INTRODUCTION

Recently, GIS related research and applications have been making rapid progress. Algorithms for spatial access of GIS data has been thoroughly studied and put into practical use(Rigaux 2002).Computers with faster CPU and larger memory enable practical applications of time consuming and memory intensive algorithms, both for spatial analysis and visualization. Storage devices become cheaper and have larger space to store data of larger volume and more varieties including image, sound and video. For GIS data capture, novel technologies such as GPS, laser range finder, synthesized aperture radar (SAR) are getting more and more popular, which bring us terrain data of higher precision at lower cost (Lu 2001). Yet, our living environment is also changing rapidly, which makes it a task of high cost and complexity to update the initial data, especially those such as cartographic or utility information, of a GIS system according to the latest changes, because of the complexity in spatiotemporal changes.

The active study of GIS in the spatiotemporal aspect has a history of only less than 10 years (Al-Taha 1994, Castagneri 1998, Yuan 1996, Ott 2000). Therefore even though many practical models have been proposed and are adopted in various types of practical systems, there are still very few commercial systems that offer handy processing abilities for time-integrated applications. Furthermore, the up to date researches related to spatiotemporal issues all emphasize on description of dynamic GIS data.

This paper proposes a system for efficient GIS data maintenance and management. It makes use of a spatiotemporal data model for describing various types of dynamic changes of GIS data, manipulating data in multidimensional space including time. The system not only has the ability of spatiotemporal representation and manipulation, but also enables distributed data updating, detection of spatiotemporal errors during data updating. The experiments performed on the initial implementation of the proposed system show its ability of efficient and accurate maintenance of dynamic GIS data.

2. THE SPATIOTEMPORAL DATA MODEL

2.1 The Data Schema

The spatiotemporal data model is based on the spatio-temporal object model proposed in the spatiotemporal GIS DiMSIS [kakumoto 1997, Hatayama(a) 2001], which adopts an implicit destription and calculation data model. In this model, all the geoemtric objects are described by elements of polyline (vector) and point (connector), each having a tuple of time factors as follows:

Time ::= { GS, GE, ES, EE }

where, G stands for generation, E extinction, S start and E end respectively. Fig. 1 illustrates the concept of time tuple, which enables the description of the chaning process of an spatial object, from the beginning of generation, end of generation, beginning of extinction to end of extinction.

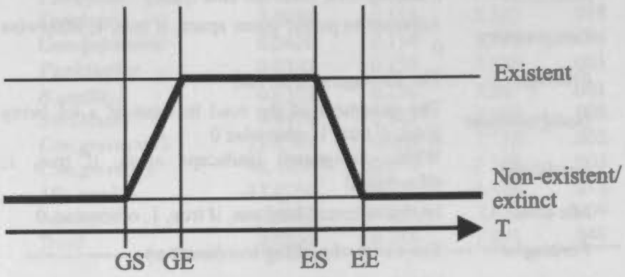


Fig.1 The concept of time tuple

---

* Corresponding author

This model is "implicit" since it has no duplicated vectors. Face objects are described by a representitive point (connector) locationed inside it. The actual boundary data of the face will be extracted from the related connector whenever necessary.
Fig.2 illustrates an example of description for dynamic spatial object with the above data model. The polylines between intersections are represented as vectors. The faces such as cadastral boundary, building shape etc are represented by connectors inside them. The value of time is represented by the nubmer of days since the initial setting time of the database.
The changes of spatial objects are recorded by their related vectors and connectors with appropriate time values. In this case, the boundary of a land is represented in dashed line. It used to have a single building before the time 1003. But after the time 1003, the land was divided into two parts, and the old building was replace by two seperate smaller buildings. The faces of old boundary and the new ones are represented by three connectors with the appropriety set time tuple. Since the boundary in dashed line and its corresponding new boundaries overlap with each other, only the dashed line is drawn here. We can notice that the chang of boundary are not simply by deletion of the data, but by modification of the extinction time. In this way, the temporal event/history can be maintained with much fewer overhead than models such as time slice or space-time composite.

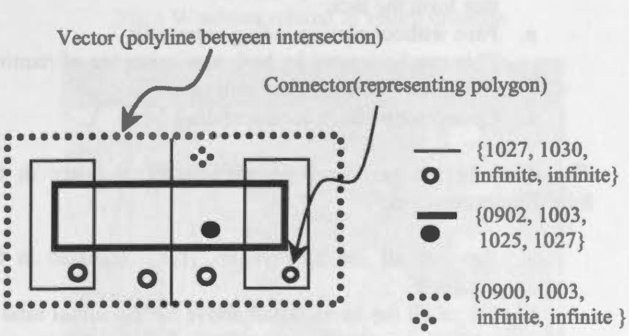

Vector (polyline between intersection)
Connector(representing polygon)
{1027, 1030, infinite, infinite}
{0902, 1003, 1025, 1027}
{0900, 1003, infinite, infinite }

Fig.2 Dynamic object descrition with implicit spatiotempral model

Among the attributes of vector and connector, there is the relation connector, which indicates inheritance, include and group relationship with other objects. Each connector can also store a text key of varible length for recording necessary informatino, such as address, statistical value, the index to an relational database.

2.2 The Description Ability of Spatiotemporal Events

With the above data mode, we can implemented the following descriptors for spatiotemporal events

a. predecessor/successor
   this can be realized by insertion of connectors with the relation attribute
b. discrete change
   this can be realized by setting GS, GE, ES and EE to the same specified time, in combination of predecessor and successor descriptions
c. stepwise change,
   this can be realized by setting GS, GE, ES and EE
d. continuous change (linear)
   this can be realized in the same way as stepwise with the temporal interpolation algorithms[Zhang 1999]

2.3 The Spatiotemporal Operations

With the same data model, we can implemented the following spatiotemporal operators:
record event of generation, by setting GS and GE respectively (ES and EE are infinite)

a. record event of extinction
   this can be realized by setting ES and EE respectively
b. record event of discrete modification
   this can be realized by the combination of recording event of extinction of old object and generation of the modification object.
c. retrieve objects existent in the specified duration
d. retrieve objects existent before/after an specified time
e. retrieve objects extinguished/modified in the specified duration
f. generate the temporal order of a serious of objects

3. The MAINTENANCE SYSTEM FOR DYNAMIC GIS

The main purposes of this system are:

a. Reduction of the overhead of data maintenance history
   The traditional maintenance systems only keep versions (snap shots) of certain interval. Management of these versions cost more resource and sometimes cause confusions.
b. Real time updating
   For applications such as those in tax related department, the actual update of real estate can only be performed once in a year. The traditional maintenance system is only feasible when all of the spatial changes during this period are collected, which will inevitably increase the work load of the maintenance department and generate a time lag.
c. Improvement of data quality
   In traditional data maintenance systems, it is difficult to keep track of the changes during data updating. As a result, careless mistakes, data contradictions are hard to avoid, and consequently, it cost more to maintain data quality.
d. Spatiotemporal analysis
   Since traditional data maintenance systems can only offer snap shots, spatiotemporal analysis that requires temporal events during the snap shots are impossible.

The details of this system will be presented in the following sections.

3.1 System Configuration

We have implemented a prototype system for data maintenance with the spatiotemporal data model described in section 2. The system configuration is shown in Fig. 3. The system consists of two main components: the kernel for fundamental processing and manipulation of spatiotemporal data, the graphical user interface for applications of designated specifications.
The major modules are as follows

a. Graphical user interface: offers the operator with easy to under stand menus and functions for spatiotemporal operations
b. Spatiotemporal interpreter: translates the requires from the operator into spatiotemporal descriptions
c. Spatiotemporal processing: realizes the spatiotemporal descriptions with the spatiotemporal operators

d. Error detection: find and visualization of and temporal contradiction

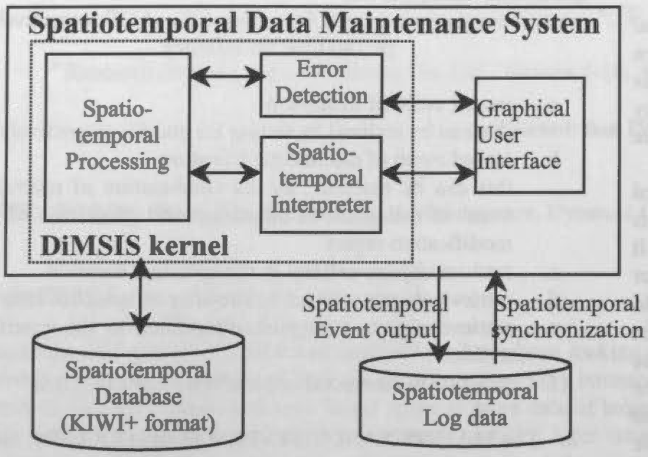

Spatiotemporal Data Maintenance System
Spatio-temporal Processing
Error Detection
Spatio-temporal Interpreter
Graphical User Interface
DiMSIS kernel
Spatiotemporal Event output
Spatiotemporal synchronization
Spatiotemporal Database (KIWI+ format)
Spatiotemporal Log data

Fig.3 The system configuration

To enable distributed or real time data maintenance operations, the spatiotemporal changes can either be reflected directly into the main database, or output into a log file as spatiotemporal events, which will be imported into the main data base when necessary. This implementation also enables real time updating of the database without affecting the main database. The later case is often required by organizations such as local municipal governments, where the daily official works rely on data that are updated in certain period, typically once every year. Updating all the spatiotemporal changes at time will cause a time lag in the freshness of the database. By updating spatiotemporal changes in real time base and outputting them as spatiotemporal events, the user can achieve instant synchronization with the latest situation when the time of database update comes.

3.2 The Details of Individual Modules

3.2.1 Graphical user interface

Offers the operator with easy to understand menus and functions for spatiotemporal operations. Some of the typical functions are:

a. Creation of new spatial objects
b. Modification of existing objects
c. Temporal spatial query
d. Settings for display, modification etc.
e. Exporting and importing temporal events in the form of log files

3.2.2 Spatiotemporal interpreter

Translates the requirements from the operator into spatiotemporal descriptions. Some of the typical examples are:

a. Deletion of objects:
   - set the ES and EE of the related objects.
b. Move/Modification-of-shape:
   - set the GS and GE of the original object.
   - create a new object with GS, GE, set to
     ($GS_{orginal}$, $GE>=ES_{orginal}$,ES=Infinite,EE=Infinite)
c. Snap shot
   - change the display time of the system

3.2.3 Spatiotemporal processing

Realizes the spatiotemporal descriptions with the spatiotemporal operators described in section 2.3. It also includes the fundamental GIS functions such as displaying the spatial objects in zoom-in/zoom-out mode, or panning the objects, changing the display colour/style etc.

3.2.4 Error detection

This module finds and visualizes both spatial and temporal contradictions/errors. Especially while maintaining historical data in the same database can help to improve the quality of data modification, it can also cause new types of errors or contradictions because of wrong setting of time values. Some of the typical functions of error detection are:

a. Duplicated vectors
   This can be caused by both negligence of deleted vectors and wrong time setting.
b. Floating terminal
   This can be caused by both incorrect snap operation and wrong time setting.
c. Intersection of vectors
d. Incomplete face
   This can be caused by both wrong operation in polygon formation or wrong time setting for vectors that form the face.
e. Face without corresponding connector
   This can be caused by both non-existence of required connector or wrong time setting
f. Connector without corresponding face

The checking are performed in spatiotemporal space in the following procedures:

a. Retrieve all the time values $\{t_0,t_1,\ldots,t_n\}$used in the database
b. Check all the items stated above for the initial time t_0, which is usually the set up time of the database.
c. For each time $t_i(0<i<n)$, find the objects with the time satisfying $GE<=t_i<=ES$, and check them against those that exist by time t_i

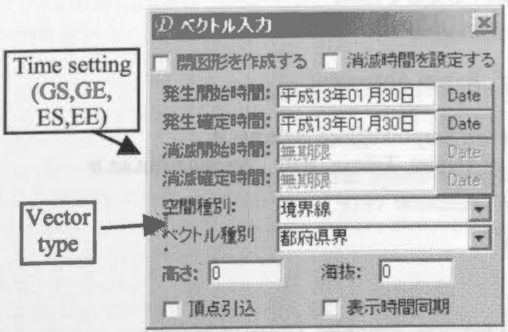
3.3 The Data Format

We have adopted KIWI+ as the data format of the system [Hatayama 2001(b)]. This format is the extension of the KIWI format under consideration as the standard of physical storage of ISO/TC204. Besides the implicit destription and calculation data model, it also enables a variety of attribute descriptions such as formated text string, symbol, relation and multimedia (attributes for presentation only; including image, vedio and sound). It also enables the asignment of the vertex of a vector as nodes for speedy searching of face from connectors. This schema can also fasten the processing of networks while maintaining precise description of their shape information.

4. The Performance Evaluation

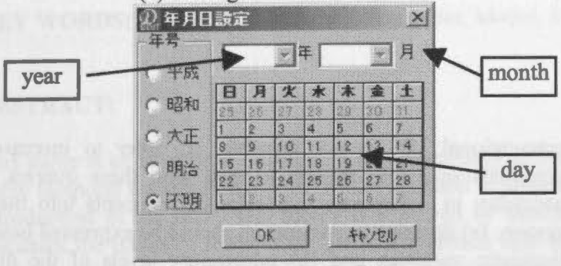
We have implemented the system based on the specifications stated so far. The kernel part is in the form of component and the graphical user interface is application dependant and is implemented in easy to customize language.
Fig.4 shows the screen of settings for vector creation and its related time setting window.
We have performed a serious of experiments with real world data to evaluate the effectiveness of the spatiotemporal data

model and the operators/descriptors. Fig. 5 shows an example of spatiotemporal data modification.



☐ 閉図形を作成する ☐ 消滅時間を設定する
発生開始時間: 平成13年01月30日
発生確定時間: 平成13年01月30日
消滅開始時間: 無期限
消滅確定時間: 無期限
空間種別: 境界線
ベクトル種別: 都府県界
高さ: 0 海抜: 0
☐ 頂点引込 ☐ 表示時間同期

(a) Settings for vector creation



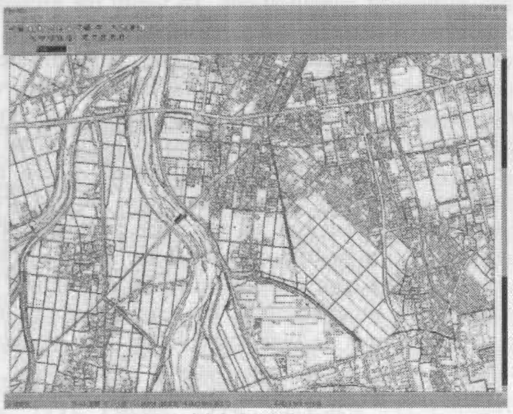
年号
○ 平成
○ 昭和
○ 大正
○ 明治
● 不明

年 月

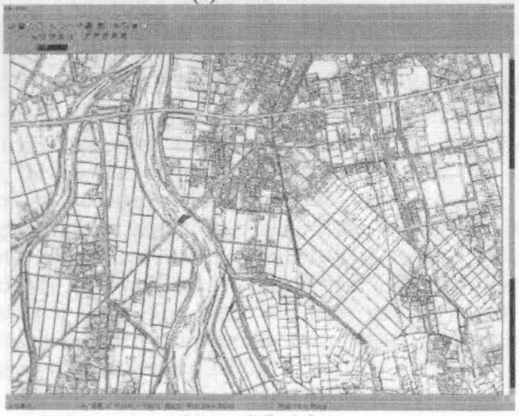
日	月	火	水	木	金	土
25	26	27	28	29	30	31
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
1	2	3	4	5	6	7

(b) time specification

Fig.4 Windows related to vector creation



(a) initial data



(b) modified data

Fig. 5 An example result of data maintenance

The results indicate the following advantages over systems of snap shot type:

a. Capability of essential spatiotemporal modification
b. Enables more intuitive data maintenance
c. Slimmer data size
d. Easier spatiotemporal data retrieval / analysis
e. High data quality with no spatiotemporal data contradictions
f. Capability of distributed data maintenance
g. Capability of data synchronization at the specified time
h. Faster reconstruction of face tracing (than older format of DiMSIS)
i.

5. Conclusions

In this paper, we have proposed a system for efficient maintenance and management of spatiotemporal data. In the initial implementation, we have only implemented modification-oriented functions. To make it easier for end users to manage the spatiotemporal database, more functions for spatiotemporal analysis are also in great demand. Further more, spatiotemporal interpolation other than linear types remains a challenge because of its complexity in scale and variety.

References from Journals:

Al-Taha KK, Snodgrass RT, Soo MD 1994: Bibliography on spatiotemporal databases. International Journal of Geographical Information Systems 8:95-103

Castagneri J. 1998: Temporal GIS explores new dimensions in time. GIS World 11

Lu W. 2001: Digital Terrain Models – An Overview of DTM Generation and Interpolation Issues, GIM International, Vol. 15, No. 12

References from Books:

Ott T. and Swiaczny F. 2000: Time-Intergrative Geographic Information Systems – Management and Analysis of Spatio-Temporal Data, Springer.

Rigaux P., Scholl M. and Voisard A. 2002: Spatial Databases – with Application to GIS, Morgan Kaufmann Publishers.

References from Other Literature:

Hatayama, M., Kakumoto, S. and Kameda, H. 2001a: Development of Dynamic Management -Temporal Information System and Application for Census Data – Toward Asian Temporal GIS(ST-GIS) (2) -, Proceedings of the 3dr ISPRS Workshop on Dynamic and Multi-Dimensional GIS, pp.123-127

Hatayama, M., Kakumoto, S., and Dohi, T. 2001b: Simple Topoloty & Temporal – Open Database Schema (ST2-ODS): Time Schema with KIWI+ Format, Proceedings of Japan Geographic Information Systems Association, Vol.10

Kakumoto, K., Hatayama, M., Kameda, H. and Taniguci, T. 1997: Development of Disaster Management Information System (DiMSIS), GIS'97 Conference Proceedings, p.595-598

Yuan M. 1996, Temporal GIS and spatio-temporal modelling. Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, Japnuary pp.21-26

Zhang, W. and Hunter G. J. 1999: Temporal Interpolation of ly Dynamic Objects, Proceedings of ISPRS, Vol.32, Part 4"12, "Dynamic and Multi-Dimensional GIS", Beijing, October 4-6

FUZZY BASED SPATIAL QUERY AND ANALYSIS IN AN URBAN INFORMATION SYSTEM

Mohammad Sarpoulaki[a], Farhad Samadzadegan[b], Rahim Ababspour[b]

[a]National Cartographic Center of Iran, Tehran, Iran e-mail: Sarpulki@ncc.neda.net.ir
[b] Dept. of Surveying Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran -samadz@ut.ac.ir

Commission IV, WG IV/1

KEY WORDS: Spatial Query, Information System, Fuzzy logic, Uncertainty, Linguistic variables, Decision Making

ABSTRACT:

With respect to the fact that nearly 60% of the world's population live in the cities and regarding the inevitable structural complexities as well as the cultural and economical complications in the cities, the intervention of a sophisticated information management systems has become a real necessity. This means that powerful city planning and management tools based on spatial information are a real need of the present situation of the world. Decision making from land administration and property valuation to simple house transactions as well as transportation decisions in a city needs spatial information management tools such as GIS and UIS (Urban Information System). An Urban information system is now internationally acknowledged to be a suitable tool for improving the management and planning potentials for the people who live in urban areas. In spite of the level of the sophistication and the services which are provided by GIS and UIS for many users in the cities, one of the main problems associated with them are the fact that these systems can answer only limited number of questions and hence a partial solution to the problem may be obtained. For example decision for purchasing a house may need to be made based on the following queries:

To be close to the shopping centers and schools,
To be far from the main roads,
To be far from the military camps,
To have a medium area size,
To have an affordable price.

With regard to the fact that there are certain level of uncertainties in these queries, incorporating them in a conventional UIS is not possible. In order to increase the potentials of a UIS system to deal with these queries, it is necessary to incorporate the following concepts into the UIS system: (a) descriptive parameters should be expressed based on linguistic variables and (b) uncertainty levels of the queries should be modeled. The method that is proposed in this paper takes advantage of the concept of the fuzzy logic to solve the above mentioned complications. To verify the proposed idea, the problem of selecting and purchasing an appropriate house is considered as a case study. To model this problem based on the fuzzy reasoning approach, in the first stage linguistic variables for decision making to buy a house is defined. In the next step the membership functions with respect to the linguistic rules for describing an ideal house is determined. These are then embedded into a fuzzy reasoning structure. The fuzzy reasoning structure is Mamdani type in which the position, dimensions, cost, and other properties of a house are defined as the input variables and the user's definition for an appropriate house is defined as the output variables. The required information layers are extracted from 1:2000 scale digital map of the city of Tehran. These information include: educational centers, entertainment areas, medical centers, etc. The descriptive information such as the property valuation, position etc. are also extracted form existing data. The preliminary tests on the implemented fuzzy data management system indicate the superiority of our proposed method over the conventional systems.

HANDLING TEMPORAL UNCERTAINTY IN GIS DOMAIN: A FUZZY APPROACH

Jonathan Li [a, *], Simon Wu [b], Gordon Huang [b]

[a] Geomatics Engineering Program, Department of Civil Engineering, Ryerson University
350 Victoria Street, Toronto, Ontario, Canada M5B 2K3
junli@ryerson.ca
[b] Environmental Systems Engineering Program, Faculty of Engineering, University of Regina
3737 Wascana Parkway, Regina, Saskatchewan, Canada S4S 0A2
{wusimon, gordon.huang}@uregina.ca

Commission IV, WG IV/1

KEY WORDS: Temporal, Uncertainty, Fuzzy Set, Model, Reasoning, GIS

ABSTRACT:

This paper is an attempt to introduce a new framework to handle both uncertainty and time in spatial domain. The application of the fuzzy temporal constraint network (FTCN) method is proposed for representation and reasoning of uncertain temporal data. A brief introduction of the fuzzy sets theory is followed by description of the FTCN method with its main algorithms. The paper then discusses the issues of incorporating fuzzy approach into current sptio-temporal processing framework. The general temporal data model is extended to accommodate uncertainties with temporal data and relationships among events. A theoretical FTCN process of fuzzy transition for the imprecise information is introduced with an example. A summary of the paper is given together with outlining some contributions of the paper and future research directions.

1. INTRODUCTION

Geographic information systems (GIS) have been popularly applied in modeling environmental and ecological systems. However, real-world system analysis is typically challenged by difficulties in handling temporal change and uncertainty. Current GIS provide little or no support for modeling dynamic and uncertain phenomena. The GIS, more broadly, lack the ability to effectively model change and vagueness. There has been much research activity over the past decade focused on dynamic applications with geographic reference and handling uncertain features of geographic applications (Leung and Leung, 1993; Fischer, 1994; Stefanakis 1997). Fuzzy set theory, as a promising logical foundation for an intelligent GIS (Stefanakis, 1997), has drawn a lot of attention in its application to handling uncertainty and temporal reasoning in spatial domain. The fuzzy approach has been applied in geography on problems related to soil classification and definitions (Burrough, 1989; Kollias and Voliotis, 1991; Banai, 1993; Davidson et al., 1994; Cheng et al., 2001). The fuzzy aspect of boundaries in geographic space is now fully recognized (Mark and Csillag, 1989; Wang and Hall, 1996). A general spatial data model was extended to deal with the uncertainty with geographic entities in a database repository (Stefanakis, 1999). The fuzzy set theory was also applied to perform temporal interpolation in a raster GIS database (Dragicevic and Marceau, 2000).

There is little research trying to handle uncertainty in temporal presentation in spatial domain. This seems to be due to the fact that the handling of time and the management of uncertainty are two distinct important issues. Each of both possesses its own specific problems, and their solutions require a lot of research effort separately. Some modeling issues in spatial domain require a representation of time and of the temporal relationships between events. Most systems rely on a mechanism in which time is associated with each piece of knowledge. In spatial data models currently used, temporal information is stored through a series of snapshots associated to particular instants in time. Relationships are then the time ordering. In the complex environmental and ecological systems modeling with GIS, the current time mechanism is not sufficient. One must be able to represent situations with relative knowledge like "precedes" or "during". Real-world modeling requires a unified model of time and uncertainty. Here is an example of the scenario in wildlife migration modeling analysis: All animals of species S_1 moved away from area A before October 1st. Animals of species S_2 entered the area A after June 1st. But between mid-July and mid-August, we seldom saw both species of animal appeared at same time in the area. This gives a sample description of temporal definitions and relationships necessary for modeling wildlife migration properly. One obvious characteristic with the description is that those definition and relationships are vague. A major task of the analysis is to translate it and capture the knowledge in a form which we can correctly process and reason over.

The purpose of this paper is to introduce a novel fuzzy approach, fuzzy temporal constraint network (FTCN) for reasoning uncertain temporal data in spatio-temporal information processing. Section 2 presents a brief introduction of the fuzzy sets theory as the basis of the model. Section 3 describes the FTCN method with its main algorithms. Section 4 addresses the issues of incorporating the fuzzy approach into current sptio-temporal processing model, and of FTCN inferences in a GIS domain. Section 5 concludes the discussion by summarizing the contribution of the paper and giving some directions for future research.

2. FUZZY SET THEORY

Fuzzy theory is a method that facilitates uncertainty analysis of systems where uncertainty arises due to vagueness or "fuzziness" rather than due to randomness alone (Evans et al., 1986). This is based on a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth - truth values between "completely true" and "completely false". It was introduced by Zadeh as a means to model the uncertainty of natural language (Zadeh, 1965). Fuzzy theory uses the process of "fuzzification" as a methodology to generalize any specific theory from a crisp (discrete) to a continuous (fuzzy) form.

Classical set theory has a "crisp" definition as to whether an element is a member of a set or not. However, certain attributes of systems cannot be ascribed to one set or another. For example, an attribute of a system can be specified as either "low" or "high". In such a case, uncertainty arises out of vagueness involved in the definition of that attribute. Classical set theory allows for either one or the other value. On the other hand, fuzzy theory provides allows for a gradual degree of membership. This can be illustrated as follows:

In the classical set theory, the truth value of a statement can be given by the membership function $\mu_A(x)$, as

$$\mu_A(x) = \begin{cases} 1 & \text{iff } x \in A \\ 0 & \text{iff } x \notin A \end{cases} \tag{1}$$

On the other hand, fuzzy theory allows for a continuous value of μ_A, between 0 and 1, as

$$\mu_A(x) = \begin{cases} 1 & \text{iff } x \in A \\ 0 & \text{iff } x \notin A \\ p;\ 0 < p < 1 & \text{if } x \text{ partially belongs to A} \end{cases} \tag{2}$$

In fuzzy theory, statements are described in terms of membership functions, that are continuous and have a range [0,1]. For example, given the measured value of a parameter, the membership function gives the "degree of truth" that the parameter is "high" or "low".

Further, fuzzy logic is defined by the following the set relationships:

$$\begin{aligned} \mu_{A'}(x) &= 1 - \mu_A(x) \\ \mu_{A\cap B}(x) &= min(\mu_A(x), \mu_H(x)) \\ \mu_{A\cup B}(x) &= max(\mu_A(x), \mu_H(x)) \end{aligned} \tag{3}$$

Using fuzzy arithmetic, based on the grade of membership of a parameter of interest in a set, the grade of membership of a model output in another set can be calculated. Fuzzy theory can be considered to be a generalization of the classical set theory. It must be noted that if the membership grades are restricted to only 0 and 1, the fuzzy theory simplifies to classical set theory.

3. FUZZY TEMPORAL CONSTRAINT NETWORK

Temporal constraint network algorithm was proposed to handle time reasoning in artificial intelligence systems with explicit temporal representation (Dechter et al., 1991). However, it is not enough for domains where knowledge about time is highly pervaded with vagueness and uncertainty. Several pieces of work have considered the representation of approximate temporal knowledge. Vila and Godo (1994) described an extended method by applying fuzzy set theory, the Fuzzy Temporal Constraint Network (FTCN).

Timeline is assumed to be a discrete and totally ordered set as the basis of temporal entities, $\tau = \{t_0, t_1, \ldots, t_i, \ldots\}$, where each t_i represents a precise point of time and t_0 represents the time origin. Time points are equally spaced and the unit of time is the distance between two consecutive time points. An imprecise time point *a* is represented by means of a possibility distribution π_a over τ. Given any precise time point $t \in \tau$, $\pi_a(t) \in [0,1]$ represents the possibility (membership) of *a* being precisely *t*. The extreme values 1 and 0 represent, respectively, the absolute and null possibilities of *a* being *t*. There is always a fuzzy set A associated with a possibility distribution π_a, whose membership function coincides with π_a. The concept of imprecise time extent was introduced to represent fuzzy quantities of time. An imprecise time extent *m* is represented by a possibility distribution π_m over the set of integer numbers **I**, where the elements of **I** represent units of time. Thus, given an $n \in \mathbf{I}$, $\pi_m(n) \in [0,1]$ represents the possibility of the quantity of time *m* being precisely n units of time. The duration of the time interval between two fuzzy time points is an imprecise time extent. Given an ordered pair of imprecise time points (*a*, *b*), the temporal distance from *a* to *b*, *D*(*a*, *b*), is represented by means of the time extent possibility distribution, $\pi_{d(a,b)}$:

$$\forall\, n \in \mathbf{I},\ \pi_{d(a,b)}(n) = \max_{n = t - s;\ t, s \in \tau} \min \{, \pi_a(s), \pi_b(t)\}, \tag{4}$$

corresponding to the fuzzy difference $D(a, b) = B \ominus A$

An imprecise time interval *i* may be defined as a triplet (*a*, *b*, *d*), in which *a* and *b* are fuzzy time points representing, respectively, the initial and final time points of the intervals, and *d* is a fuzzy time extent representing its duration. Assume we have a set of instantaneous events whose time points of occurrence are completely unknown. Each of them is represented by means of a temporal variable X_i that, in absence of any other information, can take as its value any precise time point t_i. When any imprecise information is added about the position of event X_i, the information can be represented as the possibility distribution of an imprecise time point π_i and establishes a constraint over the possible values for X_i. It is called unary constraint L_i over the variable X_i. Then assume some imprecise information is added about the relative position of two variables X_i and X_j, that is, over the time elapsed between the two events they represent. This information can be represented as the possibility distribution of an imprecise time extent and π_{ij}, establishes a constraint over the possible values of other constraints, the assignments $X_i = t_i$ and $X_j = t_j$ are possible if $\pi_{ij}(t_j - t_i) > 0$. This is called binary constraint *L* over variables X_i and X_j.

Based on the above arbitrary set of information, a fuzzy temporal constraint network (FTCN) can be defined as follows: A FTCN *L* is a finite set of n + 1 temporal variables X_0, X_1, ... X_m, and a finite set of binary constraints L_{ij}, over those variables, where X_0 is a precise origin of times. Minimized imprecision in the position X_i can be obtained when all the constraints over X_i are combined. The formalization of concept of minimizing imprecision is also supported by a few more definitions on *L*. Here we just introduce how to obtain the minimal network *M* corresponding to a network L. Between each pair of L, X_i, X_j, there is a direct constraint L_{ij}, but there will also be induced constraints. Each induced constraint will correspond to a possible path connecting to them time points associated with X_i and X_j. In order to obtain the minimal constraint between two nodes X_i and X_j, it is necessary to obtain the intersection of the direct constraint L_{ij} and the induced constraints between X_i and X_j.

Given a path k from a network node of index i to another one of index j:

$$i_0 = i,\ i_1,\ i_2, \dots,\ i_k = j, \tag{5}$$

The induced constraint is represented as $C^{k}_{i0,i1,\dots ik}$. It is given by the composition of the direct constraints between each pair of consecutive nodes belonging to the path.

$$C^{k}_{i0,i1,\dots ik} = L_{i0i1} \oplus L_{i1i2} \oplus \dots \oplus L_{ik\text{-}1ik} = \sum_{p=1}^{k} L_{ip\text{-}1ip} \tag{6}$$

L^k_{ij} is used as the intersection of the induced constraints corresponding to all the paths of length k from the node with index i to the one with index j:

$$L^k_{ij} = \bigcap_{i0=i,i1=0,\dots n,ik-1=0,\dots n,ik=j} C^{k}_{i0,i1,\dots ik}{}^{z} \tag{7}$$

Finally M_{ij} is used to present the intersection of all the induced constraints corresponding to paths of any length between 1 and n going from X_i to X_j: $M_{ij} = \bigcap_{k=1}^{n} L^k_{ij}$

It can be proved that a network L is consistent if and only if the constraints M_{ij} obtained by means of this expression are normalized. Also in this case, the constraints M_{ij} correspond to the minimal network M associated with L. These results have practical interest from the computational viewpoint. It is easy to operate with normalized distributions which are mostly common. Figure 1 shows a trapezoidal distribution represented by means of the four parameters (α, β, γ, δ). The M_{ij} can be obtained by solving constraint satisfaction problems with path-consistency algorithms (Rit, 1986; Van Beek, 1990; Dechter et al., 1991).

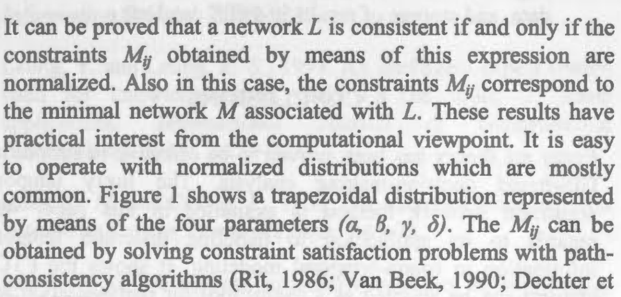


Figure 1. Trapezoidal possibility distribution for an imprecise time point

4. FUZZY TIME LAYER: MODEL AND REASONING

A critical issue for this research is how to incorporate fuzziness into a general spatio-temporal data model in GIS with the fuzzy temporal constraint network method. The most usual way of handling the time factor in current GIS is to look on time as an attribute to the objects in the same way as for other attributes. This view corresponds with the usual way of presenting spatial data, and can thus be realized for both vector and raster data (Bernhardsen, 1999). Temporal information is stored through a series of time layers (snapshots) associated to particular instants in time. To accommodate uncertain time information, the current data model must be extended to represent the partial memberships (possibilities) of each time point. Thus the fuzzy time layer concept is introduced here. Corresponding to the definition in the FTCN method, a fuzzy time layer is an imprecise time interval with initial and final time points and possibility distribution function (Figure 2). The fuzzy information can be specified in various ways based on practical observation and individual type of possibility distribution.

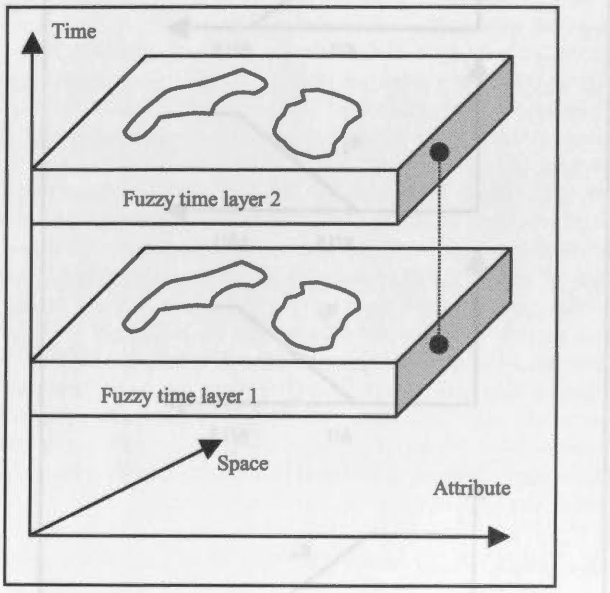


Figure 2. Uncertain times are realized in fuzzy layers (Dashed line between two layers indicates fuzzy relationship)

New components and related definitions have to be brought into data model for incorporation of the temporal fuzziness. Also the model structure has to be rebuilt to fit the change. Obviously, the number of layers increases for characterizing a temporal uncertainty. More complicated structure will be seen due to the requirements for storing and retrieving fuzzy membership functions, and if necessary, reasoning results with the FTCN method as to be discussed later. The specific concept of fuzzy temporal layer should not be unique. It may vary depending on specific uncertainty pattern on individual time layer and query requirements.

The temporal reasoning regarding fuzzy relationships among time layers has to be tackled following reconstruction of data model. This involves the incorporation of FTCN algorithms into basic data interpretation operations in GIS modeling packages. Let us still use the wildlife migration modeling example mentioned in Section 1. As we can see now, it is an ideal trapezoidal distribution problem. As shown in Figure 1, α =

June 1st, β = July 15th, γ = August 15th, and δ = October 1st. A logical translation of the fuzzy temporal problem starts with construction of separate possibility distributions (Figure 3). A simplified inference procedure in mathematical form can be as follows:

The fuzzy temporal possibility distributions π *, objects of fuzzy subsets:

$$\begin{aligned}
\pi_1^* &= \tau2 \cap \tau4 \cap \tau5 \\
\pi_2^* &= \tau2 \cap \tau3 \\
\pi_3^* &= \tau1 \cap \tau4 \\
\pi_4^* &= \tau1 \cap \tau3 \cap \tau5
\end{aligned} \tag{8}$$

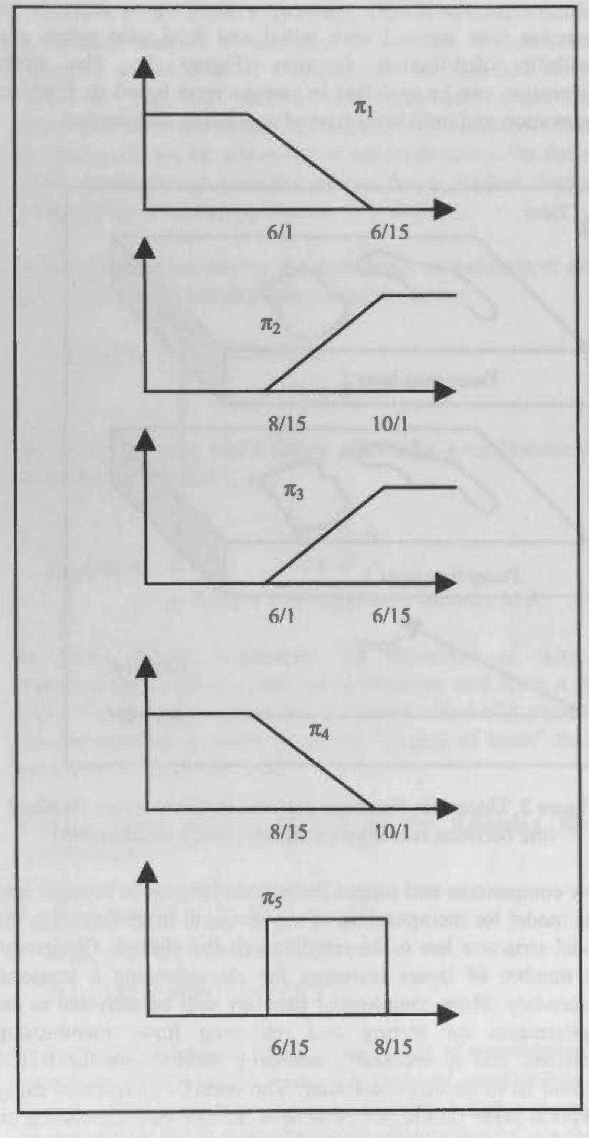


Figure 3. Separated possibility distributions

Then we can have the fuzzy complementary of the union of these four fuzzy subsets:

$$\begin{aligned}
\tilde{U} &= \pi_1^* \cup \pi_2^* \cup \pi_3^* \cup \pi_4^* \\
&= (\tau_2 \cup \tau4 \cup \tau5) \cap (\tau2 \cup \tau3) \cap (\tau_1 \cup \tau4) \cap (\tau1 \cup \tau3 \cup \tau5) \\
&= (\tau2 \cup \tau4 \cup \tau5) \cap \tau3 \cap \tau1 \cap (\tau1 \cup \tau3 \cup \tau5) \\
&= \tau1 \cap \tau3 \cap (\tau2 \cup \tau4 \cup \tau5) \qquad (9) \\
&= (\tau1 \cap \tau2 \cap \tau3) \cup (\tau1 \cap \tau3 \cap \tau4) \cup (\tau1 \cap \tau3 \cap \tau5) \\
&= \tau1 \cap \tau3 \cap \tau5 = \tau1 \cap \tau5
\end{aligned}$$

Now we can compute temporal necessity functions N *(1) and N *(2) induced by π * for the following two expressions in the example: (1) Species S_1 has certainly left the area, and (2) Species S_2 has certainly left the area.

$$\begin{aligned}
N^*(1) &= (\tau1 \cap \tau4) \cup (\tau1 \cap \tau3 \cap \tau5) \\
&= (\tau1 \cup \tau4) \cap (\tau1 \cup \tau3 \cup \tau5) = (\tau1 \cup \tau4) \cap T = \tau1
\end{aligned}$$

$$N^*(2) = (\tau2 \cup \tau3) \cap (\tau1 \cup \tau3 \cup \tau5) = \tau3 \tag{10}$$

Generally, this approach makes possible to compute the fuzzy set of instants when we are more or less certain that a proposition is true, and on the other hand, to compute the fuzzy set of instants when we are more or less certain that proposition is false. The above translation procedure and algorithm can be built into GIS as one of spatio-temporal operation functions. This would be possible if we have a well defined fuzzy data model with much knowledge on uncertainties for specifying possibility distributions and fuzzy relationships. It is necessary for large-scale dynamic modeling analysis which may have a lot of uncertainty in time dimension and interactions among events. Expected queries to related information in a GIS database should also be considered before constructing a built-in system. A set of computer programs are being developed to implement the FTCN approach to model the uncertain temporal reasoning for a case study we are working on. The major steps and tasks in the program include:

- Generation of fuzzy time layers;
- Quantification and construction of fuzzy relationships between timed events;
- Verification of generalized fuzzy parameters;
- Implementation of the FTCN process to parse the fuzzy data, and storage of results in a GIS database.

5. CONCLUSIONS

Fuzzy set theory has been proved to be effective in supporting GIS-based decision-making analysis. The fuzzy temporal constraint network method is examined in this paper with regards to its application to handling uncertain temporal information in spatio-temporal modeling. It shows the FTCN concept can be adopted as a useful tool for representation and reasoning of temporal uncertainty in real-world applications. The paper discusses how the general temporal data model is extended to accommodate uncertainty with temporal data and relationships among events. A theoretical FTCN process of fuzzy translation for the imprecise information is introduced with an example. The research is going on with an experimental study of the application of the method to environmental systems modeling. The focus will be on streamlining the FTCN reasoning procedure for various circumstances of temporal uncertainty. Then effort will be made to develop a versatile computer program built into a GIS database.

REFERENCES

Banai, R., 1993. Fuzziness in geographical information systems: contribution from the analytic hierarchy process. *International Journal of Geographical Information Science*, 4:315-329.

Bernhardsen, T., 1999. *Geographic Information Systems: An Introduction*, New York, John Wiley.

Burrough, P. A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science*, 40:477-492.

Cheng, T., Molenaar, M. and Lin, H., 2001. Formalizing fuzzy objects from uncertain classification results. *International Journal of Geographical Information Science*, 15(1):27-42.

Davidson, D. A., Theocharopoulos, S. P. and Bloksma, R. J., 1994. A land evaluation project in Greece using GIS and based on Boolean and fuzzy set methodologies. *International Journal of Geographical Information Science*, 4:369-384.

Dechter, R., Meiri, I. and Pearl, J., 1991. Temporal constraint networks. *Artificial Intelligence*, 49:61-95.

Dragicevic, S. and Marceau, D. J., 2000. A fuzzy set approach for modeling time in GIS. *International Journal of Geographical Information Science*, 14(3):225-245.

Evans, G. W., Karwowski, W. and Wilhelm, M. R., 1986. An Introduction to Fuzzy Set Methodologies for Industrial and Systems Engineering. In G. W. Evans, W. Karwowski, and M. R. Wilhelm, editors, *Applications of Fuzzy Set Methodologies in Industrial Engineering*, pp. 3-11. Elsevier, New York.

Fischer, M. M., 1994. Expert systems and artificial neural networks for spatial analysis and modeling. *Geographical Systems Journal*, 1:221-235.

Kollias, V. J. and Voliotis, A., 1991. Fuzzy reasoning in the development of geographical information systems FRSIS: a prototype soil information system with fuzzy retrieval capabilities. *International Journal of Geographical Information Science*, 2:209-223.

Leung, Y. and Leung, K. S., 1993. An intelligent expert system shell for knowledge-based GIS: 1. The tools, 2. Some applications. *International Journal of Geographical Information Science*, 7:189-213.

Mark, D. M. and Csillag, F., 1989. The nature of boundaries on 'Area-Class' maps. *Cartographica*, 1:65-78.

Rit, J., 1986. Propagating temporal constraints for scheduling. *Proc. AAAI-86*, Philadelphia, pp. 383-388.

Stefanakis, E., 1997. *Development of Intelligent GIS*. Ph.D. Dissertation, Department of Electrical and Computer Engineering, National Technical University of Athens, Greece.

Stefanakis, E., Vazirgiannis, M. and Sellis, T., 1999. Incorporating fuzzy set methodologies in a DBMS repository for the application domain of GIS. *International Journal of Geographical Information Science*, 13(7):657-675.

Van Beek, P, 1990. *Exact and Approximate Reasoning about Qualitative Temporal Relations*. Ph.D. Thesis. University of Waterloo, Ontario, Canada.

Vila, L. and Godo, L., 1994. On fuzzy temporal constraint networks. *Mathware and Soft Computing*, 3:315-334.

Wang, F. and Hall, B. G., 1996. Fuzzy representation of geographical boundaries in GIS. *International Journal of Geographical Information Science*, 5:573-590.

Zadeh, L. A., 1965. Fuzzy Sets, *Information and Control*, 8:338-253.

A NEW MULTI-SCALE MODELING APPROACH FOR SPACE/TIME RANDOM FIELD ESTIMATION

Kyung-Mee Choi [a, *], George Christakos [b], Mark L. Wilson [a]

[a]Dept. of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109-2029, USA – kmchoi@umich.edu, wilsonml@umich.edu
[b]Dept. of Env. Sci. and Eng., University of North Carolina, Chapel Hill, NC, USA - george_christakos@unc.edu

Commission IV, WG IV/1

KEY WORDS: Multiple scale, Local scale, Measurement scale, random field, BME, Spatiotemporal, Estimate

ABSTRACT:

Modern geostatistical mapping methods are being applied to various types of data to produce more realistic and flexible characterizations of a natural random process. The Bayesian Maximum Entropy (BME) is a well-known geostatistical estimation method, especially for the use of soft knowledge as well as exact measurement data. Although development in geostatistical methods helps us to solve limitations on the format of available data, real studies always present *in situ* problems. Spatial scale of mapping (grid) points in a mapping model is usually not considered at the spatial scale of measurement data, especially in the studies that involve health-related data. Moreover, the spatial scale of measurement data may not be uniform, but varies among different measurements. For example, in studies of epidemiology or environmental health exposure, spatial scale of available measurement data is often limited and becomes different from the interesting spatial scale that is sought for in the estimation of the unknown random fields. It may be difficult and unrealistic to obtain measurement data at the scale of interest. Most current geostatistical methods have difficulty explaining physical phenomenon of unknown random fields over a continuous mapping domain at a scale smaller than that available from measurement data. This study explores how we can define these different scales in a geostatistical mapping model, and attempts to generate a meaningful spatiotemporal map of estimates of unknown random fields at the scale of interest. The estimation process of this study is based on the BME method to allow the probabilistic type of "soft" data, which are not actually observed, but simulated at the local scale to the measurement scale. This new modelling approach has been called multi-scale or local-scale mapping model. With actual mortality data collected over the 58 counties of the state of California, USA, we applied this multi-scale modelling approach, and obtained more accurate and realistic spatiotemporal maps of mortality rate estimates over California. We compared these estimates with those found by another approach that did not account for multiple scales on the same data. It was verified by actual mortality data obtained at the zip-code scale. These estimates found by the multi-scale approach were considered to be more accurate than those from the other modelling approach.

1. INTRODUCTION

Spatiotemporal mapping analysis had been originally introduced by D. G. Krige in 1951, but not until 1971 was it actively developed by G. Matheron. Today, geostatistical analysis is applied to various spatial problems, although it initially addressed the area of mining engineering. Recently, applications to health-related studies have been made, for example kriging estimation of epidemiological data on influenza-like illness in France (Carrat and Valleron, 1992) and spatio-temporal mapping based on BME (Bayesian Maximum Entropy) of mortality data in the state of California in USA (Choi et al., 2001a). When mapping analysis is applied to spatial data, especially related to disease patterns, we search for transmission mechanisms, possible risk factors, and other useful information for health risk management (Christakos and Hristopoulos, 1998). With data collected at different locations for a natural process, we can find variations and uncertainties in these natural processes. For example, a stochastic process can be characterized by the spatial and temporal variations that we find based on available data for that process. We can obtain insights that might inform natural variation in processes by understanding the spatial distribution in available data of these processes.

Different stochastic estimation methods may provide different estimation results depending on the assumptions and limitations of the methods. BME method allows more flexible data to be used in the estimation of a stochastic process than other mapping methods. In this procedure of BME estimation, we can use various types of uncertain knowledge called "soft" data, such as interval, probabilistic or functional types of available data (Christakos, 2000; Choi et al. 1998). Often, meaningful characterizations of a natural process may involve the appreciation of its spatiotemporal variation at multiple scales, rather than a fixed scale over a mapping domain of space. Rigorous description of the situation depends on the scale at which phenomena are considered, rather than the scale at which measurements are taken (Choi et al., 2001b). In epidemiological studies, we often are interested in the spatial distribution of epidemiological variation at a smaller spatial scale than that of the measurements which were made. However data often are collected at a larger scale in most surveillance systems or health studies. Moreover, observed data

---

[*] Corresponding author.

may be available only at arbitrary discrete locations rather than over a continuous spatial domain.

In this study, we introduce a recent development of a new modeling approach called multi-scale modeling (Choi, 2001) which accounts for multiple scales in the estimation of spatiotemporal random fields considering those data that were at limited scales of different discrete spatial areas. The multi-scale modeling approach is developed based on the BME estimation method, and simulates soft data at the scale of mapping interest not using the data actually observed at the limited scales from the natural field. From these features, the new approach can explain more realistic stochastic variation and find more accurate estimates for the random processes than other approaches that do not account for the multiple scales. We applied the multiple-scale approach to a real-world study that involves mortality data collected at the spatial scale of counties in the State of California, USA. We generated a map of estimated mortality rates using simulated data at the local scale.

First, we review the Spatiotemporal Random Field (S/TRF) representation of a natural process and the framework of BME estimation. Then we present the multi-scale modelling approach that considers phenomena at multiple scales. Finally, we apply this approach to mortality data obtained at the scale of counties of the State of California, USA. Realistic and meaningful maps of estimates are generated. This approach is verified using mortality data collected at an actually local scale to the counties but which was not used for the estimation. The multi-scale approach produced more accurate maps than those found by other approaches.

2. SPATIOTEMPORAL RANDOM FIELDS AND OVERVIEW OF THE BME METHOD

2.1 S/T Random Fields

To describe randomness of a natural process, we define a spatiotemporal (s/t) random field $X(\mathbf{s},t)$ that takes real values at points $\mathbf{s}=(s_x,s_y)\in R^2$ and $t\in T\subseteq R^1_{0,+}$ in a domain of two-dimensional space and time. In most studies, a random field is assumed isotropic and characterized by mean and isotropic covariance functions. Stochastic trend of the isotropic random field is explained by a mean function $m_x=E[X(\mathbf{s},t)]$ at any point in the domain of interest. Stochastic variation is explained by an isotropic spatiotemporal covariance function $c_x(r,\tau)=E[X(\mathbf{p})-m_x][X(\mathbf{p}')-m_x]$ that assumes correlation between random fields at any pair of points, $\mathbf{p}=(\mathbf{s},t)$ and $\mathbf{p}'=(\mathbf{s}',t')$ for $\mathbf{p},\mathbf{p}'\in R^2\times T$ depends only on the spatial and temporal lag distances $r=|\mathbf{p}-\mathbf{p}'|$ and $\tau=|t-t'|$ between the pair of points. In this study, Greek letter χ is used to represent for the value of a random variable, and small letter x is used to denote a random variable for the random field. Bold letter $\mathbf{x}$ is used to denote a vector of variables or values.

2.2 Overview of the BME Method

The BME estimation (Christakos, 1992) consists of three main stages: prior, case-specific, and posterior. Each has corresponding knowledge. Prior stage processes general knowledge ($\mathcal{G}$), such as statistical moments or underlying physical laws that we can obtain from past experience and studies about the process. In prior stage, a prior probabilistic density function $f_{\mathcal{G}}(\chi_{\mathrm{map}})$, $\chi_{\mathrm{map}}=(\chi_{\mathrm{data}},\chi_k)$ is formulated based on available prior knowledge. The case-specific stage (also termed the specificatory stage) involves certain and uncertain types of data. A unique advantageous feature of the BME method is uncertain knowledge (called "soft" data) χ_{soft} which is considered in the estimation procedure as is data from exact measurements, or "hard" data χ_{hard}. Soft data can be any uncertain type of knowledge about the process such as probabilistic distribution functions or intervals. We denote case-specific or specificatory knowledge ($\mathcal{S}$) in terms of both hard and soft data as

$$\mathcal{S}:\ \chi_{\mathrm{data}}=(\chi_{\mathrm{hard}},\chi_{\mathrm{soft}})=(\chi_1,\ldots,\chi_m).$$

Below are several typical types of soft data χ_{soft} formulated for the BME estimation.

Interval type : $\chi_{\mathrm{soft}}\in\mathbf{I}=(I_{m_h+1},\ldots,I_m)$

Probabilistic type : $P_{\mathcal{S}}[\mathbf{x}_{\mathrm{soft}}\leq\zeta]=F_{\mathcal{S}}(\zeta)$

Functional type : $P_{\mathcal{S}}[\mathbf{x}_{\mathrm{soft}}\leq\zeta,x_{\mathrm{k}}\leq\zeta_{\mathrm{k}}]=F_{\mathcal{S}}(\zeta,\zeta_{\mathrm{k}})$

In the posterior stage, both general and case-specific knowledge are combined to generate a posterior probabilistic density function $f_{\mathcal{K}}(\chi_k)$ as

$$f_{\mathcal{K}}(\chi_k)=A^{-1}\int_D d\Xi(\chi_{\mathrm{soft}})f_{\mathcal{G}}(\chi_{\mathrm{map}})$$

where $A=\int_D d\Xi(\chi_{\mathrm{soft}})\int_D d\chi_k f_{\mathcal{G}}(\chi_{\mathrm{map}})$ and $\chi_{\mathrm{map}}=(\chi_{\mathrm{hard}},\chi_{\mathrm{soft}},\chi_k)$.

Table 1 shows integration domain **D** and operator $\Xi(\chi_{\mathrm{soft}})$ in the posterior function according to specificatory knowledge. BME estimate is found by various selection rules such as mode or mean estimate from the posterior function $f_{\mathcal{K}}(\chi_k)$. Estimation uncertainty for BME estimate is explained by a mean squared error based on the posterior function.

$\mathcal{S}$	D	$\Xi(\chi_{\mathrm{soft}})$
Interval	$\mathbf{I}$	χ_{soft}
Probabilistic	$\mathbf{I}$	$F_{\mathcal{S}}(\chi_{\mathrm{soft}})$
Functional	$\mathbf{I}\cup\mathbf{I}_k$	$F_{\mathcal{S}}(\chi_{\mathrm{soft}},\chi_k)$

Table 1. Integration domain **D** and operator $\Xi(\chi_{\mathrm{soft}})$

3. MULTI-SCALE MODELING APPROACH

When we estimate unknown random fields, stochastic variation of the random fields is derived from available measurement data at the scale that the measurements have been made rather than at the scale that is estimated. The scale of estimation points is taken uniformly and at a sufficiently fine scale that we can consider those estimation points to cover a continuous mapping domain. However, regardless of the scale we are interested in for the estimation of a process, the spatial scale of measurement data is exclusively dependent on the scale of available data. The scale of measurement data is usually large compared with the scale of estimation points and may not be uniform among the data. For example in the mortality study over the state of California (Choi et al., 2001a), mortality data were collected at the scale of counties that are different from one another in their sizes and shapes. The new multi-scale modelling approach (Choi, 2001) takes it into consideration not only the scale of estimation points, but also the scale of measurement data so that it can account for the multiple scales for the grid points of estimation and data of a natural process over a continuous mapping domain. Rigorous and more realistic description of stochastic variation becomes obtainable at the scale that phenomena are considered for the estimation of random field by the multi-scale modelling approach.

Let us denote a random field defined at a measurement scale as $Z(\mathbf{s},t)$, and a random field defined at the local scale to the measurement scale as $X(\mathbf{s},t)$ for a natural process of interest. These two random fields can be explained by each other as in Eq. (1).

$$Z(\mathbf{s},t) = \| D(\mathbf{s}) \|^{-1} \int_{\mathbf{u}\in D(\mathbf{s})} d\mathbf{u}\, X(\mathbf{u},t) \qquad (1)$$

where $\| D(\mathbf{s}) \|$ is the size of $D(\mathbf{s})$. Assuming that the random fields are isotropic, we can easily derive a same mean function $m_z = m_x$ for the random fields as each other from the equation above. Let c_z and c_x be denoted for isotropic s/t covariance functions respectively for the random fields defined at measurement scale and at its local scale. These covariance functions are also related to each other as shown below in Eq. (2), which is derived from Eq. (1) and the definition of isotropic covariance functions. For spatial and temporal lags $r = |\mathbf{s} - \mathbf{s}'|$ and $\tau = t - t'$, it is obtained

$$c_z(r,\tau) = \| D(\mathbf{s}) \|^{-1} \| D(\mathbf{s}') \|^{-1} \sum_{\mathbf{u}\in D(\mathbf{0})} \int_{\mathbf{u}'\in D(r)} d\mathbf{u}\, d\mathbf{u}'\, c_x(|\mathbf{u}-\mathbf{u}'|,\tau) \,. \qquad (2)$$

We cannot find the covariance c_x directly from measurement data since c_x represents local stochastic variation for random field $X(\mathbf{s},t)$. With an initial guess for c_x, we obtain c_z from the relation in Eq. (2). And c_z can approach the experimental covariance at the measurement scale, which we obtain from the measurement data. Change models for c_x by trial-and-error so that c_z obtained by Eq. (2) from c_x becomes close enough to the experimental variation obtained from the measurement data.

Once we have found the isotropic characteristics for the local random field $X(\mathbf{s},t)$, we simulate a set of local scale random fields $\{ X(\mathbf{u},t), \forall \mathbf{u} \in D(\mathbf{s}) \}$ to the measurement scale of $Z(\mathbf{s},t)$. By making a large enough number of simulations, we apply the central limit theorem so that any initial probabilistic distribution can be accepted for the simulation of $X(\mathbf{u},t)$. With the isotropic property of $X(\mathbf{s},t)$, the multi-scale approach takes local data that are distance-dependent between estimation and data points from the simulated random fields $\{ X(\mathbf{u},t)\ \forall \mathbf{u} \in D(\mathbf{s}) \}$ rather than taking one of the simulated values at a certain point (e.g., a centroid). Then, we find an estimate by applying the BME method with the knowledge of covariance and simulated data (not observed) at the local random fields. Local scale data can be of any type that the BME method can process (as explained earlier). In the next section, we apply this multi-scale approach to real mortality data and generate a map of estimates over the state of California. Verification of the estimated results follows.

4. THE CALIFORNIA MORTALITY STUDY

We applied the new multi-scale approach for mapping analysis to mortality data from the Statistics Health Department of the State of California. Mortality is a more certain indicator variable that can be useful in studies of environmental exposure or epidemiological risk. The spatial scale of mortality data was available only at the resolution of California counties (N=58). We aimed to obtain a map of estimated mortality rates at a uniform and smaller scale by applying the multi-scale approach. Mortality data collected at the scale of zip-code were not used for the estimation, but allowed us to later verify the estimation results based on the county scale.

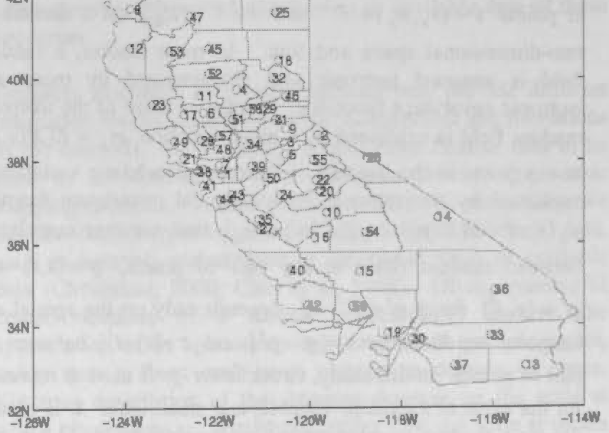


Figure 1. County map with centroids (circles) and identity numbers according to the order of county names.

4.1 Mortality Data

From the data set of death counts collected over the 58 counties of California during 1989, we selected death records only for California residents that were identified by ICD-9 (International

Classification of Death) code under 800. The total number of selected records for this study was 219,182. The map of California in Figure 1 shows county boundaries and county identity numbers on the county centroids according to the order of county names.

We obtained mortality rates by dividing the number of deaths by the population of each county, and used them as measurement data to generate estimated values. Using a smoothing technique (Choi, 2001; Choi et al., 2001a), we obtained space/time trends over these mortality rates. By leaving the trends out, we obtained a residual process of mortality rate $Z(\mathbf{s},t)$ at the county scale, which was assumed to be isotropic. Figure 2 shows the temporal trend (lined) over the mortality rate data (with circles) during the study year in the unit of days.

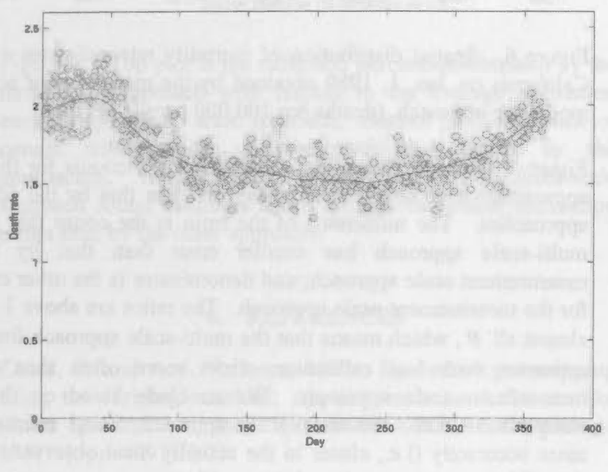


Figure 2. Mortality rates (with circles) and the temporal trend (lined) for the counties of 500,000 residents in California (deaths per 100,000 people per day).

4.2 Knowledge at the Local Scale

By applying the multi-scale modelling approach for the estimation of mortality rates over California, we first obtained experimental covariance at the county scale, shown in Figure 3 with circles. We approached the experimental county scale covariance with the covariance model $c_z(r,\tau)$ at the county scale obtained from various covariance models $c_x(r,\tau)$ at the local scale by trial-and-error as explained above. The nested exponential model in Eq. (3) is an optimal covariance function for $c_x(r,\tau)$ that we finally obtained at the local scale:

$$c_x(r,\tau) = \sum_{i=1}^{2} c_i \exp\left(-\frac{3r}{a_{r,i}}\right)\exp\left(-\frac{3\tau}{a_{t,i}}\right) \tag{3}$$

where c_1=8.09, $a_{r,1}$=1.14, $a_{t,1}$=43.5, c_2=4.37, $a_{r,2}$=51, $a_{t,2}$=990. From the values of c_1 and c_2, we can see that the covariance model $c_x(r,\tau)$ in the above is characterized more by the first nested structure that by the other. The second nested structure of $c_x(r,\tau)$ has a wider range both in space and time than the first nested structure. There was an additional third term in finding the $c_x(r,\tau)$, which was so small and ignored in Eq. (3).

Figure 3 shows the optimal covariance model $c_x(r,\tau)$ at the local scale and experimental covariance together with its fitting model $c_z(r,\tau)$ at the county scale. In the Figure 3, different covariance functions are shown on different y-axes. The left y-axis is used for the covariance at the county scale, and the right y-axis is for the covariance $c_x(r,\tau)$ at the local scale.

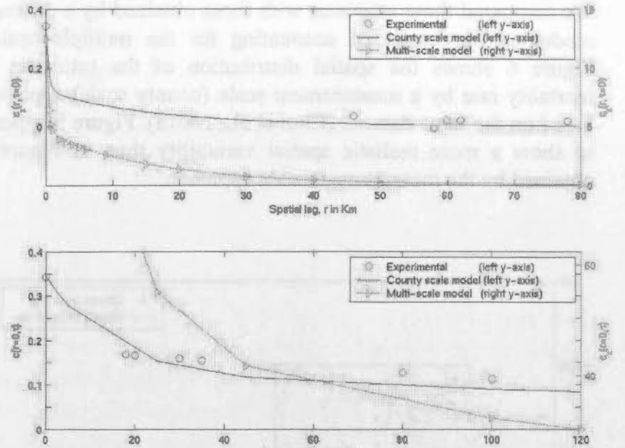


Figure 3. Isotropic spatiotemporal covariance of mortality rate (deaths per 100,000 people per day)^2 as the functions of spatial lag (top plot) and temporal lag (bottom plot).

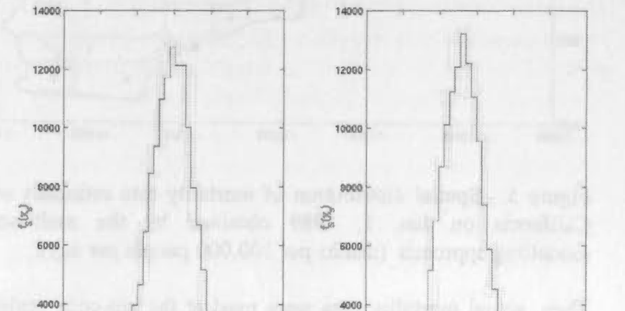


Figure 4. Probabilistic Soft data (in deaths per 100,000 people per day) at zero-lag of local distance between estimation and local data points to (a) county Alameda and (b) county Amador on January 1, 1989.

We simulated 10,000 local-scale random fields to each county and took the probabilistic type of soft data of the distance-dependent simulated values. Figure 4 shows the soft data for the residual process of mortality rates per 100,000 people per day at zero-lag of local distance between estimation and data points to county Alameda and county Amador on January 1, 1989. Since the soft data in Figure 4 is for the residual process of mortality rates, it ranges from negative to positive. With the soft data and covariance obtained at the local scale, we applied the BME estimation method and found mortality rate estimates throughout California.

4.3 Estimation Result and Verification

Using this multi-scale approach, we obtained estimates of mortality rates over a continuous spatial domain of California at any day during the study year. Figure 5 shows the spatial distribution of estimated mortality rates on Jan. 1, 1989. In this map, same mortality rates are shown on a same contour line. We compared these estimates with those obtained by a different modelling approach not accounting for the multiple scales. Figure 6 shows the spatial distribution of the estimates of mortality rate by a measurement scale (county scale) approach based on the same data set (Choi et al., 2001a). Figure 5 appears to show a more realistic spatial variability than in Figure 6 obtained by the measurement scale approach.

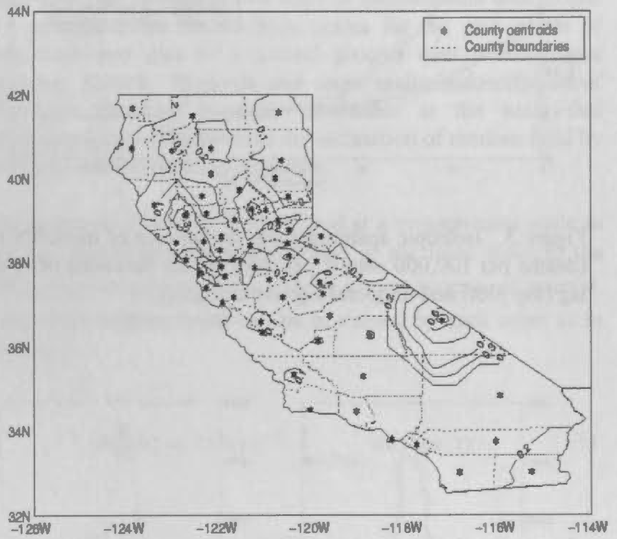


Figure 5. Spatial distribution of mortality rate estimates over California on Jan. 1, 1989 obtained by the multi-scale modelling approach (deaths per 100,000 people per day).

Then, actual mortality data were used at the zip-code scale to determine which modelling approach found closer estimates. To exclude computational error caused by mortality rates not relating to any modelling approach, we calculated the average of estimation errors that were above a threshold value θ representing the computational error. We compared in several ways the estimated mortality rates with others based on the average error as a function of θ, which were obtained by two different modelling approaches at the county scale and the corresponding local scale. We calculated the percentage of reduction in average error obtained by the multi-scale approach with respect to the average error obtained by the measurement scale approach. In Figure 7 (top), we see the values are all negative, suggesting as much reduction by the multi-scale approach as the percentage value. As we can see (Figure 7), the reduction increases as the threshold θ increases. This tells us that, as the average estimation error becomes larger, we can expect more reduction of the average estimation error by applying the multi-scale approach.

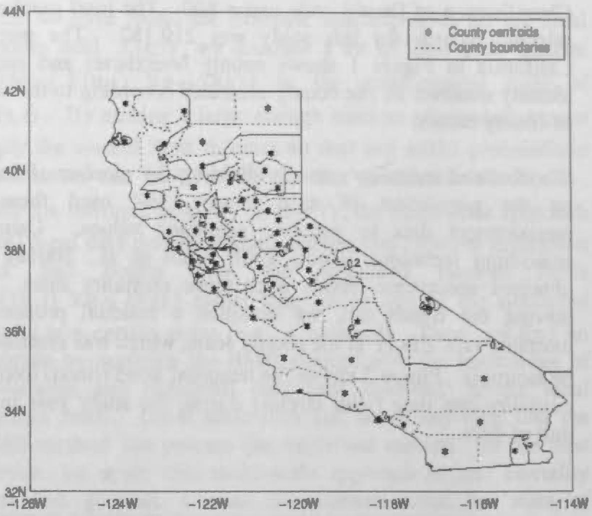


Figure 6. Spatial distribution of mortality rate estimates over California on Jan. 1, 1989 obtained by the measurement scale modelling approach (deaths per 100,000 people per day).

Figure 7 (bottom) shows that the ratio of the counts for these approaches have smaller average errors than that by the other approaches. The numerator of the ratio is the count that the multi-scale approach has smaller error than that by the measurement scale approach, and denominator is the other case for the measurement scale approach. The ratios are above 1 for almost all θ, which means that the multi-scale approach found estimates with less estimation errors more often than the measurement-scale approach. We conclude based on these comparisons that the multi-scale approach found estimates more accurately (i.e., closer to the actually local observations) than did the measurement scale approach .

5. CONCLUSIONS

To obtain a meaningful spatial distribution of a natural process from a limited scale of measurement data, we have introduced the recent development of multi-scale modelling. Stochastic variation of a random field was estimated, we believe realistically, at a scale local to the measurement scale which is usually limited by availability of data. Spatial analysis was studied conceptually over a continuous domain of space, although data were not observed in the continuous domain. Indeed, the scale of measurement data is usually not available at a uniform scale. By describing different scales of random fields for a natural process, we can represent more realistic variation for the process at the local scale to the available scale of actual measurement data. This multi-scale modelling approach seems to present more accurate spatial distribution of random field estimates than any other modelling approaches that do not account for multiple scales. This new modelling approach finds estimates based on the BME method so that realistic simulated soft data can be used. By collecting mortality rates at a local scale (California counties), it was verified that the estimates found by multi-scale approach were more accurate than those by county scale modelling approach accounting for only the measurement scale. Depending on available data and study purposes, the multi-scale modelling approach can be used to find geostatistical estimates at any scale of interest without being restricted by the scale of observed data.

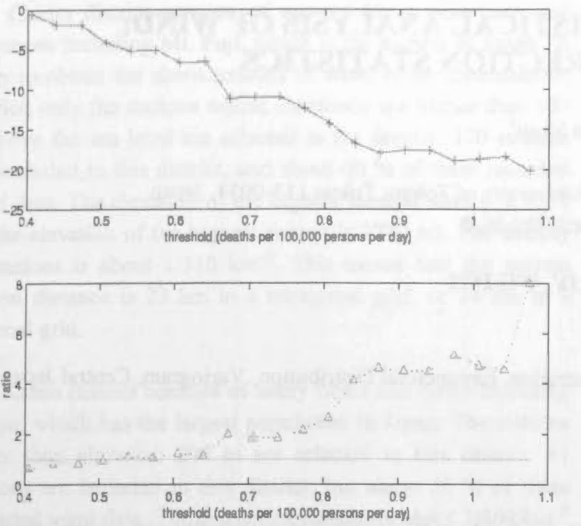


Figure 7. Top plot is the reduction percentage obtained by the multi-scale approach with respect to the average estimation error by the county scale approach. Bottom plot is the ratio of counts with smaller average estimation errors by the approaches. If the ratio is greater than 1, the multi-scale approach found estimates more often with smaller estimation errors than by the other approach.

6. REFERENCES

Carrat, F. and A.-J. Valleron, 1992. Epidemiologic mapping using the 'kriging' method: Application to an influenza-like illness epidemic in France. In: *Amer. Jour. of Epidemiology*, 135(11), pp. 1293-1300.

Choi, K-M, 2001. *BME-based spatiotemporal local scale mapping and filtering with mortality data: The California study*. Ph.D. dissertation, Dept. of Environmental Sciences and Eng., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC.

Choi, K-M, G. Christakos, and M. Serre, 1998. Recent developments in vectorial and multi-point Bayesian maximum entropy analysis. In: *IAMG 98, Proceeding of 4th Annul Confer. of the International Assoc. for Mathematical Geology* – Vol. 1, Buccianti A., G. Nardi and R. Potenza (eds.), De Frede Editore, Naples, Italy, pp. 91-96.

Choi, K-M, G. Christakos, and M. Serre, 2001a. Space/time BME analysis and mapping of mortality in California. In: *The 53rd Session of the International Stat. Institute*, Seoul, Korea pp. 22-29.

Choi, K-M, M. Serre, and G. Christakos, 2001b. Spatiotemporal BME mapping of mortality fields at different spatial scales. In: *Jour. of Exposure Analysis and Environ. Exposure* (submitted).

Christakos, G., 1992. *Random field models in Earth Sciences*. Academic Press, San Diego, CA. 474 p.

Christakos, G., 2000. A view of Modern Geostatistics. Oxford Univ. Press, New York, NY, 304 p.

Christakos, G. and Hristopulos, D.T., 1998. *Spatiotemporal environmental health modelling: A tarctatus stocahsticus*. Kluwer Academic Publ., Boston, 424 p.

DSMF (Death Statistical Master Files), 1989. Center for Health Statistics, Sacramento, CA.

7. ACKNOWLEGEMENTS

This work has been supported by a grant from the Office of the Vice President for Research, The University of Michigan.

STATISTICAL AND GEOSTATISTICAL ANALYSIS OF WIND: A CASE STUDY OF DIRECTION STATISTICS

Tetsuya Shoji[a]

[a] Department of Environment Systems, The University of Tokyo, Tokyo 113-0033, Japan
shoji@k.u-tokyo.ac.jp

Commission IV, WG IV/1

KEY WORDS: Wind, Direction, Vector, AMeDAS, Power Generation, Exponential Distribution, Variogram, Central Japan

ABSTRACT:

To study the applicability of geostatistics for vector data, wind velocity data have been analyzed statistically and geostatistically. The study area consists of two districts, the mountainous Chubu and plain Kanto districts in central Japan. For the distribution of wind speeds, exponential distribution was fitted well in both districts. Temporal experimental variograms of wind speeds, directions and velocities suggest daily duration, and wind is stronger by day than by night. While some spatial experimental variogram of wind speeds are traditional spherical schemes showing clear nugget effects, sills and ranges whose values vary 50–130 km, other variograms are not traditional schemes illustrating flat or gradually increasing curves. If variograms of momentary wind speeds have no range an empirical variograms of temporally averaged wind speeds does not also show a range.

1. INTRODUCTION

In order to find a location suitable for wind power generation, it is necessary to know spatial distribution of wind. Geostatistics is a powerful tool to estimate a spatial distribution of geosciences variables. Accordingly, geostatistical tools are applied to the estimation of wind distribution in space. However, we must overcome two difficulties for the applications. The first one is that wind data are measured at a moment in time at a spatial location. Generally, averaging in time is applied to reduce data variation. We have to check the effect of such averaging along a time axis on analytical results of originally momentary wind data. The other problem is that wind velocities are vector. Geostatistical tools have been developed for commonly scalar data. A vector datum always consists of two values at least and very different mathematical characters than the scalar data.

Vector data such as not only wind velocity, but also topological slope, folded geological structure, underground water flow, and others are very common in geoscience fields. In this paper, several new geostatistical tools are proposed for the treatment of the vector data. The analytical results from the new tools suggest several viewpoint of assessing the locations of the wind power generation stations.

2. DATA AND AREAS

The Meteorological Agency of Japan has established an automated meteorological observation system named AMeDAS (Automated Meteorological Data Acquisition System). AMeDAS has 1536 stations in the whole area of Japan (377,800 km^2). Each station records rain precipitation, wind velocity, temperature and other meteorological data at every hour. The station density is about 1/250 km^{-2} (41 stations in 100·100 km^2). If the stations are arranged on a tetragonal grid, the average distance between the neighboring stations is 16 km. If they are arranged on a trigonal grid, the average distance is 17 km. Meteorological Agency of Japan have published data of AMeDAS every year as a CD-ROM. The present study uses the data in 1999.

A wind velocity ($\vec{v}$) is a vector in a 2-D space, and is described by the combination of a speed and a direction as follows:

$$\vec{v} = s \cdot \vec{u} \tag{1}$$

where $\vec{v}$ = velocity (vector)
$s = |\vec{v}|$ = speed (scalar)
$\vec{u}$ = direction (unit vector).

Since the direction (θ) is defined as an angle measured clockwise from the north, the unit vector is given as follows:

$$\begin{aligned} u_{\mathrm{NS}} &= \cos\theta \\ u_{\mathrm{EW}} &= \sin\theta \end{aligned} \tag{2}$$

where u_{NS}, u_{EW} = north-south and east-west components of the direction.

In order to compare with wind patterns in mountainous area and plain, the Chubu district characterized by many mountains, and the Kanto district characterized by the widest plain in Japan are selected for the analysis (Figure 1). Both districts are next of each other, and the former is situated west of the latter. Accordingly, wind blows sequentially or contemporarily in both districts.

The Chubu district consists of many folding mountains and volcanoes including Mt. Fuji, which is the highest in Japan. In order to obtain the characteristics of wind in the mountainous district, only the stations whose elevations are higher than 100 m above the sea level are selected in the district. 170 stations are included in this district, and about 60 % of them recorded wind data. The elevation of the highest worked station is 1350 m (the elevation of the highest station is 2730 m). The density of stations is about 1/510 km^{-2}. This means that the nearest station distance is 23 km in a tetragonal grid, or 24 km in a trigonal grid.

The Kanto district consists of many fields and cities including Tokyo, which has the largest population in Japan. The stations lower than elevation 200 m are selected in this district. 81 stations are included in this district, but about 50 % of them recorded wind data. The density of stations is about 1/490 km^{-2}. This means that the nearest station distance is 22 km or 24 km.

3. STATISTICS

3.1 Statistical Distribution of Speeds

First, basic statistics of speed, direction and velocity were obtained from wind data in the Chubu and Kanto districts, and the Chubu-Kanto total area defined by the combination of both districts. We have empirically fitted three distribution functions: normal, lognormal, and exponential distributions. Table 1 summarizes sample correlation coefficients between wind speeds and empirical cumulative frequency distribution functions. All sample correlation coefficients are high. Especially, exponential functions (Figure 2) show high correlation coefficients: –0.999 for the Chubu district, –0.998 for the Kanto district, and –0.9996 for the Chubu-Kanto total area. Figure 2 means that wind is stronger in the plain Kanto district than the mountainous Chubu district. This feature has been consistent, even if the data are treated in each month.

The fact that the correlation coefficient between wind speeds and the cumulative frequency is extremely high ($r^2 > 0.99$) in an exponential distribution is very important for assessing wind power generation, because this means that we can estimate accurately wind speeds in a simulation based on an assumption of a random process (e.g. the Monte Carlo method).

3.2 Rose Diagram of Directions (Wind Rose)

Frequencies of wind directions are presented by a rose diagram named "wind rose". Figure 3 is an example. This diagram is most deviated among treated data. Let us define the average direction as follows:

$$\langle\vec{u}\rangle = \sum \vec{u}_i / n \tag{3}$$

where $\langle\vec{u}\rangle$ = average direction

$\vec{u}_i$ = i 'th datum

n = number of data.

Table 1. Correlation coefficients between wind speed and frequency cumulated from the high value side in some statistical distribution models.

District	Number of –	Statistical Distribution Model		
		Normal	Lognormal	Exponential
Chubu	504027	-0.9668	-0.9923	-0.9992
Kanto	779097	-0.9696	-0.9909	-0.9983
Total	1230815	-0.9657	-0.9929	-0.9996

"Total" means the total area of Chubu and Kanto.

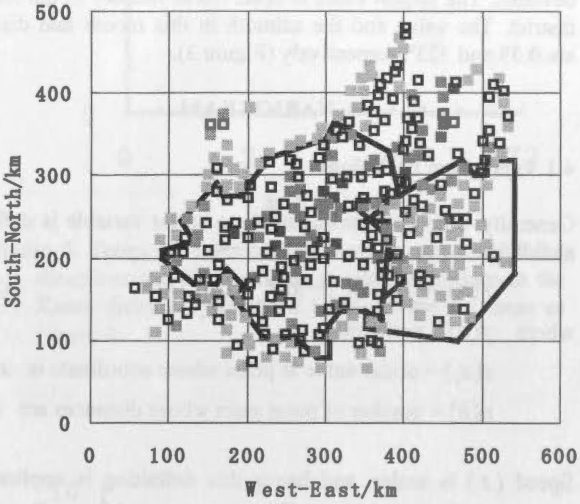


Figure 1. A map showing the locations and the elevations of the station points used in this study (open squares indicate did not work for wind data). The west and east areas by solid lines are the Chubu district and the Kanto district, respectively. Darkness of symbols represents elevations. The boundaries are 1000, 500, 200 m and 100 m.

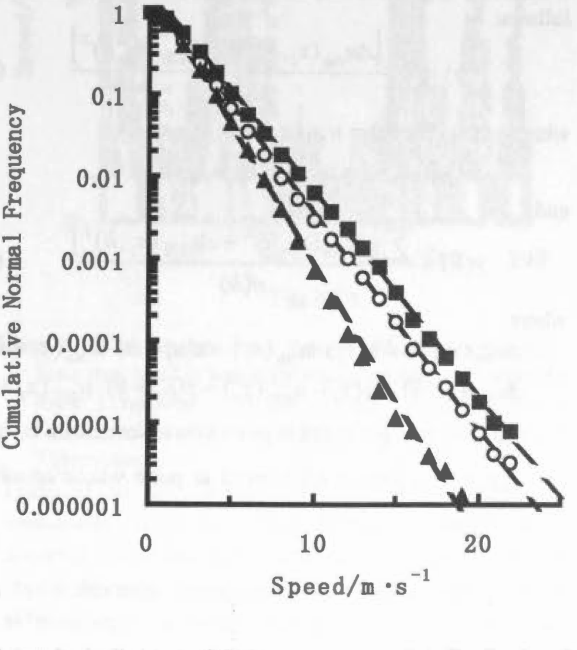


Figure 2. A diagram of fitting an exponential distribution for wind speeds in Chubu (solid triangle) and Kanto (solid square) districts, and the Chubu-Kanto total area (open circle).

If direction is presented by Equation (2), then Equation (3) gives the following equations:

$$\langle u_{\mathrm{NS}} \rangle = \sum u_{\mathrm{NS}i} / n$$
$$\langle u_{\mathrm{EW}} \rangle = \sum u_{\mathrm{EW}i} / n \qquad (4)$$

where $\langle u_{\mathrm{NS}} \rangle$, $\langle u_{\mathrm{EW}} \rangle$ = north-south and east-west components of the average direction

$u_{\mathrm{NS}i}$, $u_{\mathrm{EW}i}$ = north-south and east-west components of i 'th datum.

A large $|\langle \vec{u} \rangle|$ value means that wind directions are largely deviated. The largest value is observed in January in the Kanto district. The value and the azimuth in this month and district are 0.39 and 323°, respectively (Figure 3).

4. VARIOGRAM

4.1 Variogram Equations

Generally, empirical variogram for a scalar variable is defined as follows:

$$\gamma(h) = \sum \{s(x_i) - s(x_i + h)\}^2 / n(h) \qquad (5)$$

where $\gamma(h)$ = variogram

$s(x_i)$ = scalar value at point whose coordinate is x_i

$n(h)$ = number of point pairs whose distances are h.

Speed (s) is scalar, and hence this definition is applicable. Direction ($\vec{u}$) and velocity ($\vec{v}$) are vector, and hence Equation (5) has to be expanded as follows:

$$\gamma(h) = \sum \{\vec{u}(x_i) - \vec{u}(x_i + h)\}^2 / n(h) \qquad (6)$$

where $\vec{u}(x_i)$ = direction or velocity at point whose coordinate is x_i.

If direction is given as Equation (2), Equation (6) is written as follows:

$$\gamma(h) = \frac{\sum \left[\Delta u_{\mathrm{NS}}(x_i, h)^2 + \Delta u_{\mathrm{EW}}(x_i, h)^2 \right]}{n(h)} \qquad (7)$$

where $\Delta u_{\mathrm{NS}}(x_i, h) = u_{\mathrm{NS}}(x_i) - u_{\mathrm{NS}}(x_i + h)$

$\Delta u_{\mathrm{EW}}(x_i, h) = u_{\mathrm{EW}}(x_i) - u_{\mathrm{EW}}(x_i + h)$,

and

$$\gamma(h) = \frac{\sum \left[\Delta v_{\mathrm{NS}}(x_i, h)^2 + \Delta v_{\mathrm{EW}}(x_i, h)^2 \right]}{n(h)} \qquad (8)$$

where

$\Delta v_{\mathrm{NS}}(x_i, h) = s(x_i) \cdot u_{\mathrm{NS}}(x_i) - s(x_i + h) \cdot u_{\mathrm{NS}}(x_i + h)$

$\Delta v_{\mathrm{EW}}(x_i, h) = s(x_i) \cdot u_{\mathrm{EW}}(x_i) - s(x_i + h) \cdot u_{\mathrm{EW}}(x_i + h)$

$s(x_i) = |\vec{v}(x_i)|$ = speed at point whose coordinate is x_i

$s(x_i + h) = |\vec{v}(x_i + h)|$ = speed at point whose coordinate is $x_i + h$.

In this paper, empirical semivariograms were calculated in all cases, and are shown in figures. However, "variogram" is used instead of "empirical semivariogram", because their values are not important.

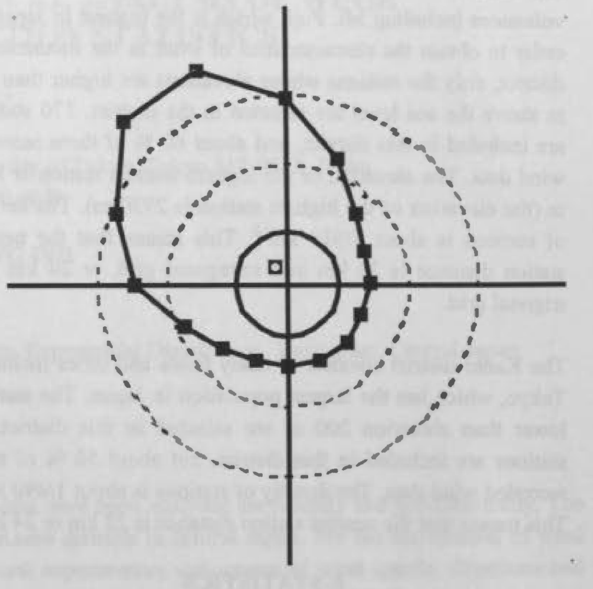


Figure 3. A wind rose showing frequencies of wind directions in the Kanto district (January, 1999). The polygon represents frequencies of directions, and the area of the central circle shows the proportion of calm (wind speed is 0). The open square shows the gravity center, when frequencies of directions are represented by columns standing at the rim of the calm's circle.

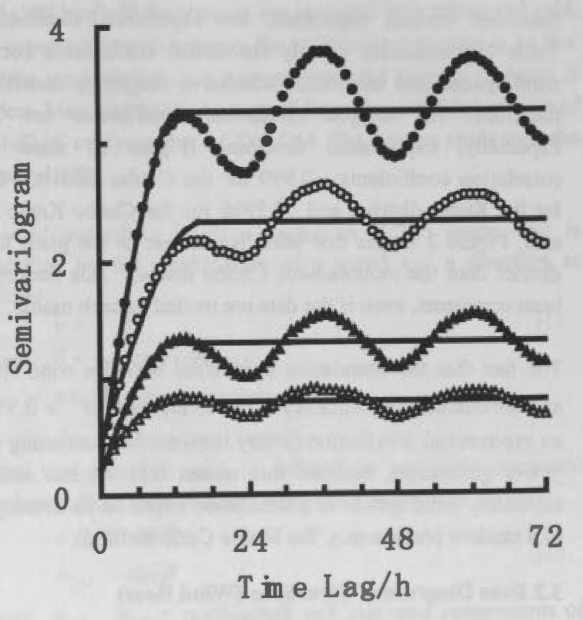


Figure 4. Temporal experimental semivariograms of wind speeds and fitted models at selected stations in the Kanto district. Hole effects showing a daily period are remarkable. Locations: open triangle = Yorii, solid triangle = Kuki, open circle = Urawa, and solid circle = Tokorozawa.

4.2 Temporal Variograms

AMeDAS records a wind velocity (speed and direction) at a moment at every hour. Wind velocities vary every moment. For this reason, in order to know temporal continuity of wind velocities, experimental variograms as a function of time were calculated using all annual data. Figures 4, 5 and 6 show temporal variograms of speeds, directions and velocities in Saitama Prefecture, the Kanto district.

The most remarkable point observed in the temporal experimental variograms is a hole effect representing a daily period. This period is not caused by sea, because every station is located more than 20 km from seashore. Figure 7 shows wind speeds from 1st to 10th, January. According to this diagram, wind was strong by day, and weak by night. The same feature is also observed in temporal variograms of wind speeds, directions and velocities in Yamanashi Prefecture, the Chubu district.

Every temporal variogram is well approximated by a traditional spherical model. The ranges of the models vary from 8 to 19 hours (Figures 4-6). This means that a wind condition continues generally a few hours

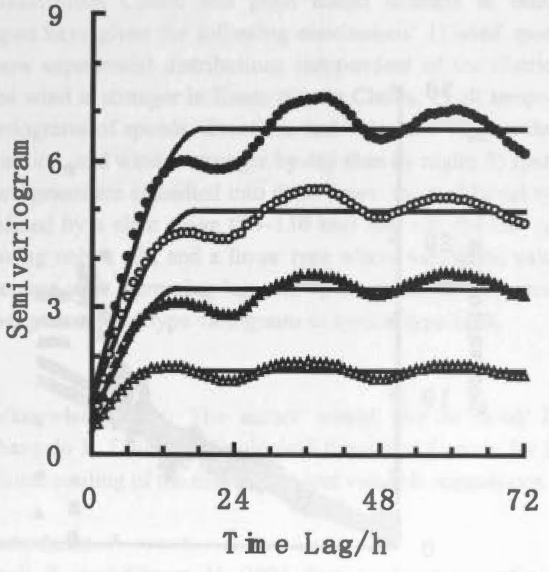


Figure 6. Temporal experimental semivariograms of wind velocities and fitted models at selected stations in the Kanto district. Symbols of locations are the same as Figure 4.

4.3 Spatial Variograms

Figures 8, 9 and 10 show spatial variograms in the Kanto district at 1 to 10 o'clock on 1st, January. We can see three patterns of variograms. The first one is traditional spherical variograms showing clear ranges and sills. The second one is flat (i.e. shows only sills). The third type is linearly increasing with increasing lag. It is not clear at present why these three types appear in this study.

Rain precipitation data are always accumulated values.

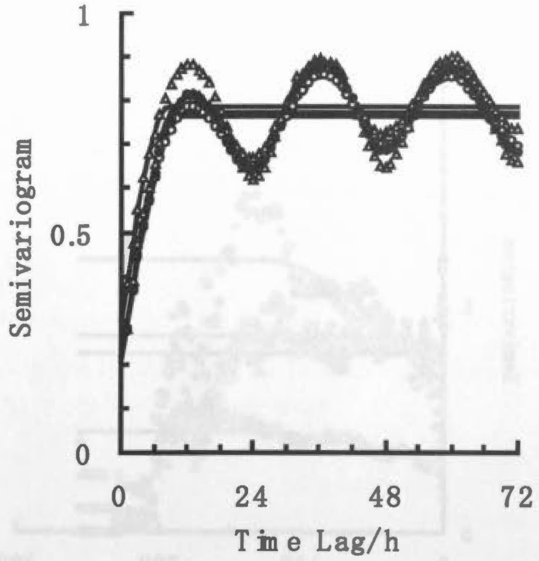


Figure 5. Temporal experimental semivariograms of wind directions and fitted models at selected stations in the Kanto district. Symbols of locations are the same as Figure 4.

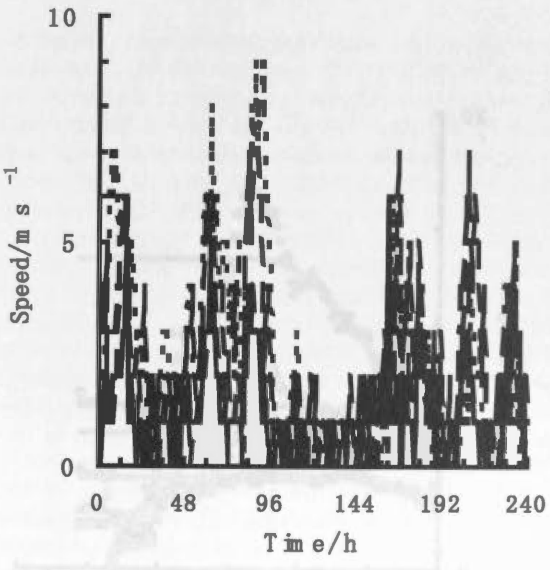


Figure 7. Wind speeds from 1st to 10th, January, 1999. Note that wind is generally strong by day, and weak by night. Locations: solid line = Yorii, dot-dashed line = Kuki, dashed line = Urawa, and dotted line = Tokorozawa.

Different to this, every wind datum is a record of a momentary condition. This seems to imply that an accumulated or averaged value has no meaning. However, the variograms in Figures 4-6 indicate some continuity existing in

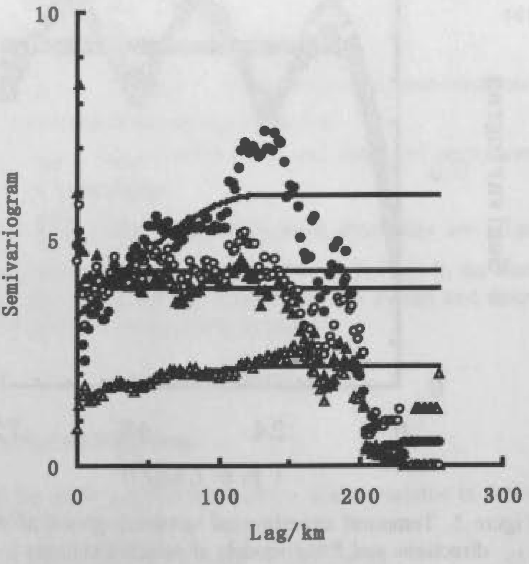


Figure 8. Spatial experimental semivariograms of wind speeds and fitted models in the Kanto district at selected times on 1st January 1999. Times (o'clock): open triangle = 2, solid triangle = 7, open circle = 8, and solid circle = 9.

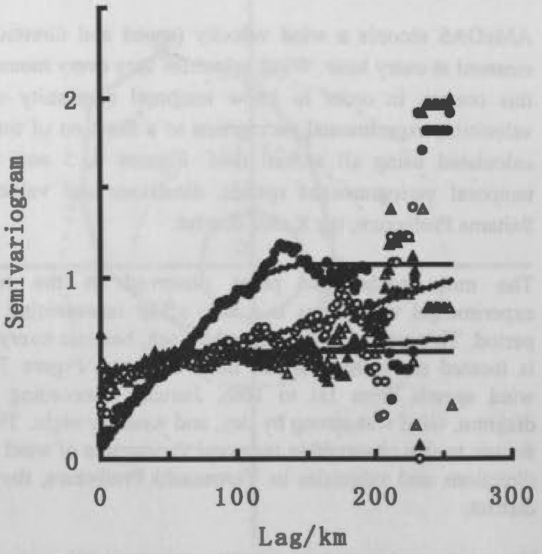


Figure 9. Spatial experimental semivariograms of wind directions and fitted models at selected times in the Kanto district at every hour on 1st January 1999. Symbols of locations are the same as Figure 8.

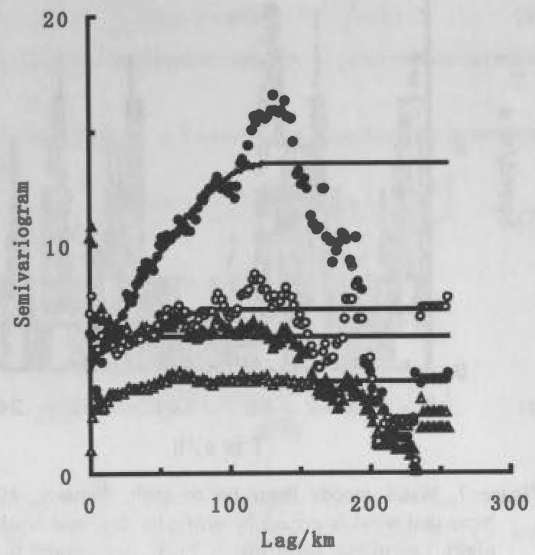


Figure 10. Spatial experimental semivariograms of wind velocities and fitted models in the Kanto district at selected times on 1st January 1999. Symbols of locations are the same as Figure 8.

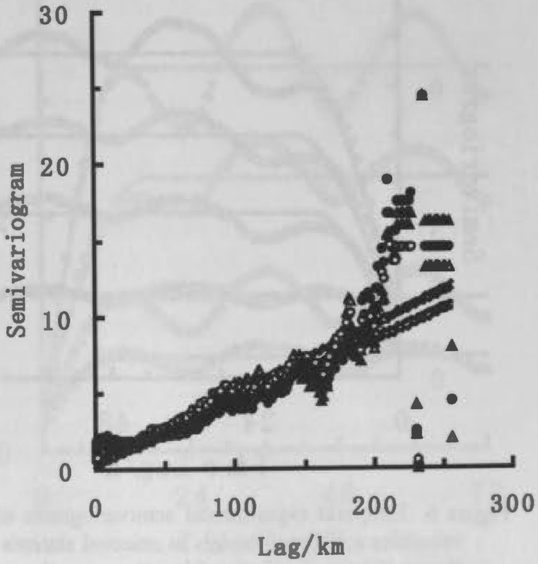


Figure 11. Spatial experimental semivariograms of averaged wind speeds and fitted models in the Kanto district from 14 to 23 o'clock on 7th January 1999. The duration (hour): open triangle = 0, solid triangle = 1, open circle = 4, and solid circle = 9. Note that all variogram are not typical.

the temporal distribution of wind, and therefore suggests a possibility that moving averages along time has meaning. Figure 11 shows variograms of temporarily averaged wind speeds. The data were obtained from 14 to 23 o'clock on 7th, January. No temporal variogram of momentary data has a clear range and sill in this duration.

The variograms shown in Figures 11 were calculated for speed data averaged during 2 to 9 hours. Any of the diagrams does not show a clear range. This point is quite different to the rain precipitation case reported by Shoji and Kitaura (2001), where variograms of accumulated data show clear ranges and sills.

The facts that some variograms of momentary wind data are not typical, and that they are not improved (i.e. do not become typical) by accumulation means that kriging will not be able to give a good estimate for wind data.

5. CONCLUSIONS

Statistical and geostatistical analyses of wind data in the mountainous Chubu and plain Kanto districts in central Japan have given the following conclusions: 1) wind speeds show exponential distributions independent of the districts, and wind is stronger in Kanto than in Chubu; 2) all temporal variograms of speeds, directions and velocities suggest daily duration, and wind is stronger by day than by night; 3) spatial variograms are classified into three types: the traditional type defined by a clear range (50–130 km) and sill, the flat type having only a sill, and a linear type where variogram values increase with increasing lag; and 4) the accumulation cannot change untypical type variograms to typical type ones.

Acknowledgments: The author would like to thank Dr. Chang-Jo F. Chung of Geological Survey of Canada for his critical reading of the manuscript and valuable suggestions.

References

Shoji, T., and Kitaura, H., 2001. Statistical and geostatistical analysis of rain fall in central Japan. In Proc. IAMG2001 (Intern. Assoc. Math. Geol.), Cancum, Mexico, Sept. 6-12, 2001, CD-ROM (Session D)

A GENERAL OBJECT-ORIENTED SPATIAL TEMPORAL DATA MODEL

Bonan Li, Guoray Cai

bonan@psu.edu, cai@ist.psu.edu, School of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802, USA

Commission IV, WG IV/1

Key Words: Spatial temporal data model, Object-oriented, Multiple semantics, Unified framework

ABSTRACT

Current research on spatial temporal data modelling meets many problems, such as the separation of space and time, application-driven, not incorporating human cognition and confusing of representation for discrete and continuous spatial change. This paper proposes a general object-oriented model that tries to overcome above problems. The model is directed by human cognition on spatial object that each object have what/where/when attributes. Attributes of theme, space, and time are tightly linked to each other in a unified object class. The precise and flexible structure of this model supports multi-semantics of space and time. Besides complex spatial relationships, it is also capable of modelling linear/branch/cyclic time order, various spatial change, and complex temporal relationships. Verified by different spatial temporal scenarios, this model is proven to be extendable to different spatial temporal applications.

1. INTRODUCTION

Space and time are most important and basic information in the real world. Many applications demand for spatial temporal support (Spaccapietra 2001), such as cadastral systems that capture the histories of land parcels, traffic management systems computing traffic network and traffic volume at different time, forecast-prediction systems recording weather change process, etc. But existing database techniques couldn't support spatial temporal information management well (Tryfona and Jensen 1999). After almost 20 years' research, representation of space and time in databases and functional applications are still problematic (Peuquet 2001). This paper presents a universal object-oriented framework for spatial temporal data modelling. Spatial temporal data modelling aims to extend the existing data models to include space and time in order to better describe our dynamic real world.

One problem with previous data models is that spatial and temporal aspects of databases are modelled separately (Sellis 1999) (Martin Erwig 1999). Spatial database focuses on supporting geometries (Guting 1994), while temporal databases focus on the past state (Tansel and Clifford 1993). But in many circumstances, such as environmental monitoring, resource management, transportation scheduling, etc, spatial and temporal attributes should be connected together. Many current systems can handle only one aspect of space and time. Spatial systems always fail to cater for many temporal aspects in a dynamic environment (Abraham and Roddick 1999). Though many people have found the necessity of integration of space and time in one environment, by far, little such work has been done.

Application-driven is another problem. Many database systems concentrate on the definition of a particular spatial temporal model that is related to certain application. The result is that more and more different models appeared. Each model focuses on a specific set of spatial temporal features (Goralwalla, Ozsu et al. 1998). When encountering other features and applications, the model doesn't work. So what we should do is to build an overall framework that can be extended to various applications, not driven by various applications generating different models (Mennis, Peuquet et al. 2000). Furthermore, application specific modelling will be more efficient if it is based on a generic model (Raza and Kainz 1999).

The third challenge for spatial temporal modelling is on the discrete representation of continuous phenomena. Many temporal systems can only capture discrete snapshot of real world (Hornsby and Egenhofer 2000), and hence couldn't model continuously changing object. But many temporal phenomena in nature are continuous, such as the clouds, rainfall, etc. Designing new model that can represent continuous process is one aspect, the other aspect is that the computer always records data in a discrete way (Peuquet 2001). How to mediate the discrete and continuous view is really a big problem when building the data model.

The fourth problem is representation of data should be natural to human. The structures of space and time are identified as essential for the realization of cognitive systems (Freksa 1998). According to Donna J. Peuquet and her group (Mennis, Peuquet et al. 2000), models of spatial temporal data in geographical database representations must incorporate human cognitive principles. Human knowledge of the dynamic geographical world comprises of three different (and interrelated) subsystems that handle *what*, *where* and *when* aspects of object properties (Mennis, Peuquet et al. 2000) (Sinto 1978). Theme-based model, location-based model and time-based model separately describe one subsystem. From this view, these 3 kinds of model are all single semantic models. According to human cognitive principle, the model we build should cover all of these 3 subsystems. Donna J. Peuquet and her group's pyramid framework show how to incorporate three subsystems of human cognition into data modelling.

From above analysis, we should build a unified infrastructure that integrate space and time, mediate discrete and continuous representation, describe when/where/what systematically, and be extensible to various applications. A multi-semantic model cannot ultimately be generated from extension of current single semantic models such as ER model and location/time/theme-based model. Object-oriented approach with its characteristic of inheritance and aggregation is capable of capturing the various notions of space and time and reflecting them into a single framework extensible to different applications.

This paper focuses on the unified models of space and time using object-oriented approach. In particular, we propose a conceptual, object-oriented spatial temporal data model based on Donna J. Peuquet's pyramid framework. Section 2 presents a unified object-oriented spatial temporal framework. Section 3 verifies the model using different spatial temporal scenarios, and discusses the advantages and disadvantages of this model. Section 4 gives a conclusion and discusses future research in this area.

2. THEORETICAL FRAMEWORK

2.1 Ontology of Space and Time

A uniform spatial temporal data model must be based on certain ontology of space and time, which is much more important than the multi-representation. Though many models are highly expressive, they present limitations to adequately model many applications because those models do not define enough or reasonable space-time primitives that would allow for a satisfactory representation of space and time (Borges, Jr. et al. 2001).

Sinto (Sinto 1978) defines geographic information with attributes as theme/location/time. He believes no geographic data should be entered into an information system without all three of these attributes present. Sinto further divides spatial data into geometric information and spatial relationship descriptions (topological, metric and algebraic relationship). Peuquet (Peuquet 1994) and Yuan (Yuan 1994) suggest a what/where/when structure. They argued that any spatial object has theme/space/time attributes. Goralwalla (Goralwalla, Ozsu et al. 1998) defines temporal primitives in a temporal object-oriented data model as combination of instant/interval/span and discrete/continuous domain and determinacy/indeterminacy. In this model, the author also defines linear/sub-linear/branching time order and valid/transaction/event time history. Wang (Wang and Cheng 2001) defines spatio-temporal behavior as continuous/discrete/stepwise change. Tryfona (Tryfona and Jensen 1999) divides spatial temporal applications into 1) objects with continuous motion, such as a moving car. 2) Discrete changes of and among object. In this type of applications, the characteristics of objects, such as shape, position may change discretely in time. 3) Continuous motion as well as changes of shape. That means a moving object with changing properties. He also divides spatial relationships into topological/directional/metric subsets. Temporal aspects in his model are valid/transaction/existence time. Raza and Kainz (Raza and Kainz 1999) defined time aspects in his model as linear/branching/cyclic subclasses and temporal operations as before/after/during/meets.

There are also many other papers and models define their ontological framework of space and time. But the main contents are same to literatures listed above. One problem of those models is they do not support a whole spatial temporal structure. Some of them support only space, some of them support only time, some of them support incomplete primitives of space and time.

Having identified the ontological structure of space and time, the next task is to model these spatial temporal primitives.

2.2 Data Modelling Approaches

Popular data modelling approaches include extending relational models, extending vector/raster models, time-based models and object-oriented models.

Driven by applications, much of the research on spatial temporal data models focus on extension to relational models (Renolen 2000), such as EER (extended entity-relation model) (Wang and Cheng 2001), STER (spatial temporal entity-relationship model) (Tryfona and Jensen 1999), ERT (entity-relationship with time model) (Renolen 2000), STSQL (Bohlen and Jensen 1998), etc. These models generalize new relations to manipulate multiple temporal as well as spatial attributes.

Extending vector/raster models are conceptually location/theme-based models. Within existing GIS, available such models include snapshot and update models (Li 2001) (Peuquet and Duan 1995). These two models are all based on a time sequence that is evenly divided. Snapshot records complete static maps at each time, while update model records complete initial map, then only changed information will be recorded at each time. This approach is straightforward. Nevertheless, because the locations are recorded in same interval, those locations where no change occurred result in the storage of redundant information, meanwhile, change occurred between two locations are not explicitly stored (Hornsby and Egenhofer 2000) (Peuquet and Duan 1995).

Time-based model records the timestamp for any change and associated details describing each specific change in temporal order, such as ESTDM (event-based spatial temporal data model) (Peuquet and Duan 1995). With certain time granularity, temporal locations will be recorded only when changes occur. For successive or continuous change, recording timestamp can be determined by certain rule or when the change accumulates to a meaningful status. So the distance between two timestamps can be different.

Object-oriented approaches originated in programming languages (Worboys and Hearnshaw 1990). In object-oriented data modelling, all conceptual entities are modelled as objects. Object class, generalization, specialization, aggregation, and association are the well-known concepts of object-oriented approach (Renolen 2000) (Guting 2000). Each class has a class ID, a list of attributes and a list of behaviours. Object is the realization of abstract class. Each object has certain attributes and performs according to certain operations. Inheritance of class includes two concepts: generalization and specialization. A class may be specialized into many sub classes, or a super class may be generalized from some classes with similar attributes and behaviours. A set of homogeneous classes can be associated to form a higher-level class, while a set of different classes can be aggregated to form a higher-level class. Not like above data modelling approach, though object-oriented approach has certain rules, object-oriented data models are largely different with each other, such as IFO (Is-a relationships, Functional relationships, complex Objects) (Worboys and Hearnshaw 1990), TOODM (Temporal Object-Oriented Data Model) (Goralwalla, Ozsu et al. 1998), MADS (Modelling for Application Data with Spatio-temporal features) (Parent, Spaccapietra et al. 1999), etc.

3. A GENERAL OBJECT-ORIENTED SPATIAL TEMPORAL DATA MODEL

The purpose of data modelling is to bring about the design of a database that performs efficiently, contains correct information, whose logical structure is natural enough to be understood by users, and is as easy as possible to maintain and extend. Based on above analysis and former work, this paper proposes an object-oriented spatial temporal data model, which is extended from Peuquet's what/where/when triad structure. Following OMT standard, this model adopted classification/specialization/association/generalization/aggregation strategy.

3.1 Class Structure

3.1.1 Super Class: Super-class of the model is shown as below. It's a void class that defines classID and basic attributes and behaviours of any object located in spatial and temporal dimension. A superclass can be inherited by subclasses in different applications contexts.

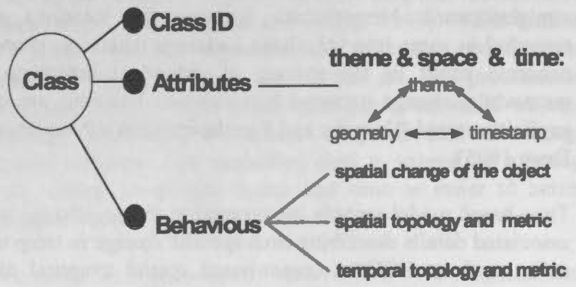

Class ID
Class
Attributes
theme & space & time:
theme
geometry
timestamp
spatial change of the object
Behavious
spatial topology and metric
temporal topology and metric

Figure 1. Superclass structure of the model

3.1.2 Structure of Attributes: Each object has attributes of theme, space and time.

Theme attributes record the information of "what is this object" and other related property information. For example, if the theme of this object class is "city", a city must have attributes such as Nation or Province it belongs to, Population, etc. Theme can be inherited according to certain classification hierarchy. For example, theme can be classified as transportation, settlement, water system, cadastre, natural resource, etc.

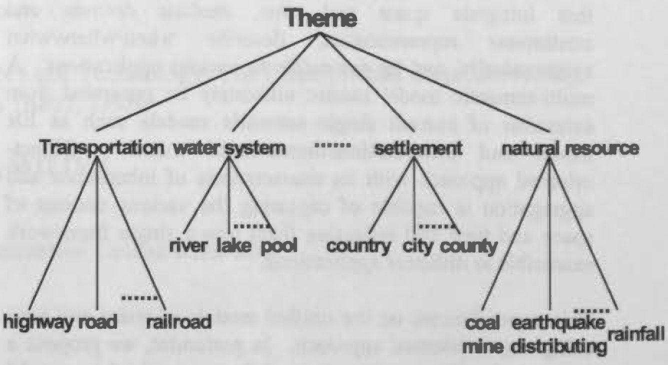


Figure 2. Theme classification hierarchy

Space attributes record geometry information of the object, such as whether it is a point, line, or polygon, and its location. Space attributes can be inherited into subclasses such as the point class, line class and polygon class.

Time attributes are represented through timestamps associated with theme and geometry information. When properties or geometry information of an object changed, a new timestamp will be recorded for that object. That means only initial information and changed information will be recorded. Each timestamp is an instant. Interval is the time between two timestamps. This model can support linear, branch and cyclic time order. The demonstration of different time order will be shown below in this paper.

In this model, theme, spatial, time attributes are recorded as 3 tables. Each attributes use one list. For example, a lake object has theme attributes as *"name"*, *"volume"* and spatial attributes as *"border"*. During different season and time of a year, the volume of the lake will change, and the border of the lake expands or contracts. Whenever the volume or the border of the lake changes, a timestamp will be recorded. Thus, as shown in fig. 3, each lake object has a name list, a volume list (*v1, v2, v3*), a border list (*b1, b2, b3*) and a timestamp list (*Jan, Mar, Apr, Jun, July, Sep, Oct, Dec*). Name list has only one element. It points to all border elements and timestamps. Each record of volume relates to a record of border and a timestamp; each record of border relates to a record of volume and a timestamp; each timestamp relates to a record of volume and border. In winter (*January*), the lake has the smallest volume and border (*v1 and b1*). It lasts for 2 months. From March to April, the volume and border change from *v1*, *b1* to *v2*, *b2*. After 2 months, it changes to *v3*, *b3*. From Sep., it goes down to *v2*, *b2*. Then, in next January, it is down to the smallest volume *v1* and border *b1*.

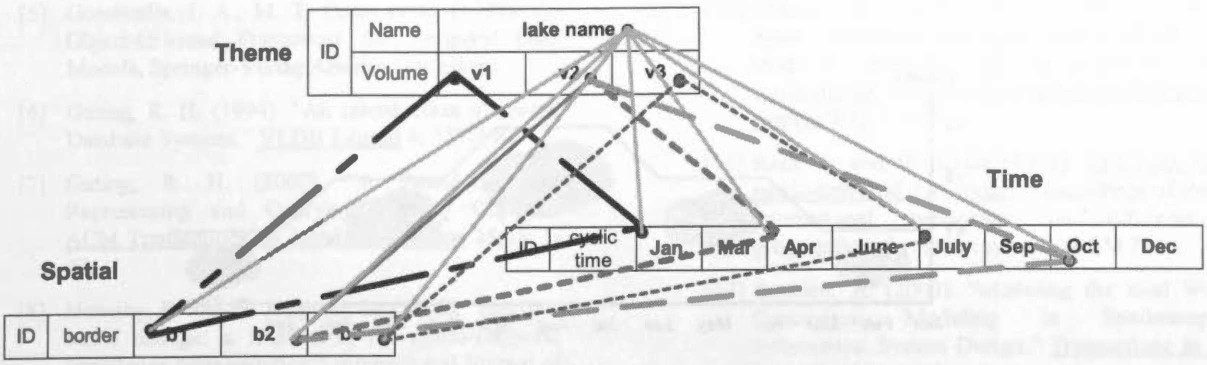


Figure 3. Relationship between theme, space and time

There are 3 kinds of relationships among theme, space and time: one-one, one-many, and many-many. In a one-one relationship, theme, space and time all changed, such as *v1-border1-Jan* and *v3-border3-July* in last example. In one-many relationship, one theme attribute may be related to more spatial attributes and many timestamps. That means that theme attribute doesn't change during this interval while the spatial attributes changed, such as the attribute *"name"*. Also, in some circumstances, spatial attributes of an object are static while theme attributes change. Another special case is that the timestamp list may have just one element. That means this object is static.

The theme list, spatial list and time list can have multiple rows, such as the theme list in this example. The theme list includes two attributes *"name"* and *"volume"*. If an object is made up of many spatial elements such as a river has many parts *(p1, p2, p3)*, its spatial list will have 3 rows to record these 3 part's spatial information.

3.1.3 Behaviours: This model defines three kinds of behaviours. They are spatial change of the object, spatial relationship (topology and metric) with other objects, temporal relationship with other objects.

Spatial change of the object includes: 1) Stay-in: object stay-in one status (location & properties) during certain time; 2) Transform-between: object transform from one status to another during certain time; 3) Appear/disappear; 4) Increase/decrease; 5) Splitting; 6) Moving; etc.

Spatial relationship between two objects includes topology relationship (such as intersection, inclusion, adjacency, disjunction, equality, merging, etc.) and metric relationship (such as the distance of two points).

Temporal relationship between two objects also includes topology and metric relationship, such as before, overlap, finish, equal, during, meet, start, etc.

3.2 Generalization, Specialization, Association and Aggregation

This model supports generalization, specialization, association and aggregation to define relationships among objects.

Generalization and specialization are contrary constructs. Generalization extracts similar properties of different classes into a super-class, while specialization extends a class to some sub-classes. Each sub-class has same properties as well as special properties. For example, in Fig.2, transportation is a super-class, which can be specialized into highway, road, railroad, etc. In the contrary, highway, road, and railroad can be generalized into transportation.

Association can also be called grouping which enables some objects of same type to form an object of higher-level type. For example, line is the association of many points.
Aggregation is a construct that enables different objects to be amalgamated into a higher-level object. The higher-level object has all of the attributes of those different objects. For example, the water system in a country is the aggregation of many rivers, lakes, and pools.

The super class mentioned above defines basic void attributes and behaviours of a spatial temporal object. This super class can be specialized into sub-classes. Each sub-class inherits the basic void attributes and behaviours from the super-class, and realizes the attributes and behaviours with special content. Each sub-class can also have its own sub-classes. With this class hierarchy and the aggregation and association of different classes, multi-semantic space and time can be expressed.

3.3 Demonstration

There exist various spatial temporal phenomena, such as "stay-in", "transform-between", "appear/disappear", "splitting", etc. In this section, we will demonstrate how our model captures some spatial changes inside an object or among objects, and temporal changes in different time orders: linear, cycle, or branching.

In the lake example mentioned above, there are two kinds of spatial change inside the object: stay-in and transform-between. As shown in Fig.4, the lake stays in volume *v1* and border *b1* from the beginning of January to the end of February. From the beginning of March to the beginning of April, the lake transform from *v1* to *v2*, *b1* to *b2*. Supposing every year, the river has same changes. During one year, the change of the river border and volume is linear, while in many years change is cyclic. In fig. 3, the cyclic time order is marked in the timetable.

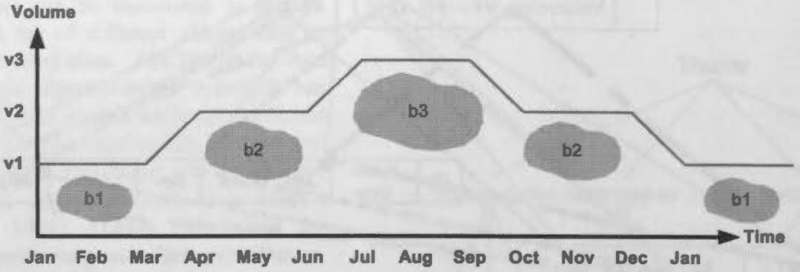


Figure 4. Volume and border change of the lake

As shown in Fig. 5, a country (*ID0*) has name *A0*, Population *P0*. It appeared at time *T0*, and split into *ID1*, *ID2* at time *T1*. "Split into" and "come from" define the relationship among these 3 instances. After splitting, *time0* branched into *time1* and *time2*. From this example, we can see, in this model, the branching time order is expressed as relationship among objects.

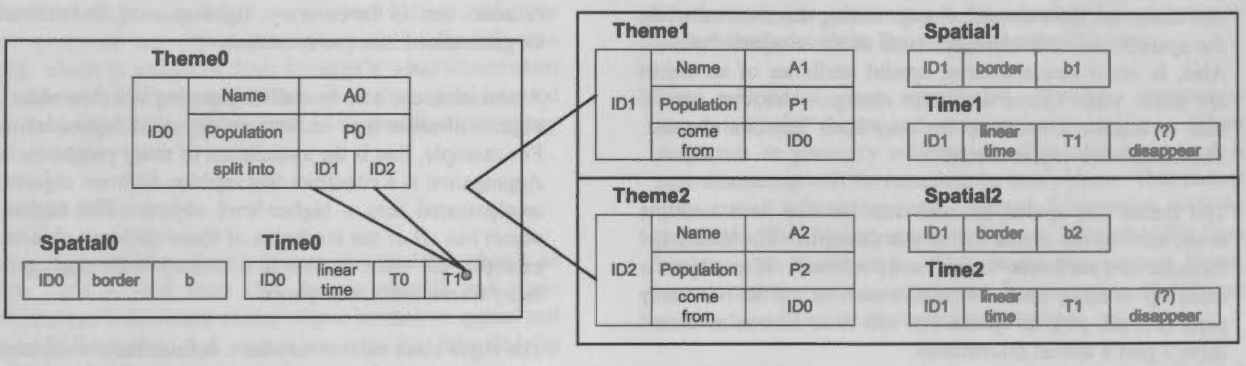


Figure 5. Splitting demonstration

4. DISCUSSION AND CONCLUSIONS

This paper reviewed current research on spatial temporal data modelling and finds some problems in this field. Then, this paper provides a unified object-oriented model that partially overcomes some problems. Based on human cognition, the model linked together theme, space and time through common object references. Such structure allows unified operations on space and time, such as navigation, tracking and query. This unification means the effectiveness both in function, consistency, and computational complexity.

Application-driven models are usually single-semantic or incompletely multi-semantic. This model is built up from the basic attributes and behaviours of spatial temporal objects. Based on the object-oriented strategy, it's extensible to any applications by generalization, specialization, association and aggregation. Thus, this model can completely support the multiple semantics of space and time.

Another advantage of this model is no matter querying time or querying space, this model has the same efficiency. Not like temporal model or spatial model. Temporal model is designed for time query. It is quite efficient when querying time through timestamp list. But when querying space, it is inefficient. Spatial model has the similar problem. In this model, each time table, each space table and each theme table has an object ID as the index. When querying time among different objects, the model will go directly to query all time tables using ID index, no need to search all objects. Things are similar when querying space and theme.

Although the research has made a progress, much work still needs to be done. Efficient implementations will be carried out in further research. Work on specification of integrating space-time query language is also under way.

REFERANCE

[1] Abraham, T. and J. F. Roddick (1999). "Survey of Spatio-Temporal Databases." Geoinformatica 3(1): 61-69.

[2] Bohlen, M. and C. S. Jensen (1998). Spatio-Temporal Database Support for Legacy. 1998 ACM Symposium on Applied Computing, Atlanta, Georgia.

[3] Borges, K. A. V., C. A. D. Jr., et al. (2001). "OMT-G: An Object-Oriented Data Model for Geographic Application." GeoInformatica 5(3): 221-260.

[4] Freksa, C. (1998). Spatial and Temporal Structures in Cognitive Processes, Springer Verlag Heidelberg.

[5] Goralwalla, I. A., M. T. Ozsu, et al. (1998). An Object-Oriented Framework for Temporal Data Models, Springer-Verlag ABerlin Heidelberg.

[6] Guting, R. H. (1994). "An Introduction to spatial Database Systems." VLDB Journal 4: 357-399.

[7] Guting, R. H. (2000). "A Foundation for Representing and Querying Moving Objects." ACM Transactions on Database Systems 25(1): 1-42.

[8] Hornsby, K. and M. J. Egenhofer (2000). "Identity-based change: a foundation for spatio-temporal knowledge representation." International Journal of Geographical Information System 14(3): 207-224.

[9] Li, B. (2001). Multi dimensional dynamic and VR visualization. Institute of Geographical Sciences and Resources, CAS.

[10] Martin Erwig, R. H. G., Markus Schneider, Michalis Vazirgiannis (1999). "Spatio-Temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases." Geoinformatica 3(3): 269-296.

[11] Mennis, J. L., D. J. Peuquet, et al. (2000). "A Conceptual Framework for Incorporating Cognitive principles into geographical database representation." International Journal of Geographical Information System 14(6): 501-520.

[12] Parent, C., S. Spaccapietra, et al. (1999). Spatio-temporal conceptual models: data structures + space + time. Conference on Information and Knowledge Management Proceedings of the 7th international symposium on Advances in geographic information systems, ACM Press.

[13] Peuquet, D. J. (1994). "It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems." Annals of the Association of American Geographers 84(3): 441-461.

[14] Peuquet, D. J. (2001). "Making Space for Time: Issues in Space-Time Data Representation." GeoInformatica 5(1): 11-32.

[15] Peuquet, D. J. and N. Duan (1995). "An event-based spatiotemporal data model(ESTDM) for temporal analysis of geographical data." International Journal of Geographical Information System 9(1): 7-24.

[16] Raza, A. and W. Kainz (1999). Cell tuple based spatio-temporal data model. Proceedings of the 7th international symposium on Advances in geographic information systems, ACM Press.

[17] Renolen, A. (2000). "Modeling the Real World: Conceptual Modeling in Spatiotemporal Information System Design." Transactions in GIS 4(1): 23-42.

[18] Sellis, T. (1999). Research issues in Spatio-temporal Database Systems, Springer-Verlag Berlin Heidelberg.

[19] Sinto, D. F. (1978). "The Inherent Structure of Information as a constraint to Analysis: Mapped Thematic Data as a Case Study." Harvard Papers on GIS 7: 1-17.

[20] Spaccapietra, S. (2001). "Editorial: Spatio-Temporal Data Models and Languages." Geoinformatica 5(1): 5-9.

[21] Tansel, A. U. and J. Clifford (1993). Temporal Databases: Theory, Design, and Implementation, Benjamin/Cummings Publishing Company.

[22] Tryfona, N. and C. S. Jensen (1999). "Conceptual Data Modeling for Spatiotemporal Applications." Geoinformatica 3(3): 245-268.

[23] Wang, D. and T. Cheng (2001). "A spatio-temporal data model for activity-based transport demand modelling." International Journal of Geographical Information System 15(6): 561-585.

[24] Worboys, M. F. and H. M. Hearnshaw (1990). "Object-oriented data modeling for spatial databases." International Journal of Geographical Information System 4(4): 369-383.

[25] Yuan, M. (1994). Wildfire conceptual modeling for building GIS space-time models. Proceedings of GIS/LIS '94, Phoenix.

MODELLING TURNING RESTRICTIONS IN TRAFFIC NETWORK FOR VEHICLE NAVIGATION SYSTEM

J. JIANG *, G.HAN, J.CHEN

National Geomatics Center of China,
No.1 Baishengcun, Zizhuyuan, Beijing, 100044, P.R.China
jjie@nsdi.gov.cn

Commission IV, WG IV/1

KEY WORDS: ITS, in-car navigation, navigable database, data modeling, turning restriction

ABSTRACT:

Vehicle navigation is currently the most widespread and successful consumer application of GIS-T. It is based on positional accurate data and comprehensive transportation-related content. A specially-made navigable database is essential for the key functions of the vehicle navigation systems, for example, route finding and driver guiding. There are lots of research challenges in generating a satisfied navigable database. Among them the data modelling is the most important comparing with representation, data processing, etc., especially the modelling of the turn restrictions of the traffic network. There are three levels of data modelling for the vehicle navigation systems, i.e., conceptual modeling, logical modeling, and algorithmic modeling. The authors of this paper think that special consideration should be put onto the algorithmic model, because it represents not only the relationships between the transportation elements but also the real-time and dynamic traffic restrictions in the traffic network. This paper discussed firstly the properties of the turning restrictions in road networks. The model in conceptual model (GDF 4.0) and logical model (UNETRANS) are introduced. Concerning the algorithmic model, the authors analyzed the disadvantages of the commonly used data structure for representing the turning restrictions. And proposed a new link-based data structure, where a node-link table is used to represent the connectivity of the road network, and a link-link table is used to represent the turning restrictions. The algorithm for route finding was also modified accordingly. The results of the experiment show that the efficiency of route finding is obviously improved with the new method.

1. INTRODUCTION

GIS-T is the application of geographic information systems in transportation. It is now one of the most important application areas of GIS technology [Waters, 1999]. Vehicle navigation is currently the most widespread and successful consumer application of GIS-T, with huge profit already in the developed countries and huge market potential in the developing countries as China. For example, the market volume of in-vehicle navigation and information systems alone in China is expected to be over 2 billion Euro in the next 12 years, and the intelligent transport systems using digital road map as a basis will reach a market volume of 10 billion Euro in China for the same time horizon. Thus lots of studies in GIS field of China are now focusing on the methodologies and technologies related to in-car navigation system. One of them is the study of navigable database for vehicle navigation system, which need to modelling traffic related features and various traffic restrictions.

Two key functions of the vehicle navigation system are route finding and driver guiding. That is, when the driver defines two points on a given road network (one is the start point and the other is the destination point), the system should provide an optimum path connect these two points based on multiple impedances, such as the geometric distance, time (speed limit), cost of travel (where data is available) , as well as the traffic conditions. Then generate turn-by-turn guidance during the driving. And further, dynamically find a new route if the current route has some problem to go through (such as due to traffic accidents or jams) of if the driver has taken a wrong turn. A specially-made navigable database is essential for such functions. Within the navigable database three kinds of information need to be modelled, including geometric data that representing the road network and position of transportation-related features, attribute data that representing various address, traffic conditions and restrictions, and dynamic date describing real-time traffic information. Other Location-Based Services (LBS) fields in ITS (Intelligent Transportation System), such as fleet and asset management, emergency services, telematics, etc., are also depended on such specially-made navigable databases.

Although there already are some commercial navigable database being put into use all over the world, especially in North American, Europe and Japan, there are still lots of research challenges in generating a satisfied navigable database, concerning with data modelling, representation, data processing response to the requirement of ITS applications [Goodchild, 2000]. With the responsibility for providing the fundamental geographic information and related service, the National Geomatics Center of China (NGCC) is now undertaking some projects funded by the State Development Planning Commission of China, State Bureau of Surveying and Mapping of China, as well as some enterprises involved in ITS technology. One of these projects is related to study and establish a navigable database for vehicle navigation. During the implementation of this project, the authors of this paper found that data modelling is the most important comparing with representation, data processing, etc.

---

* Corresponding author.

There are three levels of data modelling for the vehicle navigation systems, i.e., conceptual modelling, logical modelling, and algorithmic modelling. The conceptual model defines the potential contents of such databases (Features, Attributes and Relationships). The logical model specifies how to organize the objects in databases. In the algorithmic model, data structure (such as list or array) is defined according to the characteristics of conceptual and logical models to facilitate the searching, indexing and computing for various particular functions, such as optimum routes finding, turn-by-turn guiding, etc.

Now GDF4 (Geographical Data File) is under approving by ISO, which specifies the conceptual data model and data exchange format for the geographic databases for ITS applications. Some logical data models are also proposed, such as the UNETRANS (Unified Network for Transportation) Data Model by the UNETRANS Consortium in U.S. But for the algorithmic model, generally people think it is just an implementation of the conceptual and logical model. Actually for vehicle navigation systems special consideration should be put onto the algorithmic model, because it represents not only the relationships between the transportation elements but also the real-time and dynamic traffic restrictions in the traffic network. Fig.1 shows the components of a vehicle navigation system. There is a navigable database, which generally stored in a CDROM. A data structure should be established based on the tables in the database to facilitate the computing of the functional modules. When there is an real-time traffic message received, the lists or array in the data structure should be updated, and then a new route or instruction will be computed. So it is very important for the data structure to properly represent the relationships between the transportation elements, as well as the real-time and dynamic traffic restrictions in the traffic network.

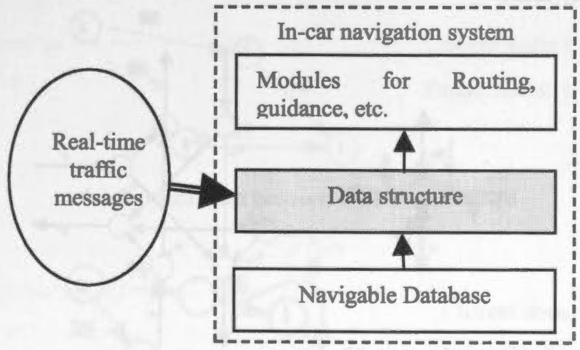

In-car navigation system
Modules for Routing, guidance, etc.
Real-time traffic messages
Data structure
Navigable Database

Figure 1 The components of the in-car navigation system

The authors of this paper studied how to model the turning restrictions in traffic network for vehicle navigation systems. The properties of the turning restrictions are discussed first. The model in conceptual model (GDF 4.0) and logical model (UNETRANS) are introduced. Concerning the algorithmic model, the authors analyzed the disadvantages of the commonly used data structure for representing the turning restrictions, which is based on the nodes relationship of "directed graph". A new link-based data structure was proposed, where a node-link table is used to represent the connectivity of the road network, and a link-link table is used to represent the turning restrictions. And the algorithm for route finding was also modified accordingly. The results of the experiment show that the efficiency of route finding is obviously improved with the new method.

2. TURNING RESTRICTIONS IN TRAFFIC NETWORKS

There are various traffic restrictions in traffic networks, including restricted period of time, vehicle type restriction, signal, turning restrictions, etc. (Fig.2). Among them the turning restrictions contributed significant to the total travel time [Dirch, 1977; Pallotinno, Scutella, 1997]. The magnitude of delays by them may even be comparable to the link travel time which account for 17~35% in the total time [Nielson, Frederiksen and Simonsen., 1998]. Ignoring the turning restrictions in the traffic network modeling may miss essential characteristics of the network and lead to sub-optimal or illogical paths.

In the GDF4 conceptual model a road network is modelled in 3 levels according to different grade of abstracting. Where level 0 describes the geometry of the road network, level 1 (Fig.2a) describes the network in terms of simple features (such as road element and junction), and level 2 (Fig.2b) describes the network with complex features (such as road and intersection). Generally turning restrictions are related to intersections or junctions (Fig.3). For a specific intersection, the turning restrictions are the generalization of the restrictions of related junctions. A kind of relationship called Prohibited Manoeuvre is used to represent the turning restrictions in GDF4.

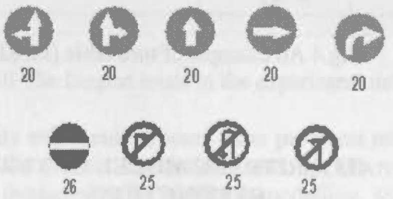


Fig2. Control signs in traffic network

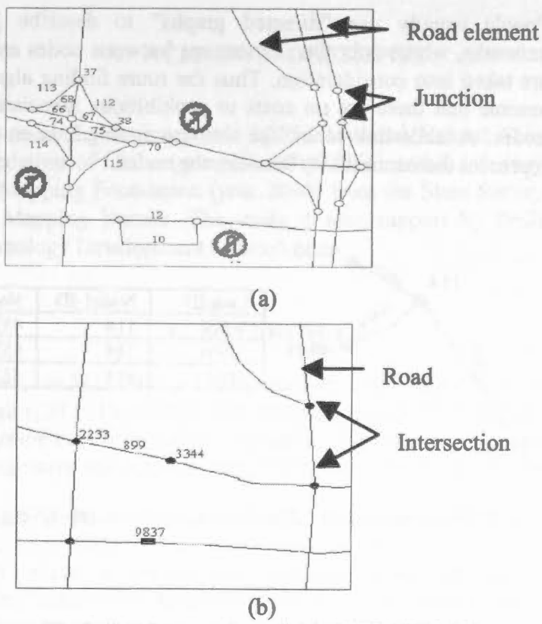


Fig 2 Road network model in GDF4 [GDF4]

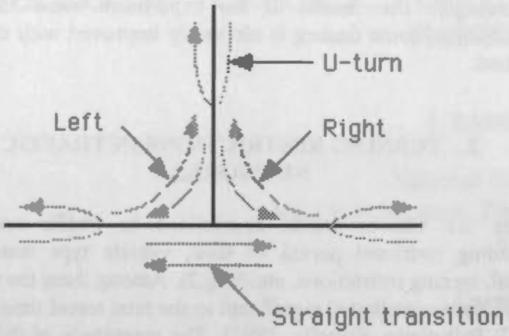
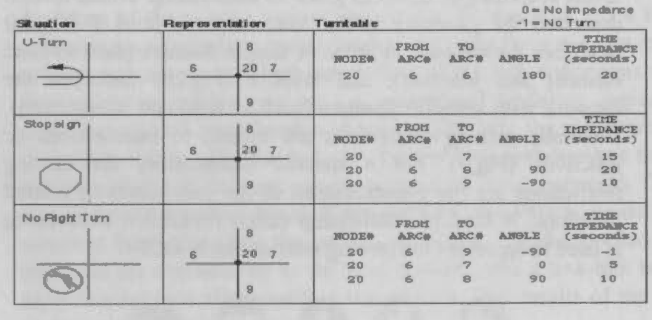


Fig.3 Turn restrictions in a junction/intersection

In the ArcInfo Network model or UNETRANS model, the turn restriction in each node (junction/intersection) is represented with turntables, where every possible turns are listed and maintained. Fig.4 is a example of turntable from ArcInfo.


Situation
Representation
Turntable
U-Turn
Stop sign
No Right Turn
8
6
20
7
9

0 = No Impedance
-1 = No Turn

U-Turn:

NODE#	FROM ARC#	TO ARC#	ANGLE	TIME IMPEDANCE (seconds)
20	6	6	180	20

Stop sign:

NODE#	FROM ARC#	TO ARC#	ANGLE	TIME IMPEDANCE (seconds)
20	6	7	0	15
20	6	8	90	20
20	6	9	-90	10

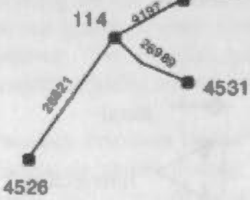
No Right Turn:

NODE#	FROM ARC#	TO ARC#	ANGLE	TIME IMPEDANCE (seconds)
20	6	9	-90	-1
20	6	7	0	5
20	6	8	90	10

Fig.4 An example of turn table [ESRI]

3. ALGORITHMIC MODEL OF TURNING RESTRICTIONS

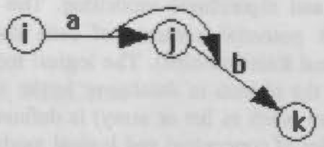
People usually use "directed graphs" to describe general networks, where only the connections between nodes and links are taken into consideration. Thus the route finding algorithms assume that there are no costs or prohibitions associated with nodes. A node-link table like the one in Fig.5 is enough to represent the connectivity between the nodes.



Link-ID	Node1-ID	Node2-ID
26821	114	4526
26989	114	4531
9197	114	4530

Fig.5 node-link table for connectivity describing

A turn represents a transition from one link to another link at a node. It represents relationships between links rather than nodes. Thus the node-link table in Fig.5 is not enough to describe such relationship between links. For example, to describe the relation between link a and link b in Fig.6, two tables are used, one for the turn and the other for the connectivity. At this moment, the classical route finding algorithms based on the nodes has to be modified to adapt to such change.



JUNCTION	From-Edge	To-Edge	Turn Value
j	a	B	TRUE

Turntable

Edge	From-Node	To-Node
A	i	j
B	j	k

node-link table

Fig.6 A node with turn between two links

Some people tried to solve the turning problem by expanding the network, that is to highlight each movement in the intersections, where the costs of the dummy arcs are the banned costs or penalties costs [Kirby and Potts, 1969]. So an intersection is split into many dummy nodes and dummy links that restructure a sub-network, as shown in Fig 7. The advantages of this approach are no turning delays and prohibitions are included in the large network, and the problem can be solved with any general data structure and route finding algorithm. But the shortage of the approach is evident, i.e. the resulting network is considerably larger than the original one. More computational time and computer memory are required and moreover, more time-consuming network updating is involved.

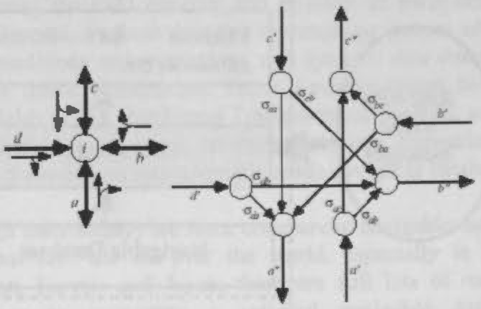
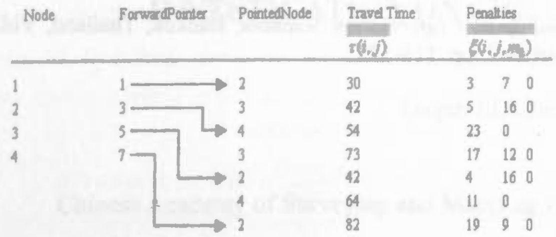


Fig. 7 An expended intersection with turns [Kirby and Potts]

To overcome the disadvantage of the network expending approach, some people tried to solve the turning problem without changing the topology of the network. For example, Ziliaskopoulos and Mahmassani proposed an extended forward star structure (EFSS) to describe the network [Ziliaskopoulos and Mahmassani , 1996], as Fig.8 shows. But when describing the turn with EFSS, the storage space expends rapidly and the efficiency of route finding will be reduced [Han and Jiang, 2001].



Node	ForwardPointer	PointedNode	Travel Time $\tau(i,j)$	Penalties $\xi(i,j,m_k)$
1	1	2	30	3 7 0
2	3	3	42	5 16 0
3	5	4	54	23 0
4	7	3	73	17 12 0
		2	42	4 16 0
		4	64	11 0
		2	82	19 9 0

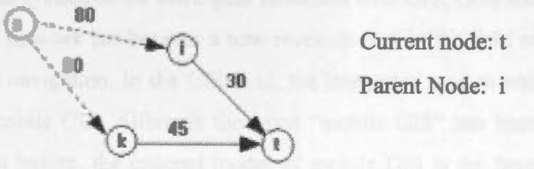
Fig.8 The EFSS including turn representation [Ziliaskopoulos and Mahmassani]

The authors of this paper proposed a new data structure for turning representation. In the data structure, the node-link table is used for connectivity representation and link-link table is used for turn describing, as defined below:

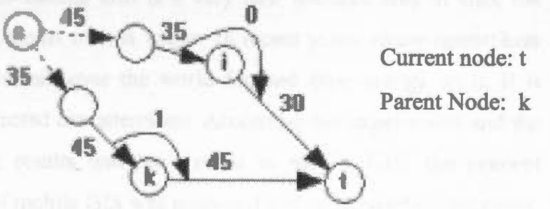
```
Struct NodeLink
{
  double m_dF_NodeID;
  double m_dT_NodeID;
  double m_dLinkID;
  double m_dLength;
};

Struct LinkLink
{
  double m_dF_LinkID;
  double m_dT_LinkID;
  BOOL m_bTurn;
};
```

In classical node-based Dijkstra algorithm for route finding, the selecting of subsequent node depends only on the current node itself, as shown in Fig.9a. With turning restrictions, the subsequent node selecting is related to the parent node of the current node (Fig.9b).



(a) Only distance between node is considered



(b) Turning restriction is considered

Fig.9 Selecting of the subsequent node

To solve this problem, the authors modified the classical node-based Dijkstra algorithm to be link-based Dijkstra algorithm. According to analysis, the time consumption of the new data structure and algorithm is O(nlogn) compared with the classical O().

4. EXPERIMENTS AND CONCLUSIONS

The proposed method is tested in a large-scale road network of Hong Kong ,with 5351 nodes, 7117 links and 43575 turn restrictions. To find the longest route (which consists of 174 links) in the network, 4323 links are searched within no more than 0.2 seconds (Fig.10). Whereas more than 2 minutes are used with the classical method. The approach will be more efficient if some strategies of data organizing, such as dividing data sections according to geographic coverage, defining data layers according to road classes, and setting searching range according to the positions of the start node and destination node.

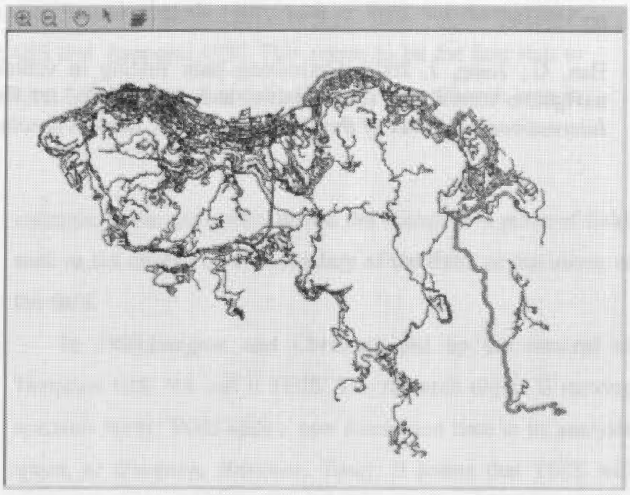


Fig.10 The longest route in the experiment network

Further study will focus on some other problems related to data modelling for vehicle navigation, such as 3D road network modelling, dynamic traffic condition modelling, as well as the efficient data organization methods.

5. ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (grant No. 40171076) and the Surveying and Mapping Foundation (year 2000) from the State Surveying and Mapping Bureau. The study is also support by Brilliant Technology Development Limited.

6. REFERENCES

Waters, N.M., 1999, Transportation GIS: GIS-T, in P.A. Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind (Eds.), Geographical Information Systems: Principles, Techniques, management and Applications, Wiley: New York, pp. 827-844.

Goodchild, M.F., 2000, GIS and Transportation: Status and Challenges, GeoInformatica,4(2),pp.127-139.

Dirck V.V, 1977, Improved Shortest Path Algorithms for Transport Networks, Transpn Res., 12, pp.7-20

Pallotinno S., Scutella M.G, 1997, Shortest Path Algorithms in Transportation models: Classical and Innovative Aspects, Technical Report's-97-06, University of Perugia, Institute of Electronics, via G. Duranti 1/A1, 06131 Perugia, Italy

Nielsen O.A, Frederiksen R.D and Simonsen N., 1998, Using Expert System Rules to Establish Data for Intersection and Turns in Road Networks, Int.Trans.Opl Res. 5(6), pp.569-581

Kirby R.F. and Potts R.B., 1969, The minimum route problem for networks with turn penalties and prohibitions, Transpn Res.3, pp. 397-408

Ziliaskopoulos A.K. and Mahmassani H.S, 1996, A Note on Least Time Path Computation Considering Delays and Prohibition for Intersection Movements, Transpn Res.-B,.30(5), pp.359-367

Han, G., Jiang, J, 2001, Optimizing path finding in vehicle navigation considering turn penalties and prohibitions, In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Bankok, Thailand, Vol. 34, part 2w2, pp. 118-122

INVESTIGATION ON THE CONCEPT MODEL OF MOBILE GIS

Luqun LI, Chengming LI, Zongjian LIN

Chinese Academy of Surveying and Mapping 100039, No.16, Road Beitaiping, District Haidian, Beijing, P.R.China,Email: liluqun@263.net Tel:.0086-10-68134229

Key Words: Moving Object, Mobile GIS, TGIS,Concept

Abstract:

Today, we hear more Mobile, and find some mobile GIS application in our everyday life, such as GPS car management However, what is mobile GIS? What is the difference among Static GIS and Temporal GIS? This seems to be the first step to study mobile. The paper, it gives the concept of mobile GIS, the system Architecture of Mobile GIS, the key research area in mobile.

1 Introduction

With the development of computer software and hardware, as well as the application of WAP successfully in wireless network, and the emergence of intelligent terminals, such as PDA and WAP phone etc., People owing these intelligent terminals can do their work on the spot at any time or anywhere, which they have to do in their office or home before. That is to say, they can "do their work on the move". In the meantime, apparatus such as GPS, GSM model made for the intelligent terminals have extended the application region of it. The integrated system of the intelligent terminals with GIS, GPS and wireless network has become a new research area in the field of GIS and navigation. In the GIS field, the integrated system was called mobile GIS. Although the word "mobile GIS" has been proposed before, the concept model of mobile GIS is far from being defined clearly.

Since mobile GIS is a very new research area in GIS, the concept model of it is vague. In recent years, many researchers distributed all over the world focused their energy on it. It is also attracted our attentions. According our experiences and the research results that have made in mobile GIS, the concept model of mobile GIS was proposed and introduced in this paper.

2 What is Mobile GIS

Traditional GIS mainly does research work on static spacious entity. We call it Static GIS. The analysis space of Static GIS is *(Postion, Attribute)*.The research object is spacious entity. As for static GIS does not involve, it can not depict the character of the nature of the moving world, for example, Static can hardly record the change of a piece of field, such as the change of the boundary of the field, or the owner of the field.

In 1988,Langran and Chrisman put up the concept of Temporal GIS. We call it TGIS. The research object is moving spacious entity. TGIS adds a new dimension time in its analysis space, or *(Position, Attribute, Time)*. It seems that TGIS will solve the dynamic state of the spacious entity. Though many work have been done on TGIS, right now, the data model of TGIS is still under research..

Both SGIS and TGIS do research work on spacious entity. In GIS spacious entity refers to an entity which has geographic aspects, such as road, mountain, building, etc, in the meantime spacious entity does not include no- geographic entity, such as a car, a desk, a book and so on.

Today, we hear more about mobile GIS. What is mobile GIS? What is the research object of mobile GIS? Here, we give the concept of Mobile GIS:

Mobile GIS is a kind of GIS, it does main research work on no-geographic moving object in geographic space, it do research work on the relationship between moving object and geographic entity, or moving object between another moving object. For an example, we can integrate GIS, GPS, wireless internet to build a mobile GIS to monitor cars. We study the moving car in geographic entity space, moving car is no-geographic entity.

3 General System Architecture of Mobile GIS

A mobile GIS is composite of a mobile client, a server, a wireless network, a mobile client position record system, such as GPS.

A mobile client can be a moving car equipment with GPS and GSM model which can get and send the geographic position to the server by SMS, or a man takes PDA, such as Palm or windows CE which equipped with GPS, the PDA can show the digital map and can comminute with the back support server by wireless network.

A wireless network can be GSM, CDMA, CDPD or GPRS, which support digital data transmitting.(See the figure)

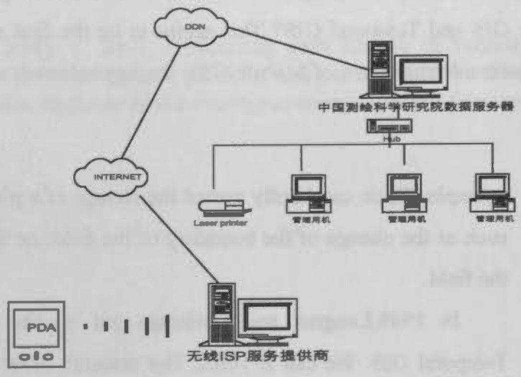


Figure 1 A typical system artichture of a mobile GIS

4 Key research area in Mobile GIS

As for mobile GIS is new research area, the following area will be mobile key research Devices that combine a hand-held computer with a GPS receiver, a cellular phone, and a digital camera will enable users to integrate spatial analysis into their daily lives, opening GIS to the mass market. We currently lack the theory necessary to tackle in real time very large volumes of spatio-temporal data in a highly-distributed computing environment. This research topic covers cognitive aspects about the interaction with Mobile GISs; spatio-temporal information in a highly distributed computing environment; and the integration and interoperation of multi-modal spatial information (such as voice, graphics, image, and video).

4.1 The data model for mobile client side in Mobile GIS

The live world is very complicated, and the people hope that the GIS include enough data to meet their demands, at the same time, they also always want to discard the data, which they do not care about (Freksa and Barkowsky, 1996).

The intelligent terminals that are used for mobile GIS hardware platform are with lower CPU speed than PC, limited externals storage and small size display screen. In general, it's hard to display map on them if we just take the normal method that we use in personal computer or workstation. Concerning about the characteristics of these intelligent terminals, a different or new line compared with traditional GIS must be taken on mobile GIS.

4.2 The mobile computing model for Mobile GIS

As for mobile client usually has limited CPU speed, many computer works such as query some place should be place to the server side. How to co-operate to finish the distributed work in wireless network environment will be a challenge in this area.

4.3 The database structure to store moving object data

To store the moving data of a moving object will cost many space in computer, today there are is not a suitable database to store moving and changeable data. The Mobile Object Database will be a research in this area.

4.4 The realization of query moving object and analysis moving object

Query and analysis are the basic function of GIS, mobile GIS's query of deal with finding the moving object, or track a moving object etc.

5 Conclusion

As for Mobile GIS is a new subject in GIS, this paper presents the basic of mobile GIS, however this is the first step in studying Mobile GIS. Further, I will present the design of mobile GIS data model

References:

[1]. Xiao leben.The study on GIS model. Journal of Wuhan University.2001.3

[2]. Freksa,C. and Barkowsky, T.On the Relations between Spatial Concepts and Geographic Objects, IN: Peter A..Burrough and Andrew U.Frank, Geographic Objects with Indeterminate Boundaries, Taylor &Francis, 1996

[3]. Gore A.The Digital Earth: Understanding Our Planet in the 21th Century [R]. The Lecture Note on the Science Center of California，1998

[4]. 2 H. Samet, Neighbor finding in images represented by octrees ,Computer Vision[J] ,Graphics, and Image processing, 1989, 46, 367~386

[5]. http://www.savaje.com 2001.7

[6]. http://www.spatial.maine.edu/~max/UCGIS.5.html 2002.1.10

RANDOM TOPOLOGICAL RELATIONS BETWEEN LINE AND AREA IN GIS

Deng Min [a*] **Li Chengming** [b] **Chen Xiaoyong** [c]

[a] Department of Geomatics, Institute of Remote Sensing Information and Engineering,
Wuhan University, 430079, No.229, Road Luoyu, District WuChang, Wuhan, P.R.China
dengmin208@263.net

[b] Chinese Academy of Surveying and Mapping, 100039,
No.16, Road Beitaiping, District Haidian, Beijing, P.R.China
cmli@casm.ac.cn

[c] Geoinformatics, School of Advanced Technology, Asian Institute of Technology
PO BOX 4, Klong Luang, Pathumthani 12120, THAILAND
xychen@ait.ac.th

KEY WORDS: Data quality, Topological relations, Uncertainty

ABSTRACT:

There is inevitably error or uncertainty in spatial data, which are used to represent reality world, it will causes a suspicion for correctness of topological relations obtained by reasoning observation data. In this paper, we firstly describe the geometric and statistical properties of random line and area, and analyzing the effect of error in spatial data on representation of topological relations between spatial entities. Then, we present the concept of uncertain topological relations sets in order to represent random topological relations between spatial data with random errors. Finally, a new approach to determining random topological relations is presented.

1. INTRODUCTION

There is inevitably error or uncertainty in spatial data, which are used to represent reality world [Blakemore, 1984; Chrisman, 1982; Guptill, 1995], it causes a suspicion for correctness of topological relations obtained by reasoning observation data [Winter, 2000; Worboys, 1995]. While most of spatial queries in GIS applications are based on topological relations between two objects, so it is obvious that a wrong answer, even a mistaken decision will be made if description of topological relations is not correct. Topological relation, as one of the most fundamental property of spatial objects, has been investigated in GIS recent years. Most of researches are based on crisp set and involve common topology such as algebraic and point-set topology [Egenhofer and Franzosa, 1991; Egenhofer and Herring, 1991; Egenhofer and Mark, 1994]. How to formalize the topological relations between two random objects is still an open question.

In this paper, we consider random error of spatial data, and assume that positional error of a node conform to a standard normal distribution, then to derive and compute positional error of any point on a line segment is based on stochastic process [Shi and Liu, 2000]. Furthermore, it is well-known that an areal object is composed of a sequence of ordered line segments, here, an areal object is regarded as a random field, and positional error of any point on its boundary line can be derived similarly. On the basis of this, we will analyze the effect of random error on description of topological relations between line and areal object in following three aspects: ① classes of topological relations; ② changes of graphical structure; ③ Randomness of value of element in 9-intersection model. Sequentially, we focus on how to describe random topological relation, defining the concept of uncertain topological relations sets. Finally, this paper ends with an approach of determining uncertain topological relations by means of relative probability as an indicator.

2. GEOMETRIC DESCRIPTIONS OF RANDOM LINE AND AREA

2.1 Definition of Random Line and Area

A mass of line entity existed in reality world such as road, river, administrative boundary and so on, are all expressed with line entity in GIS. These entities are composed of a sequence of ordered line segment, which is defined by two endpoints. The error of point coordinates acquired by measuring, scanning, digitalization etc., will lead to location uncertainty of line entity. Here, it is called random line. Let L_{12} a random line segment in 2-D space, is denoted by its two end points Z_1 and Z_2. Let $Z_1=(x_1,y_1)$, $Z_2=(x_2,y_2)$, defining following operation rule: 1) $tZ_i=(tx_i,ty_i)$ 2) $Z_1 \pm Z_2=(x_1 \pm x_2, y_1 \pm y_2)$, then random line segment can be expressed as

$$L_{12}=\{Z_1+t(Z_2-Z_1),\ 0 \le t \le 1\} \qquad (1)$$

Obviously, L_{12} is a connective close set, or denoted as a close interval $[Z_1, Z_2]$. Furthermore, the line segment not including the end points is denoted as an open interval $(\mathrm{Z}_1, \mathrm{Z}_2)$. So the random line element $L_{1,2,\cdots,n}$ composed of n random points $\{\mathrm{Z}_i, 1 \le i \le n\}$, are expressed as a union set, *i.e.*,

$$L_{1,2,\cdots,n}=\bigcup_{i=1}^{n-1}[\mathrm{Z}_i, \mathrm{Z}_{i+1}] \qquad (2)$$

any integer $i\,(1 \le i \le n-1)$, the relation between random line element and its vertex satisfy the following formula

$$L_{1,2,\cdots,i} \cap (\mathrm{Z}_i, \mathrm{Z}_{i+1})=\phi \qquad (3)$$

Here, if random line point $\mathrm{Z}_1 \ne \mathrm{Z}_n$, then $L_{1,2,\cdots,n}$ is called simple random line. If having $\mathrm{Z}_1=\mathrm{Z}_n$, it is a close polygon in geometry. Having $L_{1,2,\cdots,i} \cap (\mathrm{Z}_i, \mathrm{Z}_{i+1}) \ne \phi$, then $L_{1,2,\cdots,n}$ is a self- intersection random line. In this paper, we only discuss simple random line and area.

2.2 Geometric Constraints of Random Line

For any digital line stored in GIS after data procession and compression procedures, the data used to express the location of line can be assumed to satisfy the following conditions:

1) The distance between any two points is far larger than upper limit of the maximum error, namely,

$$Z_iZ_j >> \delta \quad (1 \le i,j \le n-1)$$

2) The distance of any point to any segment, *d*, is far larger than the maximum error up limit, namely,

$$d(Z_i, Z_{i-1}Z_{i+1}) >> \delta$$

2.3 Statistical Description of Random Line

We supposed that: (1) the coordinates vectors $\mathrm{Z}_i=(\mathrm{X}_i\ \mathrm{Y}_i)^{\mathrm{T}}$ and $\mathrm{Z}_{i+1}=(\mathrm{X}_{i+1}\ \mathrm{Y}_{i+1})^{\mathrm{T}}$ are independent and the same accuracy; (2) random point Z_i ($i=1,2,\cdots,\mathrm{n}-1$) complies with the two dimensional normal distribution

$$Z_i \sim N_2(\mu_{Z_i}, \textstyle\sum_{Z_iZ_i})\,(i=1,2,\cdots,\mathrm{n}-1) \qquad (4)$$

Here, $\mu_{Z_i}=(\mu_{X_i}\ \mu_{Y_i})^T$ is mathematical expectation of any random point coordinate vector $\mathrm{Z}_i=(\mathrm{X}_i\ \mathrm{Y}_i)^{\mathrm{T}}$ $(i=0,1)$ and $\sum_{Z_iZ_i}$ is its covariance matrix, having

$$\sum\nolimits_{Z_iZ_i}=\begin{bmatrix}\sigma_{X_i}{}^2 & 0 \\ 0 & \sigma_{Y_i}{}^2\end{bmatrix} \quad (i=0,1) \qquad (5)$$

For any point in random line, its coordinate components can be expressed as

$$\begin{cases} X_i(t_i)=(1-t_i)X_i+t_iX_{i+1} \\ Y_i(t_i)=(1-t_i)Y_i+t_iY_{i+1} \end{cases} \quad (0 \le t_i \le 1) \qquad (6)$$

It means that $X_i(t_i)$ and $Y_i(t_i)$ is a linear combination of normal random variables X_i, X_{i+1} Y_i and Y_{i+1} separately, they hereby satisfy the normal distribution too. Hence, the random line is composed of infinitely normal random variables $X(t)$ and $Y(t)$, where $\{t_{\mathrm{i}}, 1 \le \mathrm{i} \le \mathrm{n}-1\}$ and $t_i \in [0,1]$.

3. IMPACT OF UNCERTAINTY ON TOPOLOGICAL RELATION BETWEEN LINE AND AREA

3.1 Invariant of Topological Relations Types

Base on 9-intersection model, topological relations between line and area without error or uncertainty in 2-D space can be distinguished into 19 types. For random line and random area, though data describing their spatial position have inevitably error or uncertainty, the topological relation types separable is still 19 kinds based on the 9-intersection model.

3.2 Change of Graphic Structure

Definition 1. For any random line, all its vertex should satisfy Degree(N) ≥ 1, in which Degree(N) denotes the connective degree related with vertex N. If having Degree (N)=1, the vertex N is boundary point of random line too.

Definition 2. If there is a chain between two nodes, we call the two nodes connective. Furthermore, if all pairs of nodes in a planar graph are connective, the graph is connective.

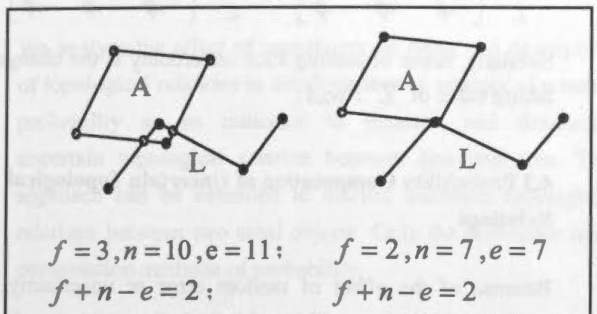
Any graph in the plane, *G*, is composed of node, edge and face, and the numbers of these elements satisfy the following formula

$$f+n-e=c+1 \qquad (7)$$

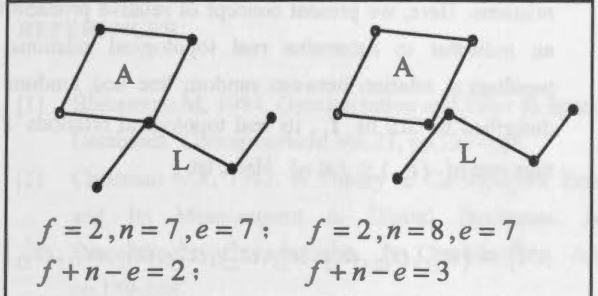
Where *f*, *n*, *e* are the numbers of the face, node and edge, and *c* is the number of connective branches of *G*. If the planar graph is connective, *i.e.*, *c*=1, the formula (7) is simplified as

$$f+n-e=2 \qquad (8)$$

The expression is the famous Eula formula, which is often used for check of topological inconsistency. Below we apply it to analysis changes of graphic structure under the effect of uncertainty. In figure 1, because of the positional error, the graphic structure composed of the random line *L* and random area A such as (a) and (b) is different. In figure 1(a), only numbers of face, node and edge are differs, while keeping the connectivity. But in figure 1(b), not only the numbers of graphic components differ, but the connectivity also changes.



(a) Graphic structure changes, its connectivity not changes



(b) Graphic structure changes, its connectivity does

Figure 1 The effect of uncertainty on graphic structure

3.3 Uncertainty of Value of Element in 9-intersection Model

Topological relations between random line (L) and random area (A) are described based on the 9-intersection model, which is defined by intersection of L's interior (L^0), boundary (∂L) and exterior (L^-) and A's interior (A^0), boundary (∂A) and exterior (A^-), *i.e.*,

$$\Im_9(L,A)=\begin{bmatrix} L^0\cap A^0 & L^0\cap\partial A & L^0\cap A^- \\ \partial L\cap A^0 & \partial A\cap\partial A & \partial A\cap A^- \\ L^-\cap A^0 & L^-\cap\partial A & L^-\cap A^- \end{bmatrix} \tag{9}$$

According to point-set topology, the topological components of a random line L_i are defined as

$$\partial L=\{b,e\};\ L^0=L/\{b,e\};L^-=R^2-L \tag{10}$$

Their geometric meanings are illustrated in figure 2.

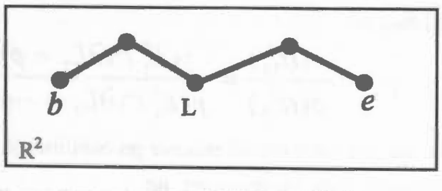


Figure 2 Topological definition of 2-D random line

We may see from the figure, that the two endpoints of random line L are its topological boundary, and the other vertex and all edges are its interior. But the measuring points described in GIS are disaccording with their real location, that is, ∂L_i, L_i^0 and L_i^- in formula (10) contain random error.Here, we define them random set. In this paper, probability is used to represent error or uncertainty of element in random set. The followings is its definition

$$p\{s\in A\}=\max\{H(s,a),a\in A\} \tag{11}$$

Where, s is a point element, A a random set, $p\{s\in A\}$ is the probability of s belonging to random set A, $H(s,a)$ is probability function of determining whether s and a are the same point. Furthermore, the probability that s does not belong to random set A is expressed as

$$p\{s\notin A\}=1-\max\{H(s,a),a\in A\} \tag{12}$$

For any point $a(x_a,y_a)$ in random line L, its probability density functon is defined as

$$f(x,y)=\tfrac{1}{2\pi\sigma^2}\exp[-\tfrac{(x-x_a)^2+(y-y_a)^2}{2\sigma^2}] \tag{13}$$

Here, parameter σ is the standard error of $a(x_a,y_a)$, so the probability that any point falls in the equal density error circle is

$$\begin{aligned} p(d) &= \iint_{C_d}\tfrac{1}{2\pi\sigma^2}\exp[-\tfrac{(x-x_a)^2+(y-y_b)^2}{2\sigma^2}]dxdy \\ &= \int_0^d\tfrac{\rho}{\sigma^2}\exp(-\tfrac{\rho}{2\sigma^2})d\rho \\ &= 1-\exp(-\tfrac{d^2}{2\sigma^2}) \end{aligned} \tag{14}$$

Where, C_d: $(x-x_a)^2+(y-y_b)^2\le d^2$, d the radiu of error circle, So we define $H(s,a)$ as

$$\begin{aligned} H(s,a) &= 1-p(d) \\ &= \exp(-\tfrac{d^2}{2\sigma^2}) \end{aligned} \tag{15}$$

here, d is the distance between the point elements $s(x_s,y_s)$ and $a(x_a,y_a)$

So, If uncertainty will be considered, then values of elements in the 9-intersection model are extended as $\phi/p_{ij}(\phi)$ or $\neg\phi/p_{ij}(\neg\phi)$, and satisfy the expression

$$p_{ij}(\phi)=1-p_{ij}(\neg\phi) \tag{16}$$

Where, $i,j(1\le i,j\le 3)$ is the number of row and collum in the 9-intersection model.

4. DETERMINING TOPOLOGICAL RELATIONS BETWEEN LINE AND AREA

4.1 Uncertain Topological Relation Set

Because there is random error in spatial data, the topological relation ($\hat{t}$) derived by spatial data in GIS sometimes does not coincide with real topological relation (t). But sometimes it is difficult to determine the real topological relation,so the uncertain topological relations can be expressed with confidence probability of t to $\hat{t}$.In addition, topological relations based on the 9-intersection model is a definite classification. So the real topological relation t will only is in the neighbourhood of $\hat{t}$. Here, we define uncertain topological relations set as

$$p(t \mid \hat{t}_i) = \begin{cases} p_{i1}, & t = \hat{t}_1 \\ p_{i2}, & t = \hat{t}_2 \\ \vdots & \\ p_{in}, & t = \hat{t}_n \end{cases} \tag{17}$$

Where, $p_{ij}(1 \le \mathrm{j} \le \mathrm{n})$ is the probability when $t = \hat{t}_j$, while any topological relations $\hat{t}_j$ is a relation existed with some probability $p_{ij}(p_{ij} > 0)$. Equivalently, formula (17) is expressed as

$$t_i = p_{i1} / \hat{t}_1 + p_{i2} / \hat{t}_2 + \cdots + p_{in} / \hat{t}_n \tag{18}$$

4.2 Description of Uncertain Topological Relations

As above mentioned, value of element in the 9-intersection model sometimes is uncertain, while the probability of taking some value is certain. Therefore, we can extend the 9-intersection model to describe such uncertain topological relations, *i.e.*,

$$\Im_9(L,A) = \begin{bmatrix} p_{11}(L^0 \cap A^0) & p_{12}(L^0 \cap \partial A) & p_{13}(L^0 \cap A^-) \\ p_{21}(\partial L \cap A^0) & p_{22}(\partial L \cap \partial A) & p_{23}(\partial L \cap A^-) \\ p_{31}(L^- \cap A^0) & p_{32}(L^- \cap \partial A) & p_{33}(L^- \cap A^-) \end{bmatrix} \tag{19}$$

As showed in figure 1(a), the two kinds of topological relations between *L* and *A* are described as

a1) Case 1

$$\begin{array}{c|ccc} & A^0 & \partial A & A^- \\ \hline L^0 & \boxed{\neg\phi} & \neg\phi & \neg\phi \\ \partial L & \phi & \phi & \neg\phi \\ L^- & \neg\phi & \neg\phi & \neg\phi \end{array}$$

a2) Case 2

$$\begin{array}{c|ccc} & A^0 & \partial A & A^- \\ \hline L^0 & \boxed{\phi} & \neg\phi & \neg\phi \\ \partial L & \phi & \phi & \neg\phi \\ L^- & \neg\phi & \neg\phi & \neg\phi \end{array}$$

Obviously, it is the change of taking value of $L^0 \cap A^0$ that causes the change of topological relation between *L* and *A* in Fig1 (a). For the same reason, two kinds of topological relations in Fig1 (b) can be described as:

b1) Case 1

$$\begin{array}{c|ccc} & A^0 & \partial A & A^- \\ \hline L^0 & \phi & \boxed{\neg\phi} & \neg\phi \\ \partial L & \phi & \phi & \neg\phi \\ L^- & \neg\phi & \neg\phi & \neg\phi \end{array}$$

b2) Case 2

$$\begin{array}{c|ccc} & A^0 & \partial A & A^- \\ \hline L^0 & \phi & \boxed{\phi} & \neg\phi \\ \partial L & \phi & \phi & \neg\phi \\ L^- & \neg\phi & \neg\phi & \neg\phi \end{array}$$

Similarly, cause of leading such uncertainty is the changes of taking value of $L^0 \cap \partial A$.

4.3 Probability Computation of Uncertain Topological Relations

Because of the effect of random error or uncertainty, real topological relations will be a kind of all possible topological relations. Here, we present concept of relative probability as an indicator to determine real topological relations. Let topological relation between random line and random area described in GIS be $\hat{t}_i$, its real topological relations t may take one of $\{\hat{t}_j, 1 \le j \le n\}$. Here, let

$$\hat{t}_i = \{\alpha_{11}^i, \alpha_{12}^i, \alpha_{13}^i, \alpha_{21}^i, \alpha_{22}^i, \alpha_{23}^i, \alpha_{31}^i, \alpha_{32}^i, \alpha_{33}^i\}$$

$$\hat{t}_j = \{\alpha_{11}^j, \alpha_{12}^j, \alpha_{13}^j, \alpha_{21}^j, \alpha_{22}^j, \alpha_{23}^j, \alpha_{31}^j, \alpha_{32}^j, \alpha_{33}^j\}$$

Where, α_{kl}^i、$\alpha_{kl}^j (1 \le k, l \le 3)$ are vales of elements in the 9-intersection model, and the index *i*, *j* are separately row and column of corresponding 9-intersection model. Then, we define

$$p_{ij} = \frac{\prod_{k,l=1}^{k,l=3} p(\alpha_{kl}^j)}{\prod_{k,l=1}^{k,l=3} p(\alpha_{kl}^i)} \tag{20}$$

p_{ij} is called relative probability of topological relation in this paper. It satisfies the following properties: (1) $0 < p_{ij} < +\infty$; (2) $p_{ij} = (p_{ji})^{-1}$; (3) $p_{ij} = 1$, if and only if $j = i$.

Here, we still take figure 1 as an example. After analyzing figure 1(a), the formula (20) can be simplified as

$$p_{ij} = \frac{p(\alpha_{11}^j)}{p(\alpha_{11}^i)} = \frac{p(L_1^0 \cap L_2^0 = \neg\phi)}{p(L_1^0 \cap L_2^0 = \phi)}$$

After analyzing figure 1(b), the formula (20) can be simplified as

$$p_{ij} = \frac{p(\alpha_{12}^j)}{p(\alpha_{12}^i)} = \frac{p(L_1^0 \cap \partial L_2 = \phi)}{p(L_1^0 \cap \partial L_2 = \neg\phi)}$$

Thus, we can compute all relative probabilities in formula (17) by means of (20). Furthermore, let

$$p_{ic} = \max\{p_{i1}, p_{i2}, \cdots, p_{in}\} \tag{21}$$

Therefore, We can determine that the real topological relation between random line and random area is $\hat{t}_c$.

5. CONCLUSIONS

We analyze the effect of uncertainty on types and description of topological relations in detail, present to concept of relative probability as an indicator to measure and determine uncertain topological relation between line and area. The approach can be extended to discuss uncertain topological relations between two areal objects. Only the difference is in computation methods of probability.

REFERENCES

[1] Blakemore M, 1984. Generalization and Error in Spatial Databases. *Cartographica,* Vol.21, pp.131-139

[2] Chrisman N.R, 1982. A Theory of Cartographic Error and Its Measurement in Digital Databases. *In: Proceedings of Auto-Carto 5*. Crystal City, VA, pp.159-168.

[3] Guptill S.C and Morrison J.L, 2000. Elements of Spatial Data Quality. Oxford: Elsevier Scientific, 1995.

[4] Winter S. Uncertain topological relations between imprecision regions. *International Journal of Geographical Information Systems,* Vol.14 (5), pp.411-430.

[5] Worboys M, 1995. GIS, a computing perspective. London: Taylor & Francis.

[6] Egenhofer M and Franzosa R, 1991. Point-set topological spatial relations. *International Journal of Geographical Information Systems,* Vol. 5(2), pp.161-174

[7] Egenhofer M and Herring J, 1991. *Categorizing binary topological relationships between regions, lines and points in geographic databases.* In: A Framework for the Definition of Topological Relationships and an Approach to Spatial Reasoning within this Framework, edited by Egenhofer M, Herring J, et.al. Santa Barbara, CA, pp.1-28.

[8] Egenhofer M and Mark D, 1995. Modelling conceptual neighborhoods of topological line-region relations. *International Journal of Geographical Information Systems,* Vol. 9 (5), pp.555-565

[9] Shi W and Liu W, 2000. A Stochastic Process-Based Model for Positional Error of Line Segments in GIS. *International Journal of Geographic Information Sciences,* Vol.14 (1), pp.51-66.

GIS APPLICATIONS USING AGENT TECHNOLOGY

N. Shahriari [a], C. V. Tao [b]
[a] Dept. of Geomatics Engineering, University of Calgary, 2500 University Dr. NW, Calgary, AB, Canada T2N 1N4 - nshahria@ucalgary.ca
**Dept. of Earth and Atmospheric Science, York University, 4700 Keele Street, Toronto, ON, Canada M3J 1P3 - tao@yorku.ca

KEY WORDS: Agent, Geographic Information System, Intelligence, Geospatial Data.

ABSTRACT:

There is an increasing interest in agent-oriented technology, spanning applications as diverse as information retrieval, user interface design and network management. Application of intelligent agents in the GIS environment is actively being explored. Many different types of agents are being employed to improve usability of GIS software as well as users access to geospatial data and services available through Internet. Geospatial information retrieval and filter, intelligent geospatial search engine, knowledge discovery, decision model assessment and optimization are typical applications based on using agents in GIS.
This paper presents a short introduction to agent and the main research areas in agent technology. The heart of the paper is a wide overview of applications of intelligent agents in GIS. These applications have been clearly classified in different categories. The paper concludes with an example of a web search tool, Special Spatial Search (SSS). SSS is developed to assist users in geospatial related searching over the web.

1. INTRODUCTION

Over the past two decades, GISs have been adopted widely in support of planning, agriculture, forestry, infrastructure maintenance, and many other fields. Distributed Geospatial Information (DGI) is the widespread distribution of geospatial information in a variety of forms, including maps, images, data sets, analysis operations, and reports. Distributed Geospatial Information Systems (DGIS), are the systems that provide DGI through Internet.

Nowadays, the Internet has become a huge source of Geospatial information and offers already many different services. The advantages of using such technology are obvious to most users of geodata and geoprocessing resources. However, effective use of information in web is becoming increasingly difficult because of the sheer size of the web and its diversity of resource, thus intelligent software technology for DGIS needs to be developed. Moreover the development of GIS technology has made it available to a growing number of people from different disciplines and with different backgrounds. However, the degree of productivity they can achieve is limited by their lack of technical knowledge about GIS tools, which are becoming larger packages with more functionalities. A better user interaction is a key issue to get a broader user acceptance of GIS. The user interface is probably the most important aspects of GIS usability. Therefore GIS software intelligent interface could be an ideal tool to help the users of GIS software in this quickly changing environment.

Application of intelligent agents in the GIS environment is actively being explored and some projects have been reported in the literature where different types of agents are being employed to improve usability of GIS software as well as users access to geospatial data and services available through Internet. Geospatial information retrieval and filter, intelligent geospatial search engine, knowledge discovery, software intelligent user interface, decision model assessment and optimization are typical applications based on using agents in GIS.

In the following section, we present a short introduction about agent and the main research areas in agent technology. In section 3, some of the projects and researches regarding the application of intelligent agents in the GIS environment will be overviewed. Section 4 describes the developed web search tool, Special Spatial Search (SSS). SSS is developed using HTML and Java Scripting to assist users in geospatial related searching over the web. The final section offers some concluding remarks.

2. AGENT TECHNOLOGY

Intelligent Agents are one of the "hot" topics in Information Systems R&D at the moment. The last ten years have seen a marked interest in agent-oriented technology, spanning applications as diverse as information retrieval, user interface design and network management.

An agent has been defined as a computational entity which:
- acts on behalf of other entities in an autonomous fashion;
- performs its actions with some level of proactivity and/or reactiveness;
- exhibits some level of the key attributes of learning, co-operation and mobility.

In the above definition of agent,
- "autonomy" means operating without the direct intervention of humans or others;
- "reactivity" means reacting on events in the environment;
- "proactivity" means exhibiting goal-directed behavior by taking the initiative.

2.1 Software Agents

Software agents (often simply termed agents) are software systems that inhabit computers and networks and assist users with computer-based tasks. Why the need for software agents is becoming so urgent? The computer is merely a passive entity waiting to execute specific, highly detailed instructions; it provides little help for complex tasks or for carrying out actions that may consume a large proportion of the user's time (e.g., searching for information). Researchers and software companies have set high hopes on software agents, which know users' interests and can act autonomously on their behalf. Instead of exercising complete control, people will be engaged in a cooperative process in which both human and computer agents initiate communications, monitor events and perform tasks to meet users' goals. In essence, we need software agents because (Green, et. al, 1997):

- more and more everyday tasks are computer-based;
- the world is in a midst of an information revolution, resulting in vast amounts of dynamic and unstructured information;
- increasingly more users are untrained;
- and therefore users require agents to assist them in order to understand the technically complex world.

2.2 Main Research Areas

The researches in agent technology have been focused on three research areas, namely, Intelligent User Interfaces, Distributed Agent Technology and finally, Mobile Agent Technology. The following three subsections will briefly summarize the main aspects of these areas in turn.

2.2.1 Intelligent User Interfaces: Intelligent User Interfaces (IUI) attempt to maximize the productivity of the current user's interaction with the system. Intelligent (User) Interface Agents are fairly recent developments that use an agent-oriented approach to the construction of such systems. The major factors that distinguish Interface Agents from any other IUI is the fact that agents are proactive and enjoy a degree of autonomy. Generally the existing interface agents can be classified in three categories, namely, information filtering agents, information retrieval agents and personal digital assistants.

2.2.2 Distributed Agent Technology: Distributed Artificial Intelligence is a sub-field of Artificial Intelligence which is concerned with a society of problem solvers or agents interacting in order to solve a common problem. Such a society is termed a Multi-Agent System (MAS), namely, a network of problem solvers that work together to solve problems that are beyond their individual capabilities. Multi agent systems have the following advantages (Purvis, et. al, 2000):

- Using a collection of problem-solvers makes it easier to employ divide-and-conquer strategies in order to solve complex, distributed problems. Each agent only needs to possess the capabilities and resources to solve an individual, local problem.
- The mental image of autonomous, human-like agents facilitates the mapping of real-world problems into a computational domain.
- The idea of several agents cooperating to solve a problem that none could solve individually is a powerful metaphor for thinking about various ways that individual elements can be combined to solve complex problems.

2.2.3 Mobile Agents: Mobile agents simply are programs, which may autonomously migrate themselves from host to host in a computer network on behalf of their users. Mobile agent leaves its computer to `go' directly to the machine offering the service. There it invokes all the operations locally and also processes the results locally, avoiding all the network traffic. Only when its job is finished does it return to its ``home machine". There are several advantages resulting from the mobility of the programs, depending on their purpose.

3. APPLICATIONS OF AGENTS IN GIS

Over the past two decades, GISs have been adopted widely in support of planning, agriculture, forestry, infrastructure maintenance, and many other fields. Together with the use of the World Wide Web (WWW), GIS could be further developed to allow many more people to have access to GIS functionality and to enhance community participation in planning. The advantages of using such technology are obvious to most users of geodata and geoprocessing resources. Because it can be very time consuming for people to cope with the overwhelming amount of data, agents could be an ideal tool to help the users of Geospatial data and services in this quickly changing environment. Agents can also be used to make GIS user interface intelligent so that non-expert users may use the GIS software more easily.

Geospatial information retrieval and filter, intelligent geospatial search engine, knowledge discovery, decision model assessment and optimization are typical applications based on using agents in GIS. The following subsections briefly describe some applications of agent technology in GIS.

3.1 GIS Software Intelligent User Interface

The development of GIS technology has made it available to a growing number of people from different disciplines and with different backgrounds. Assuming that the user is neither familiar with GIS software, nor with the content and structure of its database, an intelligent interface agent should be able to help GIS software users to retrieve the information they want, build or help them to build complex queries during the interaction.

3.1.1 Web Based Intelligent Interface To GIS Using Agent Technology: There are many instances where GIS users do not perform complex operations using a GIS application, and also don't have to write to GIS databases constantly. They simply want to access certain data elements from the GIS database in order to carry out their day to day tasks so it is sufficient for user to have read only access to GIS data. More over, there is a major push towards providing web access to GIS data; using a simple browser interface user can quickly gain access to GIS data without having to run complex GIS applications.

Sugumaran (1999) has proposed an agent based GIS environment where several types of intelligent agents coexist in a distributed environment, and these agents collaborate with one another in solving a particular problem. These agents guide the user in executing core business processes and the GIS can also be augmented with Web interface to improve access to spatial data. The following generic agents are defined for this GIS environment (Sugumaran, 1999):

- User Interface Agents: These agents keep track of user actions and acquire knowledge on tasks, habits and preferences of users. They can start and execute tasks on behalf of the user or suggest actions for them to take. They constantly learn the user preferences and profile and build a knowledge base of the way in which he/she works and also modify the user environment according to his/her preferred tools, or the commands that are executed the most. The user environment will constantly and dynamically evolve over time.

- Monitoring Agents: These agents monitor every event that occurs in the GIS and channel the relevant information to other appropriate agents. They keep track of the tasks/processes that are being executed within the GIS system and keep track of the state of objects within the GIS environment. When a specific event or user action causes the state of an object to change, the affected parties are notified.
- Business Process Agents: They help the end users with the execution of typical business processes that the user is responsible for.
- Application Program Interface (API) Agents: These agents provide interface to other software systems that may be needed in problem solving. They serve as interfaces between the GIS and external software packages like statistical packages, generalization algorithms, spatial models, etc.

One of the typical applications of this system is assisting a utility company in providing electric and gas connections to new customers (Sugumaran, 1999).

3.1.2 Interface Agent For ArcEdit: Clover Point Cartographics Ltd. , which is a GIS and resource consultants to business and government has developed a new Windows-like graphic user interface simplifying start-up for ArcEdit users. It is called Personal Edit Agent (PEA). This agent provides a start-up interface similar to Microsoft Windows Explorer; allowing user to browse through familiar directories to find the desired coverage, then open it into a standard ArcEdit session.

3.1.3 Interface Agent for ArcInfo: Campos et. al. (1996) have developed a knowledge-based interface agent whose mission is to help users without a knowledge of ArcInfo to access and process spatial data stored in ArcInfo databases. In order to help the user, the agent must be able to:

- Build sequences of ArcInfo commands that provide answers to typical queries to an ArcInfo database,
- Infer, if it is possible, best matches between user's concepts and ArcInfo data, and generate text and graphical output to the user.

Assuming the *request - response* model of interaction between a user and ArcInfo, the task of the interface agent is to (Campos, et. al, 1996):

- interpret user's request,
- translate it into ArcInfo command(s),
- receive some results back from ArcInfo, and
- present them to the user.

In order to convert user's input into sequences of ArcInfo commands, the agent must have *plans* - sequences of actions, which are taken if their preconditions are satisfied, for every type of user request. Representations of plans are stored in a semantic network knowledge base (the system's knowledge base), and used to translate natural language queries, plan definitions, and requests to sequences of ArcInfo commands. In summary the interface agent, using a client-server schema and operating on a LAN or WAN network, receives and processes requests written in plain English, interacting with the user in case of possible mismatches between his/her concepts and the representation of data in the ArcInfo database. The agent builds and sends sequences of commands oriented to provide the information requested by the user to an ArcInfo server, receives, and presents the results of those requests to the user. Campos et. al. (1996) have built a prototype of the interface agent using SNePS (Semantic Network Processing System) and Common Lisp on a Sun Sparc station.

3.2 Making Depositories of Geographical Data Usable with Intelligent Agents

Sengupta and Bennett (2001) have used autonomous agents to make the depositories of geographical data more usable. Considerable insight is often gained when multiple thematic layers are incorporated into the analysis. When such techniques are applied, however, it is assumed that all data:

1) Share a common and appropriate geographical referencing system;
2) Cover a common and appropriate geographical area;
3) Contain non-spatial attribute data of interest; and
4) Are stored in a format that can be read by the selected analytical software.

However, geospatial data stored in network accessible depositories are stored in a variety of proprietary formats, cover varying spatial extents, have different levels of spatial accuracy, map to a variety of geographic referencing systems, and contain a vast and disparate array of attribute data. It is often necessary, therefore, to apply a sequenced set of data transformations to produce a spatial database that contains the desired spatial and attribute information in a consistent and usable format. Finding applicable data and performing the necessary data transformations can be a time consuming task that requires significant expertise in geo-processing technologies. Thus, the need for such transformations places real limits on network-based data mining and exploration.

Sengupta and Bennett (2001) propose the process called "data fusion" to automate the integration of disparate datasets into consistent geographical datasets that are suitable for analysis by users or by data mining technologies. In their presented system, users of data rich digital environments are assisted by ***Intelligent Agents*** that search for relevant data and analytical tools and ***"fuses"*** these resources into usable information.

3.3 Using Mobile agents in GIS

The inevitable trend of information technology is to make things popular and easy to use. However, effective use of information in web is becoming increasingly difficult because of the sheer size of the web and its diversity of resource, thus intelligent software technology for DGIS needs to be developed. Today new mobile agent-based products are introduced in different application areas including DGIS.

GeoAgent (Wu, et. al, 1999), which is one of the mobile agents introduced in geospatial information field inhabit distributed networking environments (e.g. Internet), sense and act on the environments with different appearances, to realize intelligent acquirements, processing, storage, exploration, presentation and decision-making support for geospatial data. This agent has been defined as "an evolutionary intelligent agent with geospatial reasoning ability by utilizing geospatial knowledge" and it possesses the following new properties besides sharing the general properties of agents (Wu, et. al, 1999):

- Geospatial reasoning ability: This ability is based on a geospatial knowledge base and a geospatial reasoning mechanism. Having access to a geospatial knowledge base is necessary because in distributed networking environments, it is not practical for each agent to own its exclusive knowledge base since it will consume system resources so much and cause consistency maintenance problems of knowledge base.
- Evolutionary ability: This ability makes the agent adaptable to changing environments efficiently.

- Mobile ability: Agent travel from one machine to another in public networks securely. This ability enables agent to process data at the data source, rather than fetching it remotely, allowing higher performance operation and reducing network connection time and costs.
- Collaboration: Agent has the ability to cooperate with other agents efficiently and intelligently.

3.4 Agent Mediated Links between GIS and Spatial Modeling software

Spatial Decision Support Systems (SDSS) are designed to provide support for the analysis and evaluation of spatial problems. One method of developing a SDSS is to extend the functionality of existing GIS to incorporate analytical and simulation modeling software needed to solve a particular set of spatial problems. Sengupta et. al. (1996) have proposed a new method of linking GIS and analytical models. The links are implemented using agents that communicate via a modeling language, referred to as Model Definition Language (MDL). The MDL provides an inter-agent communication protocol for cartographic modeling. The design of the geoprocessing system contains five classes of agents: the user interface agent, the model search agents, the spatial modeling agent, the data search agents and the Spatial Operator Implementation Agents (SOIAs). Each class of agent has a specific function to perform in sequence once the user starts the modeling process. MDL constructs represent spatial operators that the agents use to query, retrieve, manipulate and store geographical datasets and spatial modeling software.

Sengupta et. al. (1996) have also proposed a conceptual framework for the development of a Distributed SDSS where the simulation software and required datasets, as well as the GIS may be located on multiple machines and agents are equipped to search, retrieve and put together appropriate data using the WWW. This system would utilize the computing power and storage capacity of several machines to make advanced geoprocessing capabilities more generally accessible. The users could interact the user interface agent on their machine, while additional agents search for datasets and simulation software distributed on the Internet. Current network technology provides the support for the development of such a distributed system thus agents implemented as objects can be transferred between machines during their searches. This conceptual framework can potentially be very useful for the novice GIS user with little system resources (in terms of hardware and software) and enable him/her to access more computational power and data available over the network.

3.5 Automated Map Generalization Using Agents

Generalization is the process that aims at simplifying geographical information to make it meet the users' needs. Generalization enhances of important information and removes of unnecessary one. That process concerns entities (preservation, removal or grouping of objects) and their geometry (preservation or removal of geometric information such as location or shapes). In automated cartography, most of the developed map production systems are complex and tedious to use, and require user intensive interaction and guidance during the design process. AGENT project on multi agent systems in cartography [14] is driven by the objective to address such shortcomings. In this project, agents in three levels are considered (Lamy, et. al, 1999):

- *Micro agents* manage the generalization of individual geographic phenomenon (such as Building);
- *Meso agents* manage groups of objects (such as District);
- *Macro objects* are at a coarser scale still, involved in the broad scale issues of map design (such as town).

Agents can become *active* when they act autonomously and chose the way to generalize by themselves or *reactive* when they are ordered by an upper-layer meso-agent. In the first case, agents have a number of methods in order to implement them as autonomously acting entities generalizing themselves and their map environment. Alternatively they can obey orders to execute plans given to them by organizations.

In an autonomous state, an agent aims at reaching its own goals that means to satisfy a set of constraints: its current "happiness" will therefore depend on the current degree of satisfaction of its constraints.

A violated constraint will be able to propose a set of behaviors to solve its violation. An agent can therefore receive a list of possible actions coming from its set of constraints. As the behaviors that are proposed by one constraint can infer another violation of constraint (to increase size damages accuracy), agent must own decision capabilities to choose and trigger one behavior. That choice is crucial and that procedural knowledge is encapsulated in the agent structure. It depends on a set of parameters which are related both to map specification and to the current state of the agent. The chosen action is triggered and the new state of the agent in respects to its constraints is re-evaluated. However the chosen solution can fail and another one can be tried through a backtracking process.

3.6 Defining the Correctness of Maps Using a Multi-Agent Simulation

Frank (2000) has used a multi-agent formalism to produce a model of map production, map communication and map use. Frank (2000) considers a small task; namely, the production and use of a small street network map for navigation. The model constructed simulates:

- The environment, which is a small part of a city street network;
- A map-maker who explores the environment and collects information, which he uses to construct a map of the area; and
- A map-user who acquires this map to gain knowledge, which he uses to navigate in this environment.

The environment represents the world in which persons live and the agents represent the persons who make and use maps. The simulation includes multiple agents – at least one map-making agent and one or several map-using agents. The approach is to model the complete process, which starts with data collection in reality, produces the map and then the process of map use: reading the map to navigate in an unknown territory. The model uses a two-tiered *reality* and *beliefs* representation, in which reality (facts) and the agents' cognition (beliefs) are represented separately.

The followings summarize the agents' tasks and the procedure of map quality assessment in this simulation (Frank, 2000):

Map Making Agent: The map-making agent is a specialized agent and explores first the environment and then draws a map. This map is then given to the map-using agent.

Map-Using Agents: The map-using agents have the task of moving from the node they are located at to another node in the environment. They intend to travel the shortest path (minimal distance). A map-user first reads the map and adds the knowledge acquired from the map to his set of beliefs about the environment before he plans the shortest path to his destination node.

Maps: Maps are artifacts, which exist in the environment. The map-making agent produces the map usually after he has collected all beliefs about the environment. The map represents these beliefs in a (simulated) graphical format. Maps are simulated in the model as a list of line segments (with start and end map coordinates) and labels at the intersection coordinates; one can think of this as suitable instructions for drawing a map with a computerized plotter. The map, in the form of the drawing instructions, is then read by the map-using agent and translated into a list of beliefs.

Map Quality Assessment: A map is correct if the result of an operation based on the information acquired from the map is the same as if the agent would have explored the world to gain the same information. The proof is in two steps: completeness and correctness. *Completeness* assures that all relevant elements – here nodes and segments – are transformed between the respective representations. The construction of the beliefs of an agent about the environment can be seen as a transformation between two data structures: the data structure, which represents the environment, is transformed into the internal structure of the beliefs. Similarly is the construction of the map a transformation between the data structure of the agent's beliefs into the list of drawing instructions; reading the map is the transformation of the data element of the map into beliefs. These transformations should be applied to all elements and nothing should be 'overlooked'. *Correctness* requires that the transformations preserve the properties important for the decision (in this case the determination of the shortest path). A street segment is added to the beliefs after it is traveled; having traveled the segment ensures that the segment is viable and the cost is the cost just observed. Map-makers translate each segment into a line drawn. The positions are based on the observed coordinate values for intersections. Map-users read the drawn line as viable segments and use the length of the line as an indication of the cost. These operations guarantee that beliefs about viable street segments by the map-maker are communicated to the map-users.

4. SPECIAL SPATIAL SEARCH TOOL

A web search tool, Special Spatial Search (SSS) is developed using HTML and JavaScripting to assist users in geospatial related searching over the web. The following sections describe about three facilities provided by this tool.

4.1 Intelligent Dictionary Search Tool

This tool assists user to use a GIS dictionary more easily and efficiently. User enters a word or phrase or even a sentence to get the exact match or related items in the dictionary. The on-line GIS dictionary provided by Association for Geographic Information is used and related items in that dictionary are shown in a window. User may get more than one item and can select each one to get its definition. The followings are some more details about Intelligent Dictionary Search Tool, explaining its intelligence and advantages with respect to using on-line GIS Dictionary itself:

- When using on-line GIS dictionary itself, if user enters a sentence such as "What is map?" or "I want to know about map", no match will be found. But using this intelligent tool all matches with the word "map" will be listed and all other non-keywords such as "what", "is", "I", "want", "to" will just be ignored.
- When using GIS dictionary itself, if user enters two keywords such as "map accuracy" or "map and accuracy", no match will be found if these combinations are not available in the dictionary. But using this intelligent tool all items matching with "map" and "accuracy" will be listed.
- User can search for only one item at once, when using on-line GIS dictionary. But using this intelligent tool, user can search for more than one item at once, just by typing the word "and" between them or just leaving a space between words.
- Using on-line GIS dictionary, if user enters "Base Mapping" will not get any matching. But using this intelligent interface user will get "Base Map" as a matched item.

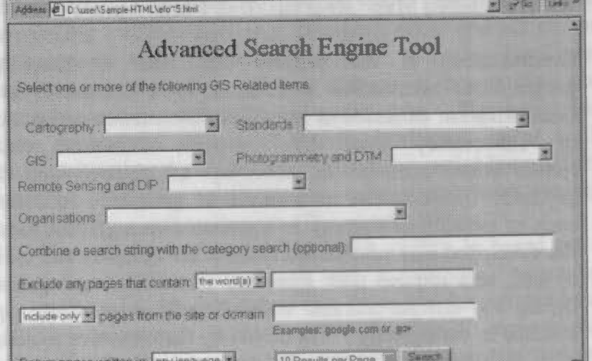
4.2 Advanced Search Engine Tool

Using this tool, user can search for one or more geospatial related item(s) over the web. Geospatial related items are divided to 6 categories and each category is associated with a pull down list containing some related items. Following facilities are provided for user to search more easily and efficiently (See Figure 1):

- Selecting one or more geospatial items from the pull down lists;
- Combining a string search with the category search;
- Excluding any pages that contain special word or phrase;
- Excluding or including only pages from special site or domain;
- Selecting pages written in one of twelve languages or in any language;
- Selecting number of results to be shown in one page (10, 20, 30, 50,100);
- Going back to the main page of SSS.

This tool uses Google search engine to search the Web. Generally, when we search for terms such as "interpolation", "filtering", "resampling", classification" , as these are general terms used in different disciplines, so the result of our search will contain many undesired pages. However, using the Advanced Search Engine Tool, the result will just contain geospatial related pages.

One of the other advantages of this tool is that if we search for a phrase such as "Digital Terrain Modeling" it will also search for its abbreviation "DTM" and if we search for the word "generalization" it will also search for "generalisation".



Advanced Search Engine Tool

Select one or more of the following GIS Related items

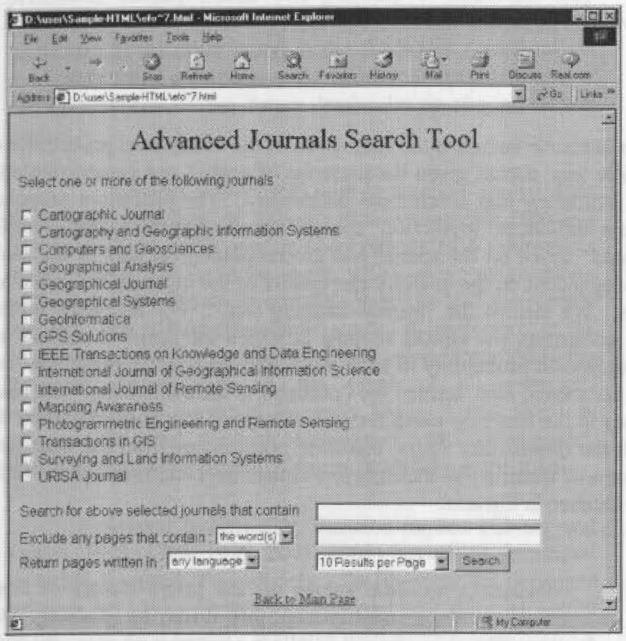
Cartography :
Standards :
GIS :
Photogrammetry and DTM
Remote Sensing and DIP
Organisations
Combine a search string with the category search (optional)
Exclude any pages that contain: the word(s)
include only pages from the site or domain
Examples: google.com or .gov
Return pages written in: any language
10 Results per Page
Back to Main Page

Figure 1: Advanced Search Engine Tool

4.3 Advanced Journals Search Tool

Using this tool, user can search in one or more geospatial related journal(s) more efficiently. Sixteen journals are available to select by the user. This tool uses Google search engine [17] to search the Web. Following facilities are provided for user to search more easily and efficiently (See Figure 2):

- Selecting one or more geospatial Journals;
- Typing a string to be searched within selected journal(s);
- Excluding any pages that contain special word or phrase;
- Selecting pages written in one of twelve languages or in any language;
- Selecting number of results to be shown in one page (10, 20, 30, 50,100);
- Going back to the main page of SSS.


D:\user\Sample-HTML\efo~7.html - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Stop
Refresh
Home
Search
Favorites
History
Mail
Print
Discuss
Real.com
Address
D:\user\Sample-HTML\efo~7.html
Go
Links
My Computer

Advanced Journals Search Tool

Select one or more of the following journals

☐ Cartographic Journal
☐ Cartography and Geographic Information Systems
☐ Computers and Geosciences
☐ Geographical Analysis
☐ Geographical Journal
☐ Geographical Systems
☐ GeoInformatica
☐ GPS Solutions
☐ IEEE Transactions on Knowledge and Data Engineering
☐ International Journal of Geographical Information Science
☐ International Journal of Remote Sensing
☐ Mapping Awareness
☐ Photogrammetric Engineering and Remote Sensing
☐ Transactions in GIS
☐ Surveying and Land Information Systems
☐ URISA Journal

Search for above selected journals that contain
Exclude any pages that contain : the word(s)
Return pages written in : any language
10 Results per Page
Search

Back to Main Page

Figure 2: Advanced Journals Search Tool

One of the advantages of this tool is that user doesn't need to remember the URL address of the journals, what he/she needs is just selecting them from the list. The other advantage is that, two or more journals can be searched simultaneously. Although this tool provides user with 16 journals, however more journals can easily be added.

5. CONCLUSIONS

With the use of the World Wide Web (WWW), many people from different disciplines and with different backgrounds have access to GIS functionality. However, the degree of productivity they can achieve is limited by their lack of technical knowledge about GIS tools. Assuming that the user is neither familiar with GIS software, nor with the content and structure of its database, an intelligent interface will be able to help GIS software users to retrieve the information they want, build or help them to build complex queries during the interaction. Moreover the Internet has become a huge source of Geospatial information and offers already many different services. However, effective use of information in web is becoming increasingly difficult because of the sheer size of the web and its diversity of resource, thus intelligent software technology for DGIS needs to be developed. Application of intelligent agents in the GIS environment is actively being explored and some projects have been reported in the literature where different types of agents are being employed to improve usability of GIS software as well as users access to geospatial data and services available through Internet.

6. REFERENCES

Campos, D. D., Naumov, A. Y., Shapiro, S. C., 1996. Building an Interface Agent for ArcInfo. ESRI User Conference Proceedings, Proceedings 96. http://www.esri.com/library/userconf/proc96/TO50/PAP049/P49.HTM

Clover Point Cartographics Ltd. Suite 202 – 919; Fort Street; Victoria, BC; Canada; V8V 3K3. cloverpoint@pinc.com.

Frank, A. U., 2000. Spatial Communication with Maps: Defining the Correctness of Maps Using a Multi-Agent Simulation. Dept. of Geoinformation. Technical University Vienna.

Google Search Engine. http://www.google.com

Green, S., Hurst, L., Nangle, B., Cunningham, P., Somers, F., Evans, R., 1997. Software Agents: A Review. Trinity College Dublin and Broadcom Éireann Research Ltd.

Lamy, S., Ruas, A., Demazeau, Y., Jackson, M., Mackaness W., Weibel R., 1999. The Application of Agents in Automated Map Generalisation. Proceedings of the 19th International Cartographic Conference, Vol 2, p. 1225-1234.

On-line GIS dictionary, provided by Association for Geographic Information. http://www.agi.org.uk/pages/dictionary/uoedictionary.html

Purvis, M., Cranefield, S., Bush, G., Carter, D., McKinlay, B., Nowostawski, M., Ward, R., 2000. The NZDIS Project: an Agent-Based Distributed Information Systems Architecture. Proceedings of the Hawai'i International Conference On System Sciences, Maui, Hawaii.

Sengupta, R. R., Bennett, D. A., Wade, G. A., 1996. Agent Mediated Links between GIS and Spatial Modeling Software using a Model Definition Language. GIS/LIS'96 Annual Conference and Exposition Proceedings; xv+1284 pp.295-309. American Society for Photogrammetry & Remote Sensing, Bethesda, MD, USA.

Sengupta, R., Bennett, D., 2001. Making Rich Depositories of Geographic Data Usable with Intelligent Systems Technologies. Chapter 5 of "geographic data mining and knowledge discovery". http://www.geog.utah.edu/~hmiller/gkd_text/

Sugumaran, V., 1999. A Web Based Intelligent Interface to Geographic Information Stems Using Agent Technology. Managing Information Technology Resources in Organizations in the Next Millennium. Information Resources Management Association International Conference. Idea Group Publishing, Hershey, PA, USA; 1140 pp. p. 569-75.

Wu, S., Huang, X., Li, Q., 1999. GeoAgents. Towards Digital Earth, Proceedings of the International Symposium on Digital Earth.Science Press. Beijing, China.

MODELING THE CONDITIONAL PROBABILITY OF THE OCCURRENCES OF FUTURE LANDSLIDES IN A STUDY AREA CHARACTERIZED BY SPATIAL DATA

Chang-Jo Chung[a]* and Andrea G. Fabbri[b]

a: Geological Survey of Canada, 601 Booth Street, Ottawa, Canada K1A 0E8
E-mail: chung@gsc.nrcan.gc.ca

b: ITC, Hengelosestraat 99, 7500 AA Enschede, The Netherlands
E-mail: fabbri@itc.nl

Commission IV, WG IV/1

KEY WORDS: Landslide hazard, conditional probability, cross-validation, likelihood ratio function

ABSTRACT:

The most crucial but difficult task in the analysis of the risk due to landslide hazard is the estimation of the conditional probability of the occurrence of future landslides in a study area within a specific time period given the presence of spatial and geomorphologic features. This contribution explores a modeling procedure for estimating that conditional probability. The procedure proposed consists of two steps. The first step is to divide the study area into a number of "prediction" classes according to the hazard level for the likely occurrence of future landslides. "Favourability Functions" based on the spatial and geomorphological data in the study area were used for the sub-division. The number of the classes is dependent on the quantity and quality of the input data. Each class represents a level of hazard with respect to the future landslides. We term it the "hazard-mapping step". For this step, several quantitative models have been developed and the strategy is to reconstruct the typical settings in which the future landslides are likely to occur. The second step is to empirically estimate the conditional probability in each prediction class given the spatial and geomorphologic data based on cross- validation techniques. For the second step, termed the "probability estimation step" the basic strategy of the cross-validation is to construct the prediction classes in the first step using the occurrences of the landslides from the first time-period and then to compare the prediction classes with the distribution of the landslide occurrences from the later time period. The statistics obtained from the comparison provides the crucial quantitative measure to estimate the conditional probability. We illustrate the modeling procedure using a case study, La Baie, Quebec in Canada.

1. Introduction

For a given study area, geomorphologists, experts in surficial earth processes have traditionally constructed a landslide hazard map identifying areas likely to be affected by future landslides. It has been achieved by geomorphological understanding of the area through aerial photographs and field works. The hazard map is usually derived from geomorphological maps containing the basic geomorphological characteristics of landforms and it includes a systematic inventory of the past landslides (Panizza et al, 1998). On the other hand, quantitative geomorphologists and civil engineers have constructed a slope stability map based on deterministic models by studying and interpreting the physical processes of landslides using slope angles, soil cohesion, water saturation capacities, shearing resistance and etc. Each point in the stability map shows a level for "the safety factor of slope failure" of the unit area surrounded the point (Terlien et al, 1995).

While the hazard maps from geomorphological maps usually show three or five levels of hazard, the slope stability maps are shown the level for the safety factor in a continuous scale. These prediction maps representing both the hazard and the slope stability maps are generated for guiding the decision makers for land-use planning. The difficulty facing the land-use planners is how to interpret the hazard levels. For example, if only a small sub-area has been assigned as "extreme hazard class" in a hazard map or has consistently extreme values for the safety factors of slope failures in slope stability map, then it may be relatively easy to make a decision not to allow any types of economic and human activities, the economic sterilization by the ban may outweigh the possible future damage due to the occurrences of future landslides in that small sub-area. However, if a sub-area is classified as "high hazard class" or has a high value for the safety factor in a relatively large sub-area, then although it obviously indicates that the sub-area is possibly be affected by a future landslide, the decision of what to do with the sub-area becomes much more difficult, because the decision makers must compare the economic sterilization with the possible damage.

What the decision makers want to have from the hazard maps or the slope stability maps is not only the relative levels of hazard but also the estimates of the probabilities of the occurrences of future landslides in any given points under certain future scenarios such as a number of future landslides are going to be occurred in the study area within the next 30 years. If we have such estimates of the probabilities, then based on a cost-benefit analysis the decision makers can quantitatively compare the economic sterilization with the possible damage under the assumptions of the scenarios, and hence make a learned and an informed decision, rather than an emotional or a "gut-feeling" decision.

We have adapted a two-step approach proposed by Chung (2002) to tackle the problem of estimating the probability of the occurrence of future landslide given geomorphological information on a sub-area under an assumption of a scenario. The first step is to construct a hazard map with a number of hazard classes as similarly done by the geomorphologists or civil engineers. Then the second step is to estimate the probability in each class given a scenario. Depending on the availability of the input data set with respect to the locations and timings of the past landslides, it may not be possible to estimate the probabilities of the occurrences of future landslides under certain scenarios (Chung, 2002).

The basic strategy is that the occurrences of the past landslides in the study area are first divided into mutually exclusive two groups. One group is used to build a prediction model, which will generate a prediction map showing several levels of prediction classes. Counting the landslides in the other group in the prediction classes of the prediction map, we estimate the conditional probabilities of the occurrences of the future landslides in the prediction classes.

Consider a study area with m layers of spatial data. Each layer consists of several non-overlapping thematic classes such as soil types or observations of a continuous measurement such as slope angle, which represent coverage of "one theme" known to correlate spatially and genetically with the type of landslides under study. At each pixel in the study area, m values are observed, one value representing the thematic classification of the pixel or a continuous measurement at the pixel for each layer. Using these m values at the pixel, we wish to construct a prediction model, which measures the hazard of future landslides at the pixel.

Several favourability function models (Chung and Fabbri, 1993, 1998, 1999, 2001, 2002) have been developed to tackle the prediction models. To illustrate the proposed strategy, we have selected the likelihood ratio function model as the prediction model. We have used a case study from La Baie area in Quebec, Canada (Chung and Perret, 2002).

2. Input spatial data matrix.

Let *A* denote the study area. Consider that we have m map-layers (causal factors) containing geomorphological information related to the occurrences of landslides in *A*. Each layer contains one particular causal factor such as surficial geological information or the slope angles. In addition to the m map-layers, we also have the landslide map-layer containing the scars of the past landslides in *A*. When the landslide scars are rather small with respect to the map scale, we may have only the locations of the past landslides as points on the map rather than the polygons of the scars. In each scar, geomorphologists typically delineate the scarp where the landslide is triggered. Whenever we refer a landslide in this manuscript, we mean it as either the scarp or the point location of the landslide.

For a quantitative study, we overlay a fine grid over *A*, such that each grid cell covers a small unit area. The size of the unit area is depended on both the original input-map-scales and the purpose of the study. The size ranges typically from 5m x 5m to 50m x 50m on ground. Each grid cell is termed a pixel. For each map-layer, one data value, termed pixel value is assigned to each pixel. Consider the case study of La Baie, Quebec. The database typically contains the binary information on the past landslides, a number of thematic classification maps such as the bedrock geology, and DEM information. From the database, a data matrix is constructed for quantitative analysis. In the data matrix, for the landslide map-layer, Y, "1" is assigned to the pixel value, when more than 50% of a pixel is covered by a scarp of a past landslide. Otherwise, "0" is assigned to the pixel value.

A digital elevation model (DEM) containing three spatial information at each pixel: (i) slope angle; (ii) aspect angle; and (iii) elevation, was included in the input data matrix. From the elevation contours, we usually obtain DEM.

The input data matrix usually consists of two different types of spatial data: (i) categorical data layers such as geological map containing rock types, and (ii) continuous data layers such as slope angles. Combining these two different types of data layers is one of the difficulties of constructing prediction maps.

La Baie, Quebec Study area covers 10km x 6km. The image consists of 2000 x 1200 pixels and each pixel covers 5m x 5m in ground. We have five layers of geomorphological information related to the landslides in the area. It contains, (1) bedrock geology (12 rock types), (2) forest coverage (binary), (3) elevations, (4) aspect angles and (5) slope angles. We have the locations of 22 landslides occurred in 1964 and 51 landslides occurred in two time-periods, 1976 and 1996. Seven of the latter 51 landslides were occurred in the same areas of the 22 landslides occurred in 1964. The average size of the 73 landslides is approximately 15m x 15m covering 9 pixels. Among 2,400,000 pixels, 445, 164 pixels covered by the lake, the rivers were excluded in the study, and the remaining 1,954,836 pixels were included in the study. For each variable, the autocorrelation of two pixels depends on the distance between two pixels. Every variable has a strong spatial characteristic.

3. Favourability function model for Step 1

Consider that we have m map-layers containing the causal factors, which are known to correlate with the scarps of landslides in a study area *A*. Consider a pixel i in *A* with m pixel values, $X_1(i) = c_1, \cdots, X_m(i) = c_m$, one for each map-layer, let Y(i) represent the presence (Y(i) = 1) or absence (Y(i) = 0) of a past landslide and let Z(i) represent the presence or absence of a future landslides at the pixel i.

As discussed in Chung (2002), suppose that, for every pixel $i \in A$, we can construct a "favourability" function ***g***:

$$\boldsymbol{g}: (X_1(i), \cdots, X_m(i)) \rightarrow \boldsymbol{R}, \qquad (3.1)$$

where $\boldsymbol{R} = (-\infty, \infty)$ such that $\boldsymbol{g}(c_1, \cdots, c_m)$ represents (or measures) a relative level of hazard of the ith pixel for given m pixel values, $(c_1, \cdots, c_m)$ for both past (Y(i) = 1) or future (Z(i) =1) landslides. For the ith pixel, we rewrite $\boldsymbol{g}(c_1, \cdots, c_m)$ as $\boldsymbol{g}_i(Y=1 \text{ or } Z=1 \mid c_1, \cdots, c_m)$ instead.

Then by computing $\hat{\boldsymbol{g}}_i$ (Y=1 or Z=1 | $c_1, \cdots, c_m$) for every pixel in *A*, we can construct a hazard map for *A*.

4. Likelihood ratio function model for Step 1

Suppose that the study area is divided into two non-overlapping sub-areas, the scarps and the remaining areas. Suppose that the slope angles provide useful information to identify the scarps, and then the slope angle data of the scarps should have unique characteristics that are different from the data for the remaining areas. This suggests that the frequency distribution functions of the scarps and the remaining areas should be distinctly different as illustrated in Figure 1(a). The likelihood ratio function, which is the ratio of the two frequency distribution functions, can not only highlight this difference as illustrated in Figure 1(b) but also be the favourability function satisfying all three conditions discussed in the previous section.

To formalize the idea, let us consider a pixel p with m pixel values, $c_1, \cdots, c_m$ in the whole study area *A* consisting of two sub-areas, the scarps **M** and the remaining area $\overline{\mathbf{M}}$.

M : set of pixels from the scarps, (4.1)

$\overline{\mathbf{M}}$: set of pixels from the remaining area.

Let $f\{c_1, \cdots, c_m \mid M\}$ and $f\{c_1, \cdots, c_m \mid \overline{M}\}$ be the multivariate frequency distribution functions assuming that the pixel is from **M**, and from $\overline{\mathbf{M}}$, respectively. Then the ***likelihood ratio*** (Kshirsagar, 1972; Cacoullos, 1973, McLachlan, 1992) at p is defined as: $\overline{\mathbf{M}}$

$$\lambda_p(c_1,\cdots,c_m) = \frac{f\{c_1,\cdots,c_m \mid M\}}{f\{c_1,\cdots,c_m \mid \overline{M}\}} \quad . \qquad (4.2)$$

For the slope angle whose distributions shown in Figure 1(a), the corresponding likelihood ratio function in logarithmic scale is illustrated in Figure 1(b). The ratio in Figure 1(b) obviously displays significant differences.

The discriminant analysis (Kshirsagar, 1972; Cacoullos, 1973; Chung, 1975; Chung, 1977) consists of estimating the likelihood ratio function $\lambda_p(c_1, \cdots, c_m)$ in (2) based on the data from Table 1. To apply discriminant analysis, we assume that: (i) all m layers are based on continuous observations, (ii) $f\{c_1, \cdots, c_m \mid M\}$ and $f\{c_1, \cdots, c_m \mid \overline{M}\}$ are the normal density functions. Many statistical packages such as SPSS (SPSS, 1994) and S-Plus (Chambers and Hastie, 1992) provide solutions to the traditional discriminant analysis and the variation of the analysis. We take the estimates of $\lambda_p(c_1, \cdots, c_m)$ in (4.2) as the favourability function $\hat{g}_i$ (Y=1 or Z=1 | $c_1, \cdots, c_m$) discussed in Section 3. We compute the estimates of $\lambda_p(c_1, \cdots, c_m)$ for every pixel in the study area. The pixel with the largest estimate is considered as the most hazardous sub-area for future landslides according to this discriminant model.

When we consider several layers simultaneously in the study area, we now have m pixel values, $c_1, \cdots, c_m$, at a pixel p. The likelihood ratio at p is the same as shown in (4.2). Suppose that the m layers provide "independent" sets of information over the scarps and the remaining area (i.e., we assume the conditional independence, as discussed Duda and et al. 1976; Heckerman, 1986; Spiegelhalter, 1986; Agterberg, et al. 1990; Chung and Fabbri, 1998), then (4.2) becomes,

$$\lambda_p(c_1,\cdots,c_m) = \lambda_p(c_1)\cdots\lambda_p(c_m) = \frac{f\{c_1 \mid M\}}{f\{c_1 \mid \overline{M}\}}\cdots\frac{f\{c_m \mid M\}}{f\{c_m \mid \overline{M}\}} \qquad (4.3)$$

The advantage of (4.3) over (4.2) is that it depends only on the univariate distribution function for each layer. The price of the advantage, however, is that the simplification requires the assumption of the conditional independence. Using (4.3), combining two different types (categorical and continuous) of data layers becomes a trivial matter.

To obtain the corresponding empirical distribution functions for $f\{c_i \mid M\}$ and $f\{c_i \mid \overline{M}\}$ from the data, we have employed the smoothed kernel method. The estimator of the likelihood ratio is obtained by:

$$\hat{\lambda}_p(c_1,\cdots,c_m) = \hat{\lambda}_p(c_1)\ldots\hat{\lambda}_p(c_m) = \frac{\hat{f}\{c_1 \mid M\}}{\hat{f}\{c_1 \mid \overline{M}\}}\cdots\frac{\hat{f}\{c_m \mid M\}}{\hat{f}\{c_m \mid \overline{M}\}} \quad , \qquad (4.4)$$

where

$\hat{f}\{c_1 \mid M\}, \hat{f}\{c_1 \mid \overline{M}\}, \cdots, \hat{f}\{c_m \mid M\}, \hat{f}\{c_m \mid \overline{M}\}$

are the corresponding empirical distribution functions. In this case, we take $\hat{\lambda}_p(c_1,\cdots,c_m)$ in (4.4) as the favourability function $\hat{g}_i$ (Y=1 or Z=1 | $c_1, \cdots, c_m$). For every pixel, we compute $\hat{\lambda}_p(c_1,\cdots,c_m)$. The pixel with the largest estimate is considered as the most hazardous sub-area for future landslides according to this model.

Using the 66 locations of the 73 landslides occurred during the past 38 years (1964 – 2002) five layers of geomorphological information related to the landslides in the area, we compute $\hat{\lambda}_p(c_1,\cdots,c_m)$ in (4.4) for each of 1,954,836 pixels in the study area. According to the rank order of $\hat{\lambda}_p(c_1,\cdots,c_m)$ for 1,954,836 pixels, we have divided the study area into 1000 classes. Each class contains 1,955 pixels (covers approximately 0.05 km^2). 1,955 pixels with the highest $\hat{\lambda}_p(c_1,\cdots,c_m)$ were assigned as the most hazardous predicted area in the study area. These classes are shown in Figure 2. As in the color legend consisting of 40 color bars in Figure 2, each color bar represents 25 classes of the 1000 original classes. The most hazardous 25 classes (48,871 pixels covers approximately 1.22 km^2 or 2.5% of the study area) were shown as purple and the subsequent most hazardous 48,871 pixels were shown as pink in Figure 2.

5. Estimation of conditional probability – Step 2

The first step is to construct a hazard map with a number of hazard classes as similarly done by the geomorphologists or civil engineers. The second step is to estimate the probability in each class given a scenario or assumptions.

Let us take an example. Suppose that we build a house of size 10m x 25m (250 m^2) within the most hazardous class (covers approximately 0.5 km^2) of the 1000 classes in Figure 2. The next logical step is to estimate the conditional probability that the house will be affected by a future landslide within the next 35 years. We are proposing to estimate the probability empirically using the cross-validation technique.

Suppose that the time of landslide hazard study in La Baie is 1967 (35 years ago from today, 2002). In the study area, we know the locations of 22 landslides and we have five layers of geomorphological information. Using these 1967 data, we have computed $\hat{\lambda}_p(c_1, \cdots, c_m)$ in (4.4) for each of 1,954,836 pixels in the study area. Similar to Figure 2, the 1,954,836 pixel values of $\hat{\lambda}_p(c_1, \cdots, c_m)$ were sorted from the descending order, and then 1,954,836 pixels into 1000 hazard classes according to the descending order. For Figure 3, we have grouped 1000 classes into 40 groups of 25 classes each. In Figure 3, we have also shown 51 landslides occurred in 1976 and 1996 as black dots. The information on these 51 landslides was not used to construct Figure 3.

The first column in Table 1 represents the portion of the whole study area assigned as "hazard" area for future landslides. The first label "Top 1%" in the column is for the group of the most hazard 10 classes of the original 1000 classes and subsequent "1 - 2%" group is for the next 10 classes. To generate the second column in Table 1, in each of the 1000 classes, we have first made a cumulative count of the 51 landslides. For the classes without the landslides, instead of the cumulative counts, we have used interpolated values. Among the 1000 pairs, we have selected the 20 pairs shown in the second column of Table 1 and it constitutes $2/5^{th}$ of red curve in Figure 4(a). It is termed as "prediction rate curve."

To estimate the probability, we need more assumptions on the future landslides within the next 35 years. We need to have the "expected" number of future landslides in the area within the next 35 years and the "expected" size of the landslides. Since we had 51 landslides for the past 35 years in the study area and the average size of the past 51 landslides is approximately 15m x 15m, we will make the following additional assumptions:

(i) 50 landslides will be occurred in the study area in the next 35 years; (5.1)
(ii) the average size of the 50 "future" landslides will be 15m x 15m.

From the assumptions in (5.1), the affected area by 50 landslides expected within the next 35 years is 50 x 15m x 15m or 450 pixels (size 5m x 5m). If we were to build a house of size 10m x 25m (250 m^2 or 10 pixels) in the most hazard 1% area ("Top 1% area), then the probability that the house will be a part of the whole affected area can be estimated by:

$$\text{An estimate} = 1 - (1 - \delta)^{10\,(=\text{size of house})} \qquad (5.2)$$

$$\text{where} \quad \delta = \frac{450\ (= \textit{size of affected area})}{1954.35}\ \text{probability},$$

"probability" equals to 0.28 shown in the corresponding row for "Top 1% area" of the second column in Table 1. The estimate is 6.26% shown in the corresponding row for "Top 1% area" of the third column. Similarly the numbers in the third column were generated from (5.2) using the corresponding probabilities in the second column in Table 1.

Theoretically speaking, the prediction rate curve must satisfy two conditions: (i) monotone increment function, and (ii) the increment rate (the target of the curve) must be monotone decrement function. Obviously, the red prediction rate curve in Figure 4(a) doesn't satisfy the second condition. For that, we have fitted a linear exponential function for the red prediction rate curve in Figure 4(a) and it is shown in Figure 4(a) as a blue curve. The equation is:

$$\text{Fitted function} = 1 - e^{-0.17 - 7.15\,\text{area}}. \qquad (5.3)$$

The "area" in the equation represents the portion of the whole area as shown in the first column of the Table 1. The corresponding fitted values are shown in the 4th column in Table 1. Using the probabilities in the 4th column, the numbers in the 5th column were generated from (5.2).

The first 20 values in the 3rd and 5th columns were plotted and shown in Figure 4(b). Under the assumptions in (5.1), they are the estimated probabilities that a house of size 10m x 25m (250 m^2 or 10 pixels) in the corresponding 1% areas will be affected by future landslides within the next 35 years. Obviously while the 3rd column is based in empirical estimates, the 5th column is based on the fitted prediction rate curve shown as blue curve in Figure 4(a).

6. Acknowledgments

We wish to thank Dr. P. Didier, Geological Survey of Canada, who has provided the spatial data for the study. The study was also partly funded from a research grant provided to the Spatial Data Analysis Laboratory of Geological Survey of Canada by PCI Inc., Richmond Hill, Canada.

REFERENCES

Agterberg, F.P., Bonham-Carter, G.F., and Wright, D.F., 1990. Statistical pattern integration for mineral exploration. *In,* Gaal, G. and Merriam, D.F., eds., *Computer Applications in Resource Estimation, Prediction and Assessment of Metals and Petroleum.* New York, Pergamon Press, p. 1-21.

Cacoullos, T., 1973, Discriminant Analysis and Applications, Academic Press, New York, 434p.

Chambers, J.M., and Hastie, T.J., 1992, Statistical Models in S, Wadsworth & Brooks/Cole, Pacific Grove, California, 608p.

Chung, C.F., 1975, An application of classification analysis for Project Appalachia data. In Proceedings of the 14th APCOM Symposium, Pennsyvania State Univeristy, p.299-311.

Chung, C.F., 1977, An application of discriminant analysis for the evaluation of mineral potential. In Geological Survey of Canada paper 75-1C, Ottawa, Canada, p.141-148.

Chung, C.F. and Fabbri, A.G., 1993, The representation of geoscience information for data integration. Nonrenewable Resources, v. 2, n. 2, p. 122-139.

Chung, C.F., and Fabbri, A.G.: 1998, Three Bayesian prediction models for landslide hazard, In A. Bucciantti (ed.), Proceedings of International Association for Mathematical Geology 1998 Annual Meeting (IAMG'98), Ischia, Italy, pp. 204-211.

Chung C.F. and Fabbri, A.G.: 1999, Probabilistic prediction models for landslide hazard mapping, Photogrammetric Engineering and Remote Sensing, 65-12, 1389-1399.

Chung, C.F. and Fabbri, A.G.: 2001, Prediction model for landslide hazard using a Fuzzy set Approach, In M. Marchetti and V. Rivas (eds.), Geomorphology and Environmental Impact Asssessment, Balkema, Rotterdam, in press.

Chung, C.F. and Fabbri, A.G.: 2002, Validation for landslide hazard using a Fuzzy set Approach, In M. Marchetti and V. Rivas (eds.), Geomorphology and Environmental Impact Asssessment, Balkema, Rotterdam, in press.

Chung, C.F. and Perret, D.: 2002, Landslide hazard mapping in La Baie, Quebec, Canada, in preparation.

Chung, C.F.: 2002, Two-step approach for spatial prediction models for landslide hazard mapping, in preparation.

Duda, R.O., 1980, The Prospector systems for mineral exploration. SRI International Final report.

Heckerman, D., 1986, Probabilistic interpretations for MYCIN's certainty factors. *In* L.N. Kanal and J.F. Lemmer Eds., *Uncertainty in Artificial Intelligence*. Elsevier Science Pub., North-Holland, pp. 167-196.

Kshirsagar, A.M., 1972, Multivariate Analysis, Marcel Dekker Inc., New York, 534p.

Panizza M., Corsini M., Soldati M. and Tosatti G.: 1998, Report on the use of new landslide susceptibility mapping techniques, In J. Corominas, J. Moya, A. Ledesma, J.A. Gili, A. Loret and J. Rius (eds.), New Technologies for Landslide Hazard Assessment and Management in Europe (NEWTECH). Final Report, October 1998 of CEC Environment Programme Contract ENV-CT96-0248, UPC, Barcelona, pp.13-31.

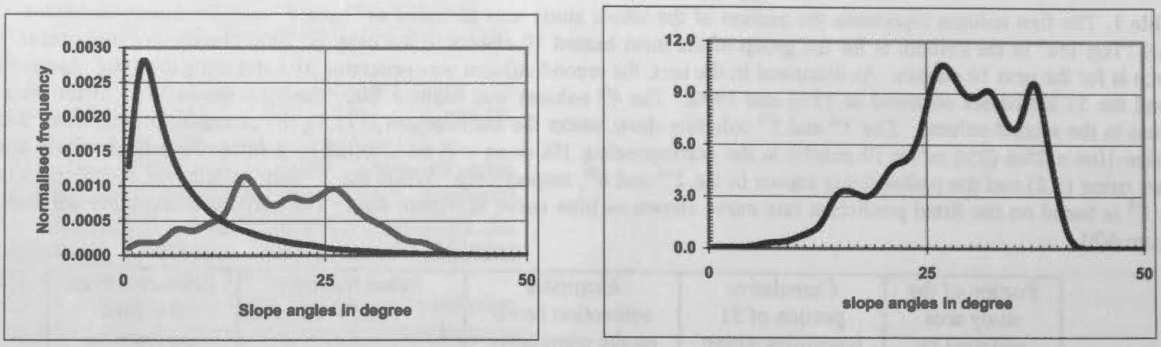
Spiegelhalter, D.J., 1986, A statistical view of uncertainty in expert systems. *In* W.A. Gale, ed., *Artificial Intelligence and Statistics*, Addison-Wesley Pub., Reading, Mass., pp. 17-55.

SPSS, 1994, SPSS Professional Statistics 6.1, SPSS Inc., Chicago, Illinois, 385p.

Terlien, M.T.J., van Westen, C.J., and van Asch, T.W.J.: 1995, Deterministic modeling in GIS-based landslide hazard assessment, In A. Carrara and F. Guzzetti (eds.), Geographic Information Systems in Assessing Natural Hazards, Kluwer, Dordrecht, pp. 57-78.

Table 1. The first column represents the portion of the whole study area assigned as "hazard" area for future landslides. The first label "Top 1%" in the column is for the group of the most hazard 10 classes of the original 1000 classes and subsequent "1 – 2%" group is for the next 10 classes. As discussed in the text, the second column was generated by comparing the 1000 classes for Figure 3 and the 51 landslides occurred in 1976 and 1996. The 4th column was based a fitted function shown in (5.3) for the empirical values in the second column. The 3rd and 5th columns show, under the assumptions in (5.1), the estimated probabilities that a house of size 10m x 25m (250 m^2 or 10 pixels) in the corresponding 1% areas will be affected by a future landslides within the next 35 years using (5.2) and the probabilities shown in the 2nd and 4th, respectively. While the 3rd column is based in empirical estimates, the 5th is based on the fitted prediction rate curve shown as blue curve in Figure 4(a). The corresponding plots are shown in the Figure 4(b)

Portion of the study area assigned as hazard area	Cumulative portion of 51 landslides within the class	Empirical estimation based on the cumulative portion	Fitted function $1 - e^{-0.17 - 7.15\,\text{area}}$	Estimated from the fitted exponential function
Top 1%	0.2800	0.0626	0.2177	0.0362
1 – 2%	0.0373	0.0085	0.0540	0.0133
2 – 3%	0.0704	0.0161	0.0503	0.0124
3 – 4%	0.0476	0.0109	0.0468	0.0115
4 - 5%	0.0452	0.0104	0.0436	0.0107
5 – 6%	0.0480	0.0110	0.0406	0.0100
6 – 7%	0.0559	0.0128	0.0378	0.0093
7 – 8%	0.0186	0.0043	0.0352	0.0087
8 – 9%	0.0455	0.0104	0.0327	0.0081
9 – 10%	0.0120	0.0028	0.0305	0.0075
10 – 11%	0.0150	0.0034	0.0284	0.0070
11 – 12%	0.0207	0.0047	0.0264	0.0065
12 – 13%	0.0186	0.0043	0.0246	0.0061
13 – 14%	0.0186	0.0043	0.0229	0.0056
14 - 15%	0.0171	0.0039	0.0213	0.0053
16 – 17%	0.0100	0.0023	0.0198	0.0049
17 – 18%	0.0100	0.0023	0.0185	0.0046
18 – 19%	0.0090	0.0021	0.0172	0.0042
19 – 20%	0.0090	0.0021	0.0160	0.0040



(a) (b)

Figure 1. (a) Two empirical frequency distribution functions of 73 landslides area (in red) and the remaining area (in blue) using a kernel method. (b) The empirical likelihood ratio function based on two empirical distribution functions in (a).

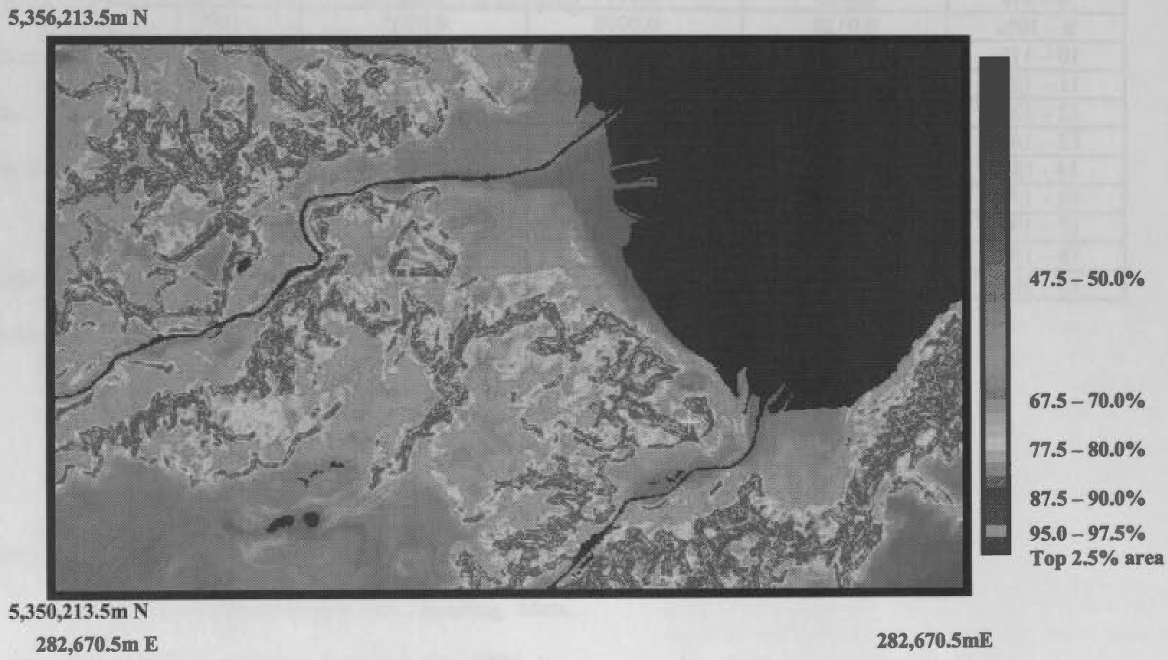

5,356,213.5m N
47.5 – 50.0%
67.5 – 70.0%
77.5 – 80.0%
87.5 – 90.0%
95.0 – 97.5%
Top 2.5% area
5,350,213.5m N
282,670.5m E
282,670.5mE

Figure 2. Landslide hazard prediction map based on 73 landslides (22 in 1967, 51 landslides in 1976 and 1996) and five layers (bedrock geology, forest coverage, elevation, aspect angle, slope angle maps) of geomorphological map information using likelihood ratio function model.

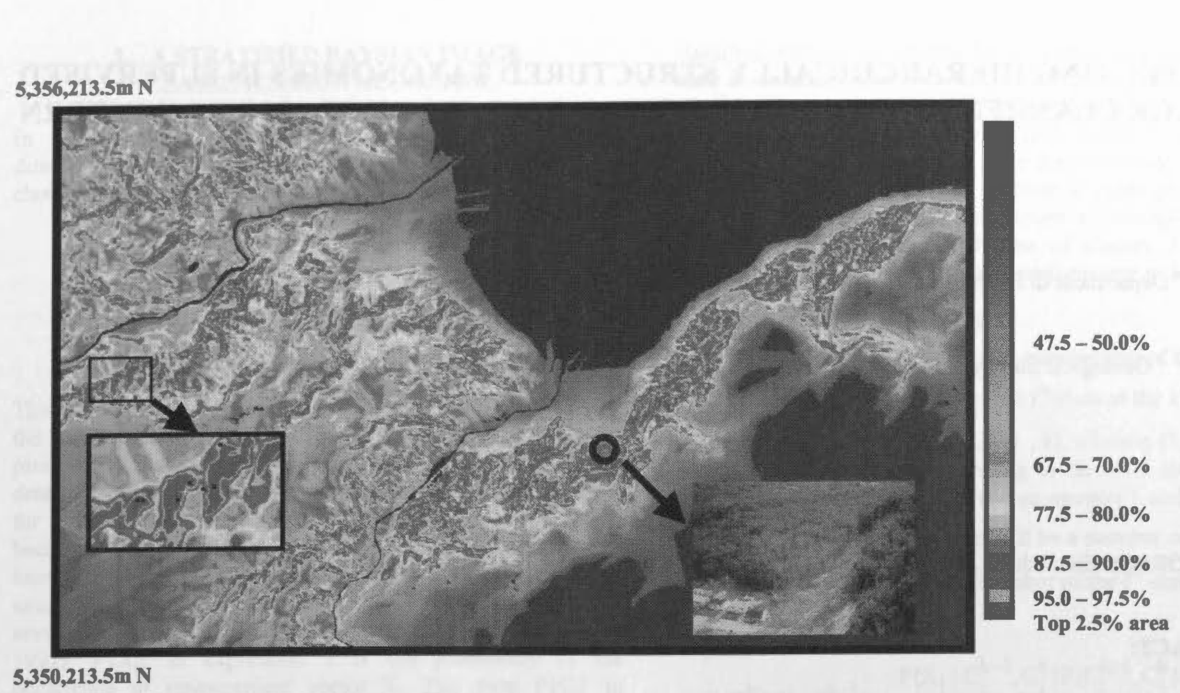


Figure 3. Landslide hazard prediction map based on 22 landslides occurred in 1967 and five layers (bedrock geology, forest coverage, elevation, aspect angle, slope angle maps) of geomorphological map information using likelihood ratio function model. The 51 black dots represent 51 landslides occurred in 1976 and 1996. The left side inset is an enlargement of a small area in black rectangle area in the middle left side. The right side inset with "Year 1996" is an image showing a photograph of a landslides occurred in 1996 at the black circle area in the middle area.

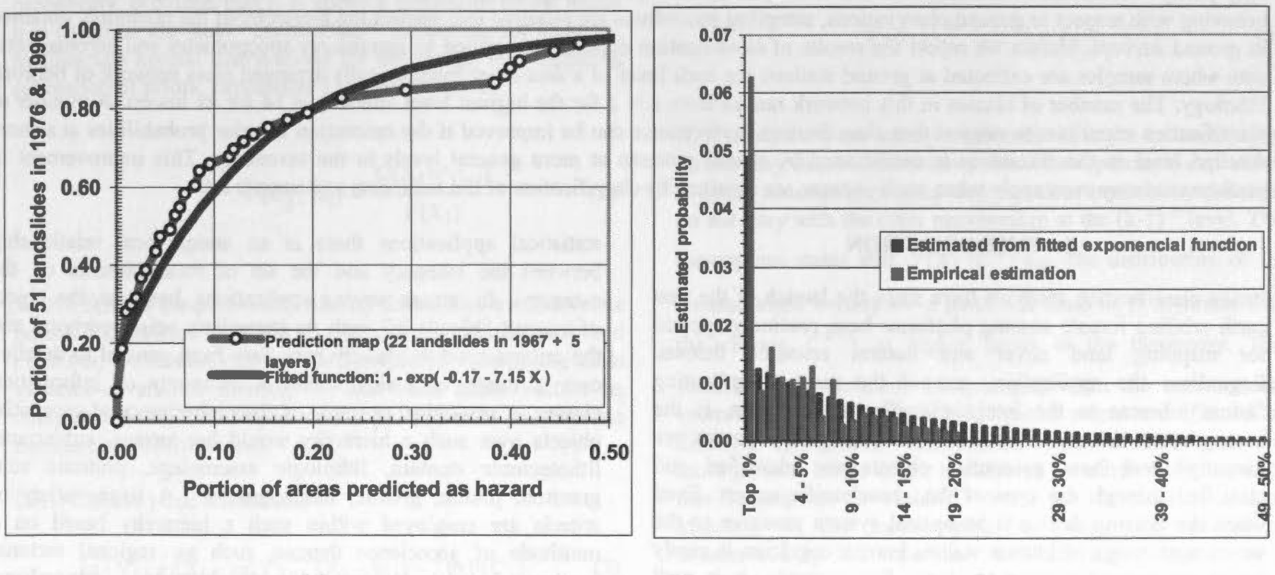


Figure 4. (a) Prediction rate curve for the prediction map shown in Figure 2. It was obtained by comparing the 1000 hazard classes generated for Figure 3 and the 51 landslides occurred in 1976 and 1996 as discussed in the text. The 20 pairs shown in the second column of Table 1 constitutes 2/5th of red curve. The fitted function shown in (5.3) is shown as blue curve. (b) It shows, under the assumptions in (5.1), the estimated probabilities that a house of size 10m x 25m (250 m^2 or 10 pixels) in the corresponding 1% areas will be affected by a future landslides within the next 35 years using (5.2) and the prediction rate curves shown in (a). Obviously while the red histogram is based in empirical estimates, the blue histogram is based on the fitted prediction rate curve shown as blue curve in Figure 4(a). The corresponding table values are shown in the 3rd and 5th columns in Table 1.

RESPECTING HIERARCHICALLY STRUCTURED TAXONOMIES IN SUPERVISED IMAGE CLASSIFICATION: A GEOLOGICAL CASE STUDY FROM THE WESTERN CANADIAN SHIELD

E. M. Schetselaar [a,*] and C. F. Chung[b]

[a] Department of Earth Systems Analysis, International Institute for Geoinformation Science and Earth Observation, Hengelosestraat 99 7514 AE Enschede The Netherlands, schetselaar@itc.nl

[b] Geological Survey of Canada Geological Survey of Canada, Ottawa, 601 Booth St. Ottawa, Canada K1A 0E8, chung@NRCan.gc.ca

Commission IV, WG IV/1

KEY WORDS: Hierarchical, Knowledge Base, Classification, Integration, Data Mining, Geology, Geophysics

ABSTRACT:

Supervised image classification is based on assembling statistics between site-specific ground observations and remotely sensed measurements. If supervised image classification is applied within the context of a particular theme (e.g. vegetation, soil, lithology, land use), one is often confronted with extracting the statistical correlations from a hierarchically arranged network of taxonomic classes spatially abstracted and hierarchically generalized over a range of mapping scales. In practice, however, supervised image classification often appears to be based on a pragmatic approach, a priori categorizing the samples into classes from various levels or from a subset of the hierarchic network. Such approaches are suspect, since the sampling often appears to be biased towards maximizing the discrimination potential of the multivariate data set at cost of representing the categories identified by direct ground observation. The classification performance is, as a result, often assessed within the context of arbitrarily defined class schemas that only partly correspond to the schemas obtained by field surveys. Clearly, to gain more insight in how supervised classifiers are behaving with respect to ground observations, sampling procedures are required that respect the hierarchy of the taxonomy obtained in ground surveys. Herein we report the results of classification experiment applied to gamma-ray spectrometry and aeromagnetic data where samples are extracted at ground stations for each level of a four-level hierarchically arranged class network of bedrock lithology. The number of classes in this network ranges from $n = 2$ for the highest level and to $n = 14$ for its lowest. A number of classification experiments suggest that classification performance can be improved if the estimation of prior probabilities at a more detailed level in the taxonomy is conditioned by spatial patterns at more general levels in the taxonomy. This improvement in performance may even apply when such patterns are obtained by classification of the same data and sample set.

1. INTRODUCTION

Image classification methods have since the launch of the first earth orbiting remote sensing platforms been routinely applied for mapping land cover and natural resource themes. Regardless the application, one of the most complicating factors inherent to the image classification problem, is the incongruence between what and how physical properties are measured and how geospatial objects are identified and classified through the eyes of the geoscientific expert. Even when the sensing device is an optical system sensitive to the wavelength range of human vision, human cognition is rarely based on spectral properties alone. For example, it is well known to earth scientists that colour (i.e. perceived relative spectral properties between 400 and 700 nm) is often one of the least diagnostic and most misleading criteria for classifying a particular mineral or rock type (as it is frequently a function of the abundance of trace elements in the crystal lattice respectively state of weathering).

Clearly, image classification is no exception to any other application of multivariate statistics, namely to explore the universe between what an expert sees and categorizes with what is measured. Unique to the image classification task, however, is the uncertainty to what extent observed features match patterns in fields of measurements. In most multivariate statistical applications there is an unequivocal relationship between the category and the set of measurements on the category. In remote sensing applications, however, the object of interest "blends-in" with its immediate neighbourhood and the criteria used to classify may vary from general to detailed over a range of scales within a taxonomy of information classes. A geological example of class inheritance of geospatial objects over such a hierarchy would be: terrane, subterrane, lithotectonic domain, lithologic assemblage, plutonic suit, granitoid pluton, granite, monzogranite. A large variety of criteria are employed within such a hierarchy based on a multitude of geoscience themes, such as: regional tectonic synthesis, geochronology, lithogeochemistry, mineralogy, texture and structure.

The problem is that it is unclear, how the above criteria are related to the patterns in the remote sensing measurements that provides the discrimination potential. Any attempt to assign ground based information classes to measurements by multivariate statistical methods, therefore, should not only be based on samples of information classes at a particular level within a taxonomy, but also exploit the a priori knowledge on how the information classes are ordered within it. This paper presents image classification methods and experiments that respect the taxonomic knowledge of earth scientists by which geospatial objects have been classified.

2. A STRATIFIED BAYSIAN IMAGE CLASSIFICATION ALGORITM

In general form, the posterior probability that an m-dimensional vector of measurement X at pixel p belongs to class C_j can be expressed in Bayesian formula:

$$P\{C_j \mid X_l\} = \frac{P\{X_l \mid C_j\}P\{C_j\}}{P\{X_l\}} \qquad (1)$$

The term $P\{ X_l \mid C_j \}$ in this expression is the probability that the measurement vector will take on the value X_l given that the pixel is a member of class C_j. This probability can be determined by sampling a population of measurement vectors for pixels known to be a member of class C_j. In practice, because of the limited availability of measurement vectors with known class membership, this probability is estimated by assuming a particular distribution, such as the multivariate normal or the multivariate student t-distribution (McLachlan, 1992). $P\{X_l\}$ in expression 1 is the probability of the occurrence of measurement vector X_l. The term $P\{C_j\}$ in expression 1 is the prior probability that a pixel will be a member of class C_j. This probability is estimated by computing the mixing proportions of the total number of samples N_j of class C_j to the total number of samples over all the classes.

Dependent on the assumption for the distributions to estimate $P\{X_l| C_j\}$, a classification decision rule is defined whereby the pixel p is allocated to the class with the highest posterior probability, provided that it is above a threshold, below which p is assigned the label "unclassified". Assuming, for example, multivariate normal distributions for the classes and including estimation of priors, expression 1 can be written as:

$$P\{C_j \mid X_l\} = \frac{f_j(X_l)P\{C_j\}}{P\{X_l\}} \qquad (2)$$

Where $f_j(X_l)$ is the probability density function of a multivariate normal distribution for class C_j. Using a Maximum likelihood classifier, a decision rule can be formulated substituting class variance-covariance matrices V_j and class mean vectors m_j computed from the sample set to parameterize multivariate normal class distributions:

DR1: Choose j that minimizes:

$$\ln |V_j| + (X_l - m_j)' V_j^{-1} (X_l - m_j) - 2\ln P(C_j) \qquad (3)$$

The incorporation of prior probabilities in the decision rule led some workers to the development of methods to improve on the estimation of prior probabilities. Strahler (1980) showed how thematic classes from one or more additional variables can be used to refine the estimation of prior probabilities. Gorte (1998) extended such concepts to non-parametric classification methods (based on the k-nearest neighbour method) and used iterative estimation of class mixing proportions to obtain local priors. In this paper, the estimation of prior probabilities is based on a stratified classification over a taxonomy based on geoscientific knowledge of the study area. Thematic maps or classified patterns at general levels within this hierarchy are used as collateral variables to estimate prior probabilities for the classification at a more detailed level within the hierarchy. The classification is called stratified, because it proceeds stepwise from general to detail over the taxonomy. Distinctive properties of the classification problem at more general levels in the hierarchy leads to the formulation of assumptions on the sample set and spatial distribution of classes that can be exploited to improve classification performance at lower levels in the hierarchy.

We start by redefining C_j, the class membership of the jth class, as C_j^k : the class membership of the jth class at the kth level in a taxonomy ranging from [k-q, ...,k-1 , k], where q {1, .. , k-1} is an index variable between referring to the level above the kth level in the taxonomy. We rewrite expression 1 and define the posterior probability that a pixel will be a member of class C_j^k given X and the fact that p is a member of the ith class at the $(k-1)^{th}$ level:

$$P\{C_j^k \mid X_l, C_i^{k-1}\} = \frac{P\{X_l \mid C_i^{k-1}, C_j^k\}P\{C_i^{k-1}, C_j^k\}}{P\{X_l, C_i^{k-1}\}} \qquad (4)$$

If the jth class descends directly from one and only one class at the $(k-1)^{th}$ level, then it can be shown that:

$$P\{X_l \mid C_i^{k-1}, C_j^k\} = P\{X_l \mid C_j^k\} \qquad (5)$$

If, instead, there are multiple inheritances in the taxonomy (e.g. a class may descend from more than one superclass) expression 5 is the assumption stating that the class probability densities do not vary with the class membership at the $(k-1)^{th}$ level. This assumption states that $P\{X_l \mid C_j^k\}$ e.g. the distribution of the measurement vectors for a particular class C_j is invariant with the classes C_i^{k-1} at higher levels in the taxonomy. This assumption is violated, for example, when the measurement vectors of classes with limited spatial extent are affected by different 'background' distributions of measurement vectors at more general levels in the taxonomy. Another example where this assumption is violated would be chemical alterations affecting X_l (by metamorphic or metasomatic processes) conditioned by the superclass in which a particular class is enclosed.

The joint probability $P\{C_i^{k-1}, C_j^k\}$ in expression 4 can be rewritten in the form of a conditional probability:

$$P\{C_i^{k-1}, C_j^k\} = P\{C_j^k \mid C_i^{k-1}\}P\{C_i^{k-1}\} \qquad (6)$$

Substituting expressions 5 and 6 into 4 we obtain:

$$P\{C_j^k \mid X_l, C_i^{k-1}\} = \frac{P\{X_l \mid C_j^k\}P\{C_j^k \mid C_i^{k-1}\}P\{C_i^{k-1}\}}{P\{X_l, C_i^{k-1}\}} \qquad (7)$$

Since $\displaystyle\sum_{j=1}^{J} \frac{P\{X_l \mid C_j^k\}P\{C_j^k \mid C_i^{k-1}\}P\{C_i^{k-1}\}}{P\{X_l, C_i^{k-1}\}} = 1$,

$$P(X_l, C_i^{k-1}\} = \sum_{j=1}^{J} P\{X_l \mid C_j^k\}P\{C_j^k \mid C_i^{k-1}\}P\{C_i^{k-1}\}$$

So that $P\{C_i^{k-1}\}$ cancels from de numerator and denominator after substitution in expression 7, we obtain:

$$P\{C_j^k \mid X_l, C_i^{k-1}\} = \frac{P\{X_l \mid C_j^k\}P\{C_j^k \mid C_i^{k-1}\}}{\sum_{j=1}^{J} P\{X_l \mid C_j^k\}P\{C_j^k \mid C_i^{k-1}\}}$$

$$= \frac{f_j(X_l)P\{C_j^k \mid C_i^{k-1}\}}{\sum_{j=1}^{J} f_j(X_l)P\{C_j^k \mid C_i^{k-1}\}} \qquad (8)$$

The term $P\{C_j^k \mid C_i^{k-1}\}$ in expression 8 can be directly computed from random samples that are hierarchically stratified into k levels.

Accordingly, the classification decision rule can be states as:

DR2: Choose j that minimizes:

$$\ln|V_j| + (X_l - m_j)'V_j^{-1}(X_l - m_j) - 2\ln P(C_j \mid C_i^{k-1}) \qquad (9)$$

Note that because of the assumption stated in expression 5, this expression could have simply been obtained by substituting the conditional prior into expression 3. If this assumption is not valid, however, the joint conditional probabilities of the form $P\{X_l \mid C_i^{k-1}, C_j^k\}$ must be computed from all the occurring combinations between j classes at the kth level and i classes at the (k-1)th level.

3. A CASE STUDY PREDICTION OF UNITS FROM A BEDROCK TAXONOMY OF THE WESTERN CANADIAN SHIELD

The above-derived method was tested in a number of classification experiments to predict a 4-level taxonomy of bedrock units from gamma-ray spectrometry and aeromagnetic data gridded on 100 x 100 meter pixels acquired over the western margin of the exposed Canadian Shield. The five grid (image) variables are: K: potassium channel (fig. 1a), eTh,: thorium channel (fig. 1b), eU uranium channel (fig. 1c), total magnetic field and residual magnetic field (fig 1dl). These five grids were augmented by their 9 x 9 average filtered derivatives, in analogy to the approach of Switzer (1980). This method exploits the spatial autocorrelation between pixels within each of the variables under the assumption that the alternation of bedrock units occurs on a large scale than that of a single pixel (Switzer, 1980). The statistical relationships between the image variables (Figure 1) and bedrock units was estimated at the field stations of several geological mapping projects, yielding 3528 samples, each having values of the K, eTh, eU, total magnetic field and residual magnetic field. The bedrock taxonomy is shown in Figure 2.

In a previous study samples of bedrock units were amalgamated from the second and third level of this taxonomy (Schetselaar et al., 2000). This classification coincided for 70% to geological maps of the study area compiled on 1 : 50.000 scale (McDonough et al., 2000 and references therein). As can be seen from Figure 2 the relationships between the first, second and third levels within this taxonomy are defined by single inheritance. The relationships between the fourth and third level (between mylonitic units and protoliths) however is defined by multiple inheritance. These were forced to single inheritance by masking the shear zones with the mylonitic units from the three levels above it. This was necessary because the relationships between original rock units (protoliths) and mylonitic units could not be recovered from the digital map database of the study area. The number of classes in this network ranged from n = 2 for the highest level and to n = 14 for its lowest. The hierarchic network was structured downwards according to lithotectonic domains (n = 2), basement-cover-plutonic assemblages (n = 4), bedrock units (n = 12 & n = 14). The bedrock classification was based on mineralogical, textural and structural field diagnostics.

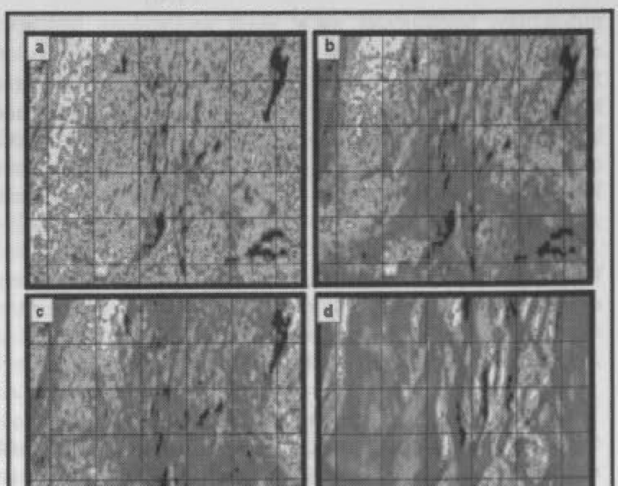


Figure 1. Four grid (image) variables used in the classification experiments, a) Potassium Channel (K); b) Thorium Channel (eTh); c) Uranium Channel (eU); residual magnetic field grid.

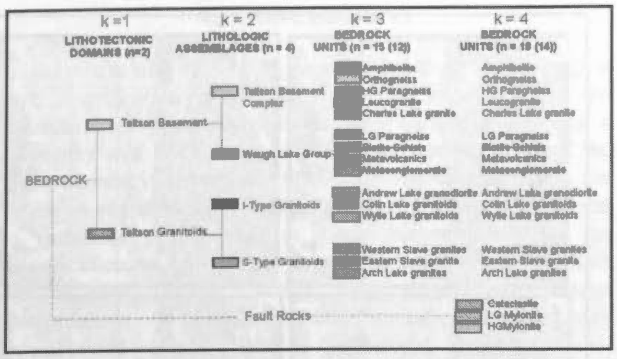

k =1
LITHOTECTONIC DOMAINS (n=2)
k = 2
LITHOLOGIC ASSEMBLAGES (n = 4)
k = 3
BEDROCK UNITS (n = 15 (12))
k = 4
BEDROCK UNITS (n = 18 (14))
BEDROCK
Taltson Basement
Taltson Granitoids
Taltson Basement Complex
Waugh Lake Group
I-Type Granitoids
S-Type Granitoids
Amphibolite
Orthogneiss
HG Paragneiss
Leucogranite
Charles Lake granite
LG Paragneiss
~~Biotite-Schists~~
Metavolcanics
~~Metaconglomerate~~
Andrew Lake granodiorite
Colin Lake granitoids
Wylie Lake granitoids
Western Slave granites
Eastern Slave granite
Arch Lake granites
Fault Rocks
Cataclasite
LG Mylonite
HGMylonite

Figure 2. Bedrock taxonomy of the Taltson magmatic zone (Canadian Shield of NE Alberta) after McDonough et al. 2000 (and references therein). Classes that are striped out do not have enough samples (n < 10) for estimating the class distributions.

At the fourth level two mylonitic units were added to the third level. This additional differentiation was made because the fabric and mineralogy within these units are altered by high shear strains to an extent that they their field diagnostics are not representative to their undeformed equivalents at the third level.

First a standard Bayesian classification was applied to all levels separately, computing overall priors from the samples. The classified map patterns and their coincidence with compiled geological maps resulting from these classification experiments are shown in Figure 2a-h respectively Figure 3. Note that the fall-off rate of the coincidence percentages for the classes is higher than the overall percentages. This is obviously due to the large bias towards the classes of great spatial extent at the first and second levels of the taxonomy. The average class coincidence percentage are rapidly decreasing with increasing number of classes, ranging from 80 at the first level to 40 percent at the fourth level.

Next, two types of stratified classification experiments were conducted were class information at more general levels conditioned the classifications at more detailed levels of the taxonomy:

1. The computation of priors at k=2 and k=3 conditioned by units of map patterns at k=1 and k=2 of the taxonomy. This experiment is comparable to situations where detailed bedrock lithology classification is conditioned by regional geological maps showing lithotectonic domains and lithologic assemblages. This scenario is considered realistic in reconnaissance mapping projects where regional geological maps (typically between 1:250,000 and 1,000000 scales) are available.
2. The computation of priors at k=2 and k=3 conditioned by units of a map pattern obtained from a non-parametric classification method at k=1. Note that in this case no additional map layers or interpretations are used in the classification. The stratification is based on the same sample set used for the non-stratified classification experiments and proceeds stepwise from general to detailed levels in the taxonomy.

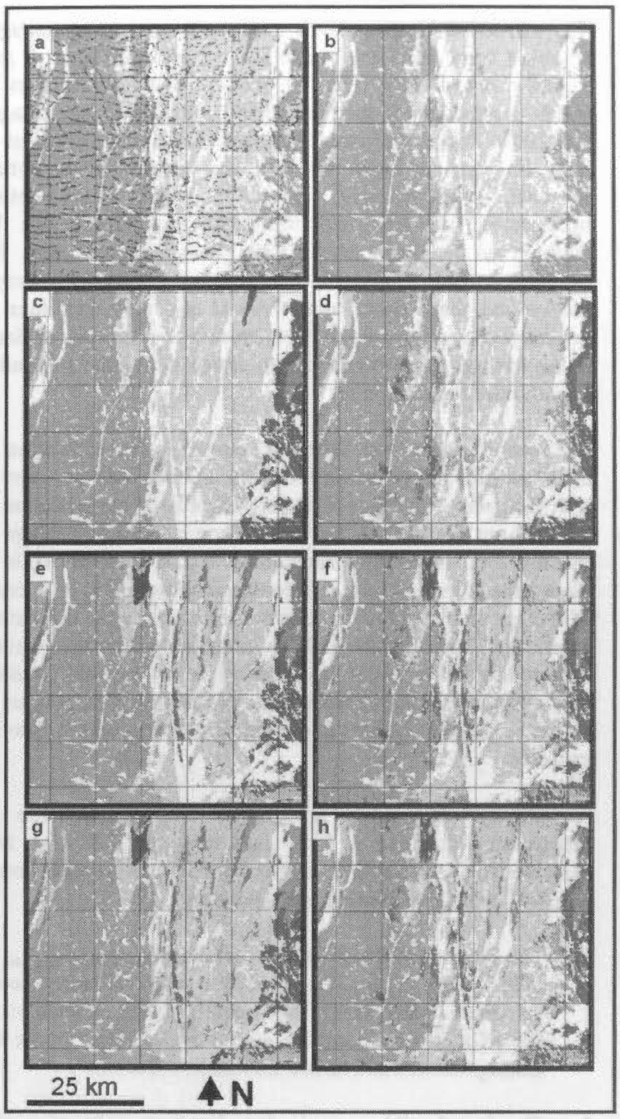

a
b
c
d
e
f
g
h
25 km
N

Figure 3. Classified patterns and image variables. (a) map compilation lithotectonic domains with overlay of sample locations; (b) classification lithotectonic domains; (c) map compilation lithologic assemblages; (d) classification lithologic assemblages; (e) map compilation bedrock units; (f) classification bedrock units; (g) map compilation bedrock units (including fault rocks); (h) classification bedrock units (including fault rocks).

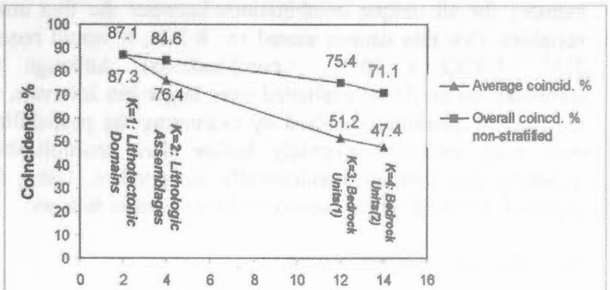


Figure 4. Overall and average coincidence percentages between classifications and map compilations for the four levels of the bedrock taxonomy.

The classified map patterns and coincidence percentages resulting from these experiments are shown in Figure 5 and Figure 6 respectively. The first method obviously appears to yields a considerable increase of the coincidence percentages (ca. 10%). This initial result suggests that it may be useful to integrate regional scale maps in a stratified classification approach. Alternatively, spatial patterns representing lithotectonic domains and lithologic assemblages can be outlined on remotely sensed data. Such regional units, for example, can often be easily outlined using anomaly patterns, texture and shape on colour-enhanced grid representations of aeromagnetic and gravity data, whereas it is often very difficult to use such grid representations to assign individual anomalies to particular bedrock units.

The second method was based on non-parametric estimation of the class probability distributions for lithotectonic domains at k=1. It appeared that the Maximum Likelihood classification at k=1 did not result in an increase of coincidence percentage at k=2 and k=3. Apparently the number of misclassified pixels of large units at k=1 was not compensated by reduction of overlap of class probability distributions or refinement in the estimation of priors at lower levels. Only in situations where the number of misclassified pixels is low with respect to the misclassified pixels due to overlap between class probability distributions at higher k, an improvement in classification performance is to be expected.

An attempt was made to improve the classification of lithotectonic domains at k=1, exploiting the following two characteristics of the classification problem at general levels (e.g. k=1 or k=2) of the bedrock taxonomy:

1. The large number of samples available ($n_{\text{Taltson Basement}}$ = 1708 and $n_{\text{Taltson Granitoids}}$ = 1512) at k=1 well spread over the entire study area. This permits direct estimation of $P\{C_j|X_l\}$ from the samples instead of estimation under the assumption of multivariate normal distribution.

2. The fact the lithotectonic domains form units of spatial dimensions at least two orders of magnitude larger with respect to the pixel size of 100 x 100 metres. This permits extensive post-classification smoothing without the risk of eliminating classes or to exploit the spatial distribution of the samples themselves in the classification.

In theory direct estimation could proceed by cross labeling all combinations between the five image variables. In practice, however, this results in computationally inhibitive in allocating memory for all unique combinations between the five image variables. (for this dataset stored in 8 bits, it would require 255^5 =1.0782 x 10^{12} combinations). Although the combinations could be evaluated over larger bin intervals, we selected an alternative method by evaluating the probabilities over each variable separately before their multiplication, assuming that they are conditionally independent. Using this approach a classification decision rule is stated as follows:

DR3: Choose j that maximizes:

$$P\{C_j\}\prod_{m=1}^{M} P\{X_l^m \mid C_j\} \qquad (10)$$

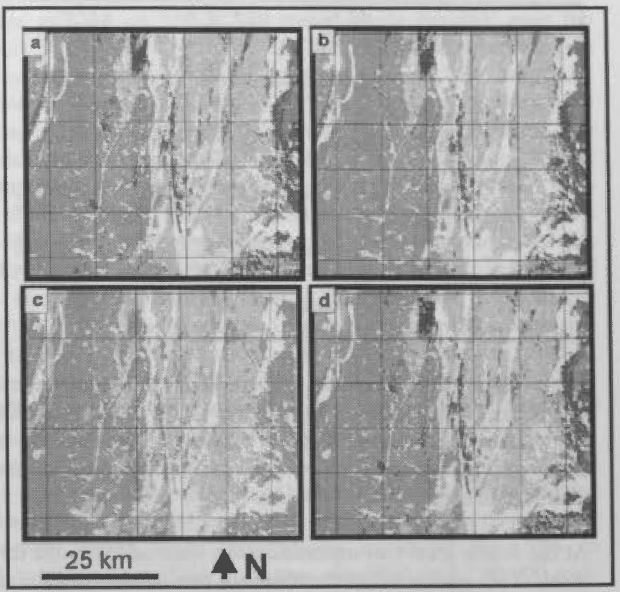


Figure 5 Results of stratified image classification experiments. a) classification bedrock units stratified on map compilation of lithotectonic domains; (b) classification bedrock units stratified on map compilation of lithologic assemblages; (c) non-parametric classification lithotectonic domains; (d) classification bedrock units stratified on classification lithotectonic domains (figure 5c).

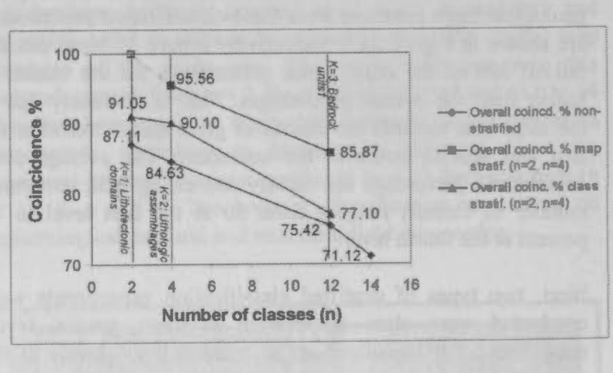


Figure 6. Coincidence percentages with geological map compilations for non-stratified Bayesian classification, for Bayesian classification stratified over map patterns at k=1 and k=2., for Bayesian classification at k=2 and k=3 stratified over non-parametric classification at k=1.

Because the lithotectonic domains are large contiguous units and the samples are evenly distributed, distance to samples was used as an additional variable to improve the discrimination potential of the multivariate dataset. By considering M variable, and computing coincidence percentages for all 20 combinations, it was found that the distance, potassium and magnetic grids provided the highest coincidence percentage (91%) improving the prediction of bedrock units with ca. 4 percent (figure 5c). The propagation of this classification result to levels k=2, and k=3 for computing priors, yielded an increase in coincidence with the geological map compilations of respectively 5.5 and 1.7 percent (figure 5d). Alternative non-parametric methods, such as the k-nearest neighbour classifier or algorithms based on artificial neural nets will be tested in the future to investigate if similar or higher classification performances can be obtained for the classifications at k=1 and k=2.

4. CONCLUSIONS

Initial results of a number of classification experiments suggest that classification performance can be improved by the estimation of prior probabilities at a particular level of a taxonomy that is conditioned by a more general level of the same taxonomy. In comparison to conventional approaches, the Bayesian stratified classification method provides mechanisms to introduce spatial data at more general levels in the classification, to which users have often better access or which can be easier derived through visual or automated image interpretation. An increase in classification performance may be obtained even when no additional data or visual interpretations are introduced, and spatial patterns associated to general levels in the taxonomy are also obtained by image classification. In addition the user can adapt algorithms and combination of image variables to each level within the hierarchy. This enhances the potential to adapt the classification methods to available map data and better exploit the intrinsic hierarchical structure of field knowledge.

REFERENCES

Gorte, B, 1998, *Probabilistic Segmentation of Remotely Sensed Images*, PhD thesis, ITC, 143 pp.

McDonough, M.R.M. McDonough, M.R.M., McNicoll, V.J. and Schetselaar, E.M., 2000, Age and kinematics of crustal shortening and escape in a two-sided oblique-slip collisional and magmatic orogen, Palaeoproterozoic Taltson Magmatic Zone, North-eastern Alberta, *Canadian Journal of Earth Sciences, special issue LITHOPROBE Alberta Basement Transect* 37(11), pp. 1549-1573.

McLachlan, G.LJ., 1992, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, 300 pp.

Schetselaar, E.M., Chung, C.F., Kim K., 2000. Classification of bedrock units in vegetated granite-gneiss terrain by the integration of airborne geophysical images and primary field data. *Remote Sensing of Environment* 71, pp. 89-105.

Strahler, A.H., 1980, The use of prior probabilities in maximum likelihood classification of remotely sensed data, *Remote Sensing of Environment* 10, pp. 135-163.

Switzer, P, 1980, Extension of linear discriminant analysis for statistical classification of remotely sensed satellite imagery, *Mathematical Geology* 12, pp. 367-376.

THE STUDY
ON APPLICATION MODE OF GIS ASSISTANT DEMARCATION SYSTEM

Guizhi Wang

The National Geomatics Center of China, No.1 Baishengcun, Zizhuyuan, Haidian District, Beijing 100044, China
wangmiaojie@163.com

KEY WORDS: Application Mode, Demarcation System, GIS, Boundary, Dispute Regions, Negotiation

ABSTRACT:

The administrative demarcation work among countries, provinces, cities are a complex, important and urgent task. First, author briefly talk about the main content, the basic task, the spatial behavior process on the administrative demarcation, and the application functions and the characteristics of the administrative demarcation system based on GIS in this article. Then, emphatically introduce the application mode of the administrative demarcation system. The presentation would be devided into three phases which is before, in and after negotiation. Finally, summarize the application benefit acquired by using the demarcation system.

Before negotiation, mainly prepare and analyze the materials for next turn of negotiation, study and set down multi-set, feasible plans for the negotiation. In negotiation, mainly inquire and search the materials for negotiation making use of the special negotiation tools, express demarcation standpoint, advance reason of demarcation standpoint, make and plot thematic map, apply the analysis functions and so on. After negotiation, clear up negotiation summary, white book and treaty, print attached drawing of treaty.

1. BRIEF INTRODUCTION OF ADMINISTRATIVE DEMARCATION AND GIS ASSISTANT DEMARCATION SYSTEM

1.1 The Main Content of Administrative Demarcation

On account of two power departments have different opinion on the boundary between adjacent administrative regions, both parties need to negotiate in order to make sure their confirmed boundary.

With the precondition of respect history, the two parties of demarcation put forward respectively own boundary claim-line and compare them. Coincident boundary segment reform confirmed boundary, inconsistent boundary segment reform dispute block. Represented in Figure 1:

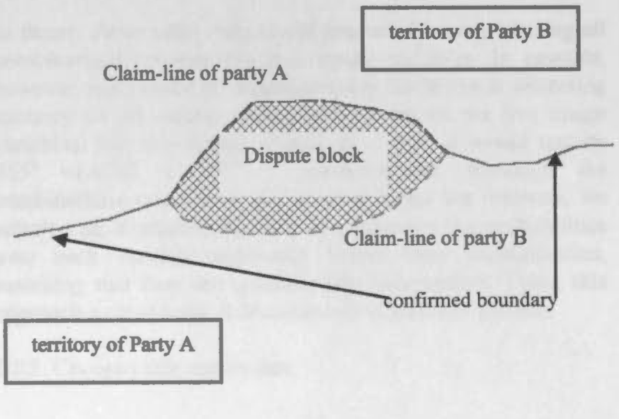


Figure 1: constitution of boundary dispute

During the negotiation, the two parties refer to a great deal of boundary materials and resolve these dispute blocks. Finally, the whole boundary line is reformed, which is accepted by the both parties.

1.2 The Basic Task of Administrative Demarcation

The basic task of administrative demarcation is that confirm the spatial trend of administrative boundary lines in the dispute regions and definitude the spatial popedom of the both parties on administrative management. The idiographic handling object is dispute block.

During the demarcation, the both parties negotiate about (consult with) the dispute regions one by one. In order to come into being same opinion, one side, need to analyze various resources of dispute region and put forward own claim with the precondition of striving for ultimate profit. On the other hand, need to analyze various historical materials, domination situation, law code and search the gists for own claim.

1.3 The Spatial Behavior Process of Administrative Demarcation

Definition: the spatial behavior sequence of leading administrative boundary to change in spatial trend constitutes the spatial behavior process of administrative demarcation. Among them, the spatial behavior process indicates the management, the analysis and the decision activities that demarcation professional have carried through. Strictly speaking, the spatial behavior sequence is the processes from the analysis of spatial data one by one to the visual expression of administrative boundary spatial trend. Simulating the spatial behavior process of administrative demarcation is virtually simulating the dispose steps that make up of the spatial behavior process of administrative demarcation. Because the task which every spatial behavior would complete is different and the operation which it would progress is independence relatively, the whole spatial behavior process of administrative demarcation may be divided different sub-processes in term of the spatial data content of administrative demarcation. Represented in Figure 2:

sub-process	operation
Inquiry of the demarcation information on the dispute region	Selecting the various materials of demarcation
Demarcation analysis on the dispute region	Contrast analysis of profit distribution on the different boundary
	Contrast analysis of the claim-line with the domination-line
	Contrast analysis of the claim-line with the historic boundary line
	Contrast analysis of disagreement between the historic boundary line
	Reason analysis of disagreement between the text illumination and the map
Visual expression of demarcation opinion on the dispute region	expression of the claim-line analysis conclusion and the conclusion gist (demarcation standpoint)
	expression of the trend analysis conclusion
	(advancing reason of demarcation standpoint)

Figure 2: Primarily breaking down on the spatial behavior process of administrative demarcation

1.4 The Characteristic and Application Function of GIS Assisted Demarcation System

Aiming at practice demand of the boundary negotiation, The land boundary system of negotiation and management has been designed and established. It includes the boundary database with hypermedia structure and the application functions of assistant boundary negotiation.

Realized integrative management of the fundamental geographical information, the demarcation material, the related treaty and law statute in the boundary belt. They are the data of multi-scale, multi-time and multimedia. Among them, the multi-media data type includes vector graph, raster image, real landscape photograph, aviation photo, three-dimensional digital elevation model, text material, video and radio material. The multi-scale data includes 1:10000-scale, 1:25000-scale , 1:50000-scale, 1:1000000-scale topographic maps. The multi-time data includes the treaty and attached drawing 100 years ago, the topographic maps in various time after establishing R.P.of China, the aviation photo in the near future, and so on.

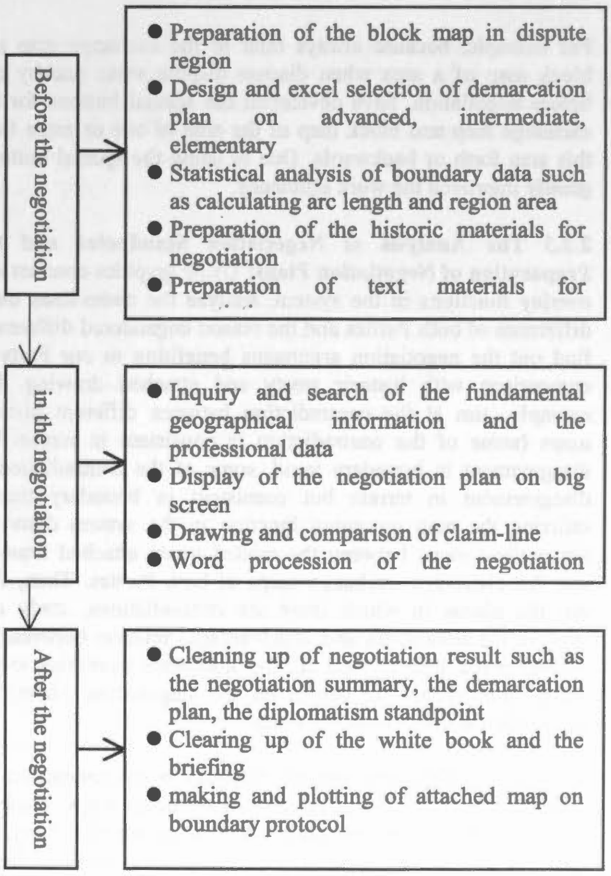
Develop the application functions assisted boundary negotiation such as the information inquiry functions based on "card" and "node" of hypermedia database, the statistical analysis functions, the demarcation analysis functions based on the digital maps considering area, the visual expression functions of demarcation opinion and claim, the manufacture functions of thematic map about demarcation and negotiation result, etc..

2. THE APPLICATION MODE OF DEMARCATION SYSTEM

2.1 The Three Phases of Appling Demarcation System

The demarcation system, it is demanded, meet the staggered requirement of boundary negotiation because it go along turn by turn. To meet the imminence requirement of the boundary negotiation, on the principle that the system tryout is began when the system development is going on, firstly establish a rudiment system that would be used and verified in the negotiation. Then the system functions are improved on and the system interfaces are perfected aiming at the questions found and the advice advanced in negotiation, supplement the new content at the same time which would be demanded in next turn negotiation.

Through the application and practice of the system, had groped a suit of effective application mode. It is divided into three phases. They resolve the problems before, in and after the negotiation. Before the negotiation, mainly prepare and analyze the materials for the negotiation, study and set down the plans for the negotiation. In the negotiation, mainly inquire and search the materials for the negotiation making use of the special negotiation tools, express the demarcation standpoint, put forward the reason of demarcation standpoint, make and plot the thematic map, apply the analysis functions and so on. After the negotiation, clear up the negotiation summary, the white book and the treaty, print attached drawing of the treaty. Represented in Figure 3:



- Preparation of the block map in dispute region
- Design and excel selection of demarcation plan on advanced, intermediate, elementary
- Statistical analysis of boundary data such as calculating arc length and region area
- Preparation of the historic materials for negotiation
- Preparation of text materials for

- Inquiry and search of the fundamental geographical information and the professional data
- Display of the negotiation plan on big screen
- Drawing and comparison of claim-line
- Word procession of the negotiation

- Cleaning up of negotiation result such as the negotiation summary, the demarcation plan, the diplomatism standpoint
- Clearing up of the white book and the briefing
- making and plotting of attached map on boundary protocol

Figure 3: 3-phases of system application mode

2.2 Clearing Up Information and Preparing Plans Before the Negotiation

2.2.1 The Preparation of Negotiation Arguments: Because the negotiation standpoints and the negotiation plans of our Party should be confirmed in according to the schedule of negotiation, need to clear up related fundamental geography data, the demarcation plans, the negotiation arguments, the negotiation documents and form special spatial database for the negotiation in notebook computer in order to take and refer in the negotiation.

The general negotiation arguments mainly include the agreement documents of both Parties, the claim-line exchange maps of both Parties, the block maps in dispute areas, historic treaty and attached drawing, the real landscape photographs, the newest aviation photo, the mere stone photographs, the video and radio boundary material, the reference text for negotiation, the negotiation outline and so on.

2.2.2 The Fast Referring of Boundary Information: To refer to the boundary information related to the negotiation as fast as possible, have developed a series of the special buttons in the system. They are the special buttons for anyone material of a dispute area and for several usual materials of anyone dispute area.

For example, the special button for the block map, for historic maps, for the real landscape photographs, for the aviation photo, for the mere stone photographs, for the video and radio material and for the reference text in one dispute area.

For example, because always refer to the exchange map and block map of a area when discuss dispute areas one by one before negotiation, have developed the special buttons for the exchange map and block map of the area of one or more from this area forth or backwards. Due to using the special buttons, greatly improved the work efficiency.

2.2.3 The Analysis of Negotiation Standpoint and the Preparation of Negotiation Plans: Using graphics contrast and overlay functions in the system, analyze the claim-lines trend difference of both Parties and the reason engendered difference, find out the negotiation arguments benefiting to our Party in comparison with historic treaty and attached drawing. For example, aim at the contradiction between different historic maps (some of the contradiction is consistent in terrain but disagreement in boundary trend, some of the contradiction is disagreement in terrain but consistent in boundary trend), utilizing the map grooming function in the system draw the comparison maps between the related treaty attached drawing and the claim-line exchange maps of both Parties. Then, find out the places in which there are contradictions, study and analyze the advantages and disadvantages relation between the various maps. Finally, find out the arguments benefiting to our Party from them and confirmed the negotiation plans, the negotiation standpoint of our Party.

Making use of the area analysis function in the system, on the claim-line exchange maps or the dispute block maps, combine the various boundary lines such as the claim-line of A Party, the claim-line of B Party, the practice domination line, the treaty line, the demarcation plan line and calculate the different areas. These are the gist of ensuring demarcation plan on advanced, intermediate, elementary so that hold measure and adjust plans at any moment in the negotiation.

Analyze and collect a great deal letter material and maps on any dispute region. Concluded negotiation outline serve for the important arguments which the head negotiation delegate hold in the negotiation.

2.3 Inquiring Information and Assisting Decision-making in the Negotiation

Need to search for the various fundamental materials, the plans and the arguments at any moment in the negotiation process. Simulate the demarcation plans based on the digital maps at the negotiation site and carry out assistant demarcation analysis.

2.3.1 Inquire and Search the Negotiation Material: Base on the negotiation course, inquiry the prepared fundamental material, the plans and the arguments at any moment. They include the treaty letter and attached drawing, the photograph of dispute region, the aviation photo, the three-dimensional digital elevation model, the text material, the video and radio material in order to explain and confirm our negotiation standpoint.

There are various inquiry functions based on the hypermedia database. For example, from node symbol to node content: from issue area symbol on 1:50,000-scale map to issue area card, continual to issue block map or issue area photography and so on; from node content to node content: from block map of certain issue area to three-dimensional digital elevation model of the same issue area and so on; from Menu to node content: from issue block item and three-dimensional item on menu to issue block map and three-dimensional digital elevation model, thus, every node can be take for start-node; from button to node content: When push any one of special purpose button bar for quick-inquiry, we can call-out the corresponding material.

2.3.2 Display the Negotiation Standpoint and Arguments of Our Party: Display at any moment the claim-lines of both Parties, the treaty line, the practice domination line, the plan line expressed our Party standpoint, the place of mere stone on the 1:50000-scale claim-line exchange maps of both Parties or dispute block maps. Refer to the multimedia information related with the negotiation. The condition of anyone area can be clear at a glance, very definitude and reality. It is easy for the both Parties to negotiate and discuss. It had brought into active play for the negotiation work group to analyze the boundary condition, discuss the negotiation plan, finally confirm the negotiation standpoint.

2.3.3 Simulate the demarcation plans at the negotiation site: The instance which both Parties contest endlessly or our Party is difficult to express own standpoint often occur in the negotiation process, therefore, using the system functions fleetly exhibit the standpoint of our Party with computer so that advanced the negotiation course.

For example, on the exchange maps or the dispute block maps, simulate the demarcation plans at the negotiation site using the special cartographic tools. That is drawing the demarcation standpoint of one Party or both Parties and carrying out explain, analyze, confirm, modify in the computer in order to understand opposite standpoint one another and manage to come into agreement.

2.4 Suming up the Result after the Negotiation

2.4.1 Sign of the Negotiation Conclusion: Using the drawing tools of the system, draw the negotiation conclusion on the digital maps. Some of them need to be printed out.

Mainly draw the boundary lines on digital maps. Besides set the type, the color, the size of the boundary line according to the request, can adjust and edit formed boundary line so that change the trend of boundary or combining or splitting boundaryline.

Can label the symbols of mere stone and the annotation on the digital maps, also intercalate their color, type(font), size.

Manufacture the thematic map about negotiation result and print out. It includes the layout of map-face, the selection of content, the grooming of map-outside and so on.

2.4.2 Updating of the Database: In any turn negotiation, the problem which would be resolved is different. Therefore, need to update database constantly and prepare the new information for next turn negotiation.

2.4.3 Process and Printing of the Negotiation Document: After the negotiation, need to process or print a lot of negotiation document such as the negotiation summary, the white book, the agreement, the demarcation plans, the diplomatism standpoint, the negotiation briefing, the report and so on.

2.4.4 Archives Management of the Negotiation Result: The conclusion of this turn negotiation would be clean up and archived by kind so that serve as the origin gist for next turn negotiation.

3. CONCLUSIONS

The system mainly serves national charge department of foreign affairs administration. Its economic benefit is hard to calculate, but its societal benefit is very markedness. The system had been used for the land boundary negotiation between both countries since October, 1996. Along with continuance improvement of the system functions, the system have taken a play more and more in the negotiation and gradually become the absolutely necessarily work system for boundary negotiation and management.

This is first special land boundary system in our country so far. It have become a pilot for wide application of GIS technology in boundary management in the future. Through studying and development of the system, have summarized a suit of advanced technology method and effective application mode. The lücubrating and practising had been done in spacial decision support system, hypermedia database, entity-relation model, multimedia information processing, visual expression.

Because the system was used to assist the land boundary negotiation, our country greatly improved the work efficiency and the work plane of preparating plan before negotiation, the reliability of expatiating argument in negotiation, the velocity and quality of clearing information after negotiation. Thus, picked up negotiation course and improved on our diplomatism negotiation work from traditional handwork operation to modern technology means.

Acknowledgments:

The author is grateful to the leaguers of this project.

References:

HaiQing Wang, 1998, Research for GIS-aided Administrative Demarcation Based on Spatial Behavior Process(graduate student paper).

GuiZhi Wang, 2000, Boundary Management Database based on Hypermedia, International Archives of Photogrammetry and Remote Sensing. Amsterdam. Vol. XXXIII, Part B4/3. Commission IV.1146-1152.

Jun Chen, 1995, the Concept and Question of GIS Special Data Model, the Integration and Application of RS, GIS, GPS, the Publish House of Surveying and Mapping

Fuling Bian, 1996, the Principle and Method of Geographical Information System, the Publish House of Surveying and Mapping.

MODELING MULTIDIMENSIONAL SPATIO-TEMPORAL DATA WAREHOUSES IN A CONTEXT OF EVOLVING SPECIFICATIONS

Maryvonne Miquel [a,b], Yvan Bédard [a], Alexandre Brisebois [a], Jacynthe Pouliot [a], Pierre Marchand [a], Jean Brodeur [a,c]

[a] Centre for Research in Geomatics Laval University, Quebec City Canada G1K 7P4 {maryvonne.miquel, yvan.bedard, alexandre.brisebois, jacynthe.pouliot, pierre.marchand }@scg.ulaval.ca,
[b] LISI-INSA de Lyon Bât B. Pascal - 7 av. Capelle, 69621 Villeurbanne Cedex France
[c] Natural Resources Canada, Center of Topographic Information, 21144 King Street West, Sherbrooke Canada J1J2E8 brodeur@RNCan.gc.ca

Commission IV, Working Group IV/1

KEY WORDS: Spatial On-Line Analytical Processing, Multidimensional Data, Data Modelling, Forestry

ABSTRACT:

In this paper, we study some problems linked to the integration of data in a spatio-temporal data warehouse. In many cases, the specifications of the data sets have evolved over time, especially when the observed period is large. Under those circumstances, data sources have temporal, spatial and semantic heterogeneity. In order to explore and analyse spatio-temporal data sets in a SOLAP (Spatial On Line Analytical Processing) application, we propose two approaches to model heterogeneous data in multidimensional structures. The first solution consists in a unique temporally integrated cube with all the data of all epochs. The second solution consists in creating a specific cube (data mart) for each specific view that users want to analyse. The final objective is to support geographic knowledge discovery through data exploration of detailed data for an epoch and of integrated comparable data for time-variant studies. Using a practical example in the field of forestry, we evaluate the implementation of these two models.

RESUMÉ :

Dans cet article, nous abordons les problèmes liés à l'intégration de données spatio-temporelles au sein d'un entrepôt de données. Dans de nombreux cas, notamment lorsque la période d'étude est relativement longue dans le temps, les spécifications des jeux de données évoluent. Dans ce cas, les données sont hétérogènes à la fois des points de vue temporel, spatial et sémantique. Afin d'explorer et d'analyser des jeux de données spatio-temporels dans une application SOLAP (Spatial On Line Analytical Processing), nous proposons deux approches pour modéliser ce type de données dans des structures multidimensionnelles. La première solution consiste à intégrer toutes les données dans un seul cube. La deuxième solution propose de créer un cube spécifique (un marché de données) pour chaque vue que l'utilisateur veut analyser. L'objectif final est de permettre l'extraction de connaissances géographiques par l'exploration des données détaillées associées à une époque et des études temporelles sur les données intégrées et comparatives. A partir d'un exemple pris dans le domaine de la foresterie, nous évaluons l'implémentation de ces deux modèles.

1. INTRODUCTION

On-Line Analytical Processing (OLAP) technology enables users to quickly analyse large sets of data. Hence decision-making is facilitated. OLAP systems are generally based on a three-tiers architecture including a data warehouse with integrated data, an OLAP server for the dimensional view and an OLAP client, i.e. a user interface for the rapid and easy exploration of data (Han, 2001). Similarly, SOLAP (Spatial On-Line Analytical Processing) systems are built to support the rapid and easy spatio-temporal analysis as well as the exploration of data according to a multidimensional approach typical of data warehouses (Bédard, 1997). This approach is comprised of aggregation levels supporting cartographic displays as well as tabular and diagram displays at various levels of detail. SOLAP systems provide the exploration and navigation tools required to analyze and explore spatial data, identify potential clusters, discover potential trends and build hypothesis (Rivest, 2001). However, building spatio-temporal data warehouses for SOLAP applications implies significant work of data integration especially when the data acquisition specifications have evolved over time. In this case, databases sources differ from one epoch to another not only in data coding and structures, but also in semantic contents. In order to permit temporal comparative studies, data must be integrated following a compatible set of temporal, spatial and semantic definitions. In this paper, we focus on some of the data warehouse supply difficulties in the case of conventional multidimensional models coupled to high heterogeneity issues. We expose a practical example in the field of forestry that considers forest maps of three 10-years periods elaborated following different acquisition specifications. We propose two solutions that are implemented and evaluated. At last, we discuss their advantages and disadvantages.

2. MULTIDIMENSIONAL MODELING OF SPATIAL DATA IN A DATA WAREHOUSE

To perform spatio-temporal analysis, multidimensional database modelling is very useful. Multi-dimensional views are produced when measures are analysed against the different dimension categories of a cube (Marchand, 2001). The most popular multidimensional model for relational OLAP (ROLAP) is certainly the star schema (Kimball, 1996). This model is centred on a fact table containing measures with related dimension tables, which characterize these facts. Each dimension has a number of attributes used for selection or grouping. A dimension is usually organized in hierarchies supporting different levels of data aggregation as well as multiple inheritances. The snowflake schema is a variant model where the hierarchies in the dimensions are explicit following normalized tables. Pedersen and al. (2001) analyse 14 multidimensional data models including star and snowflake models and show that these models do not support requirements such as multi dimension in each dimension, non strict hierarchies, handling change and time and handling different levels of granularity. They define an extended multidimensional data model for these requirements; their model is adapted for imprecise data. Some papers propose solutions for multidimensional structures with spatial data. Han and al. (1998) use a star/snowflake model to build a spatial data cube. They propose the idea of spatial measures with a method to select spatial objects for materialization. Papadias and al. (2001) use the star-schema with spatial dimensions and present methods to process arbitrary aggregations. In both cases, hierarchies of the spatial dimension are unknown at design time and are arbitrarily created by the user. Allowing change in aggregation hierarchies, Eder (2001) introduces a temporal multidimensional data model allowing the registration of temporal versions of dimension data. This solution is based on structure version and supports functions to transform data from one structure version to another. Espil (2001) and Hurtado (2001) also study multidimensional schemas with redefinition of aggregation hierarchies and heterogeneous schemas. In their solutions, they describe a new framework for modelling dimensions.

3. DATA INTEGRATION ISSUES

Spatial information is often organized according to spatial objects with geometric and descriptive attributes. For example, forest inventory information is usually vector-based and represents the boundaries of forest stands with their associate attributes. A forest stand is defined as a part of territory with homogeneous characteristics (species, height, age and so on). In Quebec, Canada, for each inventory, a new map is made by aerial photo-interpretation. Thus, each inventory is a complete new set of spatial data with no reference to the precedent set. Furthermore, from one inventory to another (usually every 10 years), the definition, the acquisition mode and specifications can change significantly, which make temporal, geometric and semantic integration as well as multidimensional modelling a difficult task. To integrate and model spatio-temporal data in a multidimensional structure with evolving specifications, several inter-related problems appear. We present the ones we have encountered once we have described our data sets.

3.1 The Data Sets Used

Three inventories of Montmorency forest compose the data set used in this study. Table 1 gives the characteristics of each inventory.

Year	Number of forest stands
1973	≈ 1 700
1984	≈ 2 400
1992	≈ 3 800

Table 1. Composition of inventories

The number of forest stands varies according to the evolution of the characteristics used for the photo-interpretation because the legislation changes. As specifications evolve over time, the definition and the number of attributes and classes vary from one inventory to another. Over the 20 years period covered by these 3 inventories, we have determined that only 12 % of data types can be compared over time and temporal analysis is hardly achievable.

For example, we present here the different classifications of the Age attribute for each inventory. In 1973, this attribute is composed of five qualitative classes (In regeneration, Young, Regular Mature, Irregular Mature, By stage). In 1984 classes have become quantitative data segmented in 20-years periods whereas in 1992 it is segmented in 15-years periods.

Descriptive data attributed to forest stand characterize the dominant trees. Even if the measured attributes of forest stands change over time, descriptive data can be grouped in time-invariant items such as Age, Density, Height, Diseases and slope.

3.2 Integration Problems

To support spatial and temporal analysis and exploration in the case of these forest inventories, we must cope with four issues: 1- each map is made independently of the previous one for the same territory, leading to delineating forest stands without relationship to the stands in the preceding epoch; 2- spatial referencing systems may change from a map to the other, sometimes without proper metadata, creating spatial matching problems 3- geometric heterogeneity caused by the temporal variation of forest stands definitions and 4- the descriptive heterogeneity caused by the evolution of the definitions and specifications.

Exploration of spatial dimension is the first requirement of our application. In order to supply time-invariant spatial reference, we propose to translate spatial data in an arbitrary regular tessellation. For each inventory, an overlay of the spatial vector-based objects and a fixed tessellation representation is produced. The cells become objects with invariant geometric characteristics and, as such, can be used as spatial reference. Figure 1 shows forest maps of the same area spanning the 3 studied epochs. At the top, we have initial forest stands as they were delimited in each inventory. Below we have the result of the overlay with a regular tessellation. The spatial reference is composed of regular cells as represented at the bottom of Figure 1. Each cell inherits the descriptive attributes of the dominant forest stand of a given inventory. With this solution, a new inventory is easy to add because it does not modify the existing data.

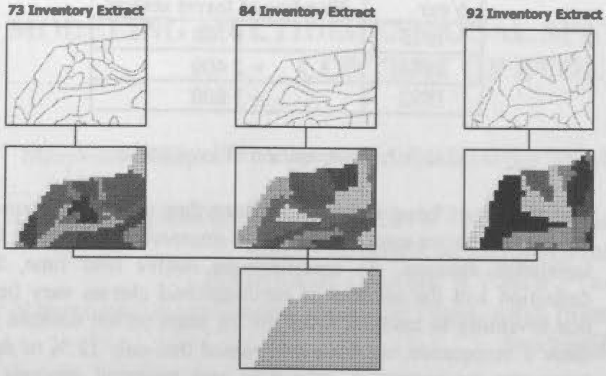

73 Inventory Extract
84 Inventory Extract
92 Inventory Extract

Figure 1. Data transformation using a regular tessellation

The hierarchy of the spatial dimension is derived from forest management rules. The hierarchy has three levels, by grouping cells in forest stands, then in compartments and in units or forest station (figure 2). A cell belongs to different forest stands depending on the inventory. The surface attributed to a cell is pondered in order to maintain correct surface measure of the forest stand. The results of different tests demonstrate that a grid of 20x20 meter is suitable to depict spatially the forest stands in Montmorency forest, it produces 162 000 cells to cover the territory of interest.

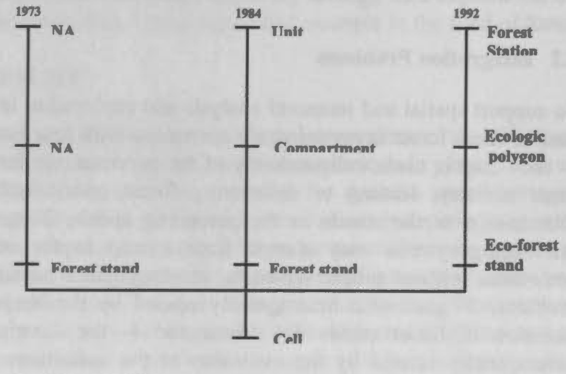

1973
NA
NA
Forest stand
1984
Unit
Compartment
Forest stand
1992
Forest Station
Ecologic polygon
Eco-forest stand
Cell

Figure 2. Hierarchy of the spatial dimension based on inventories

With this structure, users get access to the detailed data for each specific epoch i.e. 1973, 1984 and 1992 (it is especially useful for the most recent one). They also get access to evolution of the forest over the 20 of aggregated data. In other words, users access both detailed data for an epoch and to integrated comparable data for time-variant studies. So, we keep all the detailed data even if it is not homogeneous over time and it is up to the user to wisely navigate in the data set (or to the developer to restrain navigation).

Based on a multidimensional conceptual model of the data, data integration consists of building new data groupings by aggregating initial classes in a temporally compatible manner (Rebout, 1998). For each type of attributes, we define new generalized classes based on similar meanings. The aim of these new classes is to obtain comparable categories at the aggregated levels from one inventory to another. Each grouping is a level in a dimension hierarchy. At the lowest level (finest granularity) of a descriptive dimension, data is not time-comparable whereas the upper levels (coarser granularity) are time-comparable hierarchies with roll-up and drill-down functions. In Figure 3, the hierarchy of the dimension Age is shown. The attribute values are grouped in seven Age classes, and three Age Groups. At these two hierarchical levels, measures are time-comparable. For example the mature or out of age territory is identified as a category existing for all the inventories. Hence the evolution of mature or out of age areas during the last 20 years can be studied.

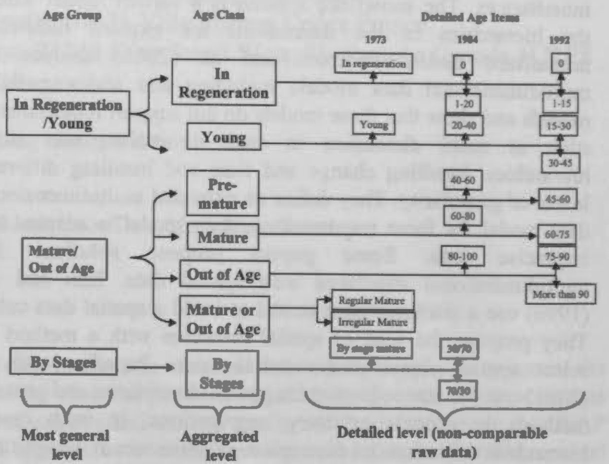

Age Group
Age Class
Detailed Age Items
1973
1984
1992
In Regeneration /Young
In Regeneration
Young
Mature/ Out of Age
Pre-mature
Mature
Out of Age
Mature or Out of Age
By Stages
By Stages
In regeneration
Young
Regular Mature
Irregular Mature
By stage mature
0
1-20
20-40
40-60
60-80
80-100
30/70
70/30
0
1-15
15-30
30-45
45-60
60-75
75-90
More than 90
Most general level
Aggregated level
Detailed level (non-comparable raw data)

Figure 3. Hierarchies of Age dimension

Figure 4 shows other descriptive dimensions with detailed data at the lowest levels. The hierarchy labels have been chosen according to the vocabulary used in forestry. One can note that the Density dimension is a single level dimension. The codification of forest stand density was maintained over time. Hence there is no integration issue in this case. In all other cases, these new aggregated classifications integrate the data semantically and over time. All the raw data are cleaned and encoded with these classifications in a common database. Then, different cubes can be designed and produced.

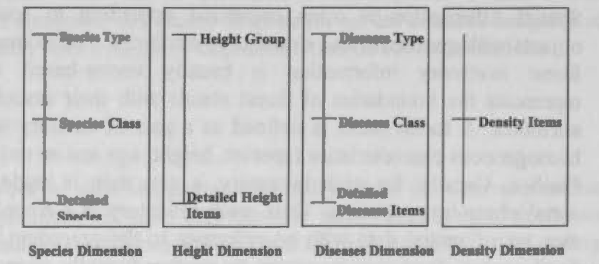

Species Type
Species Class
Detailed Species
Height Group
Detailed Height Items
Diseases Type
Diseases Class
Detailed Diseases Items
Density Items
Species Dimension
Height Dimension
Diseases Dimension
Density Dimension

Figure 4. Hierarchy of the Descriptive Dimensions

4. MULTIDIMENSIONAL MODELS

Using these spatial and descriptive dimensions, we design two models that show the link between facts and dimensions. In the star multidimensional schema, the fact table groups measures at the finest granularities of all dimensions. For forest inventories data, facts are measured areas with given characteristics (species, height, age…). In order to allow easy and complete exploration and analysis of this large set of data, we propose two approaches for modelling the spatio-temporal multidimensional data. The first solution is based on a unique temporally integrated cube. The second solution is a combination of 4 multidimensional structures that share dimension, one for each epoch and one for the 12% of data that are time-comparable at their finest level.

SOLUTION 1: One temporally integrated multi-epoch cube

In the first model, fact table is associated to a cell and to the descriptive dimensions (Figure 5). In this simplified schema, only Age, Height and Species are represented. The temporal variation is integrated within the descriptive and spatial dimensions.

This simple conceptual schema hides some implementation difficulties. The first lies in the storage of the large set of data resulting from the regular tessellation representation. Because of the homogeneous descriptive characteristics of a cell, the resulting multidimensional structure is sparse. The OLAP server will have to optimise the storage of this sparse structure.

The second difficulty lies in the management of the exploration paths (i.e. allowing a priori all dimension combinations). An additional user interface layer needs to be implemented to capture the user actions in order to control the exploration of the warehouse. Time varying queries using detailed descriptive dimensions granularities will be prohibited until allowed for upper granularities studies dealing with several epochs.

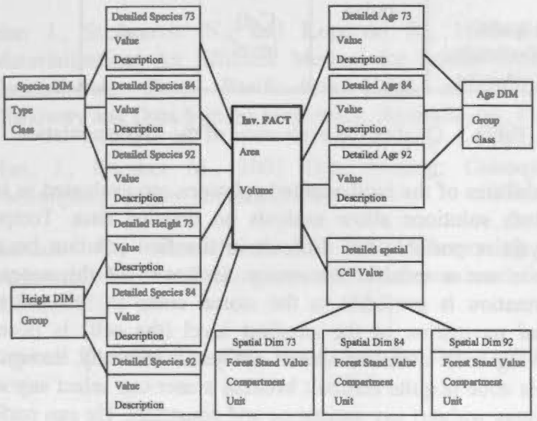

Detailed Species 73
Value
Description
Species DIM
Type
Class
Detailed Species 84
Value
Description
Detailed Species 92
Value
Description
Detailed Height 73
Value
Description
Height DIM
Group
Detailed Species 84
Value
Description
Detailed Species 92
Value
Description
Inv. FACT
Area
Volume
Detailed Age 73
Value
Description
Detailed Age 84
Value
Description
Detailed Age 92
Value
Description
Age DIM
Group
Class
Detailed spatial
Cell Value
Spatial Dim 73
Forest Stand Value
Compartment
Unit
Spatial Dim 84
Forest Stand Value
Compartment
Unit
Spatial Dim 92
Forest Stand Value
Compartment
Unit

Figure 5. A temporally integrated cube

SOLUTION 2: spatially-specific/ mono-epoch cubes and spatially-unified/multi-epoch cube

The second model is based on the fact constellation schema (Figure 6). In this case, multiple fact tables share dimension tables. In our application, a fact table is implemented for each of our 3 ten-year inventories and the measures are associated to forest stands. The descriptive dimensions include the Age, the Height and the Species. The finest granularities correspond to the detailed attributes for one inventory.

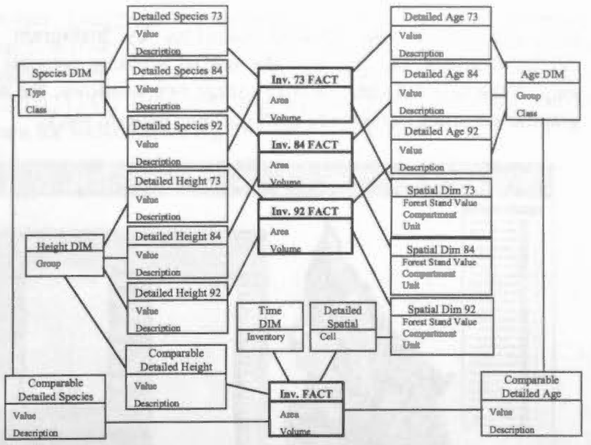

Detailed Species 73
Value
Description
Species DIM
Type
Class
Detailed Species 84
Value
Description
Detailed Species 92
Value
Description
Detailed Height 73
Value
Description
Height DIM
Group
Detailed Height 84
Value
Description
Detailed Height 92
Value
Description
Comparable Detailed Height
Value
Description
Comparable Detailed Species
Value
Description
Inv. 73 FACT
Area
Volume
Inv. 84 FACT
Area
Volume
Inv. 92 FACT
Area
Volume
Time DIM
Inventory
Detailed Spatial
Cell
Inv. FACT
Area
Volume
Detailed Age 73
Value
Description
Detailed Age 84
Value
Description
Detailed Age 92
Value
Description
Age DIM
Group
Class
Spatial Dim 73
Forest Stand Value
Compartment
Unit
Spatial Dim 84
Forest Stand Value
Compartment
Unit
Spatial Dim 92
Forest Stand Value
Compartment
Unit
Comparable Detailed Age
Value
Description

Figure 6. Fact Constellation Schema with 4 Multidimensional Cubes

The heterogeneity of dimensions hierarchies is explicit. A fourth fact table integrates the data which are comparable between the three inventories. This table shares a part of the hierarchical dimensions with the other fact tables. Measures are associated to single cells and one inventory (lowest levels of the spatial and time dimensions). This schema defines 4 multidimensional structures: one for each inventory and one for the integrated data. Each multidimensional cube associated to each inventory is based on forest stands as opposed to the multi-epoch cube which is based on the regular cells. The multidimensional cube with integrated data is unified both from a spatial and a time point of view. This schema optimises the data storage but requires the use of 4 multidimensional structures by the users. From a conceptual point of view, this model is a better representation of the complexity of the data and its intrinsic links.

5. PROTOPTYPE AND EVALUATION

The proposed solutions have been implemented using MS SQL Server and Analysis Services. These models are similar to the logical models previously presented. We defined each hierarchical dimension by one descriptive dimension for each inventory. All the multidimensional structures have been stored in a Multidimensional OLAP (MOLAP) database. In a MOLAP database, information (raw and aggregated data) is stored as series of multidimensional arrays. MOLAP databases respond faster to multidimensional queries as the structures are highly denormalized and stored in RAM.

The SOLAP prototype has been designed to support the spatio-temporal exploration of multidimensional structures using Proclarity and Intergraph Geomedia software components. A Visual Basic Application drives Proclarity functionalities to perform OLAP operators on descriptive data and drives Geomedia functionalities for spatial navigation and cartographic displays.

Figure 7 shows the interface of the SOLAP application. It is composed of two windows. The first one, on the right, is a classical OLAP interface with selection of dimensions to analyse, the navigation operators and the data displays. Using this interface, a user explores his data and obtains the representation of the selected measures by histogram and tabular displays. On the left, the spatial window presents the map of the selected data. Between these two windows, the same graphic semiology codes are used (colour, pattern, etc.).

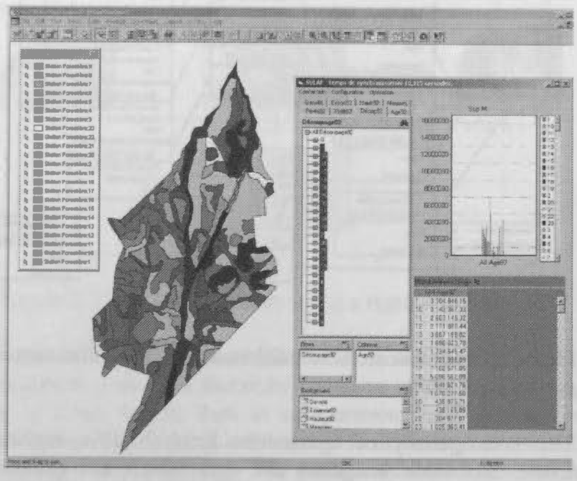


Figure 7. Interface of the SOLAP Application

Table 2 presents a quantitative evaluation of the two previous described solutions. The number of elements in the fact table depends on data organisation. The cube size is the storage size. Finally, performance gain is estimated by a ratio of pre-calculated data over stored data.

When we built cubes, we used an SQL-Server Analyse Services functionality to determine how many aggregations to create. The wizard adds aggregations until the performance gain reaches a specified percentage. We tried several values and chose a compromise between a percentage and a time running to optimise the structures.

	Elements of Fact Table	Cube Size (MB)	Performance gain
First Solution 1 temporally integrated cube	162 000	17.73	9 %
Second Solution 3 spatially-specific / mono-epoch cubes	From 1 700 to 3 800	From 0.15 to 0.65	100 %
1 spatially-unified /multi-epoch cube	496 000	18.5	50 %

Table 2. Quantitative evaluation of the implementation

With the first solution, the cube has a fixed number of elements into the fact table because the structure is stored by cell. By adding a new inventory, this value will not vary. Only the size of the multidimensional structure (actually 17.7 MB) will increase. The performance gain is very low (9%) indicating that almost any aggregation is pre-calculated. Here, the wizard failed to optimise structure in an acceptable delay.

With the second solution, four physical multidimensional structures are implemented. For the structure associated to each inventory, the number of table fact elements and the structure size depend on the number of forest stands because descriptive data are grouped according to forest stands. The fact table of the spatially-unified structure has an element for each cell and each inventory. Half the aggregated values are calculated and stored. In this solution, data are duplicated between the cubes associated to each inventory and the spatially-unified and time-integrated cube. More aggregations are pre-calculated, increasing the total storage size. The comparative table 1 displays the link between the size of the storage structure and the performance gain.

	Temporal Analysis	Spatial Analysis	Detailed data	Naviga -tion
First Solution 1 temporally integrated cube	Possible	Cell level	Yes	For expert
Second Solution 3 spatially-specific / mono-epoch cubes	Easy for the multi-epoch cube	Forest stand level	Yes	Easy and intuitive
1 spatial-unified/multi-epoch cube		Cell level		

Table 3. Qualitative evaluation of the implementation

Capabilities of the implemented solutions are evaluated in table 3. Both solutions allow analysis on detailed data. Temporal analysis is possible but difficult in the first solution because time is not a specific dimension. However all the necessary information is available in the stored cube. In this solution, spatial navigation in the smallest level (the cell) is possible allowing very abundant spatial analysis. Generally, navigation in this cube is quite difficult because a user can select any view he wants without any assistance and constraint. He can perform faulty analysis by comparing non-comparable detailed data and the result of complex queries depends on the order of the elementary queries. On the other hand, only possible temporal comparisons are implemented in the second solution, so the navigation is easy and secure in the multi epoch cube. To analyse data on one epoch, a user simply selects the cube associated to the inventory he wants to study. All the detailed data are easily available with classical OLAP operators. Users can navigate in hierarchical dimensions with different levels. These aggregated levels are used also in the spatially-unified cube in which temporal analysis can be performed.

6. CONCLUSION

In this work, we propose solutions to design multidimensional structures when source data sets have temporal, spatial and semantic variations. First, time-variant objects (forest stands) are converted in a fixed, invariant tessellation (cells) for proper spatial referencing. Then, new classifications of descriptive data are introduced by grouping initial detailed classes in comparable time-invariant super-classes. At last, several multidimensional structures are designed in order to explore data with a SOLAP application. Two opposed approaches have been implemented. The first solution consists in a temporally-integrated cube with all the data (detailed and aggregated), even if a lot of comparisons are meaningless at the lowest level. This solution is valid only if the SOLAP front-end tools guide data exploration by imposing navigation constraints. The design of the OLAP server is very simple but front-end tools need to be specifically designed. The second solution consists in creating a specific cube (data mart) for each specific view that users want to analyse. In this approach, navigation constraints are taken into account immediately in the design of the multidimensional structures. The analysis and the exploration of data can be done with standard front-ends tools for the descriptive part of data.

REFERENCES

Bédard, Y. 1997. Spatial OLAP, Vidéoconférence. 2ème Forum annuel sur la R-D, Géomatique VI: Un monde accessible, 13-14 novembre, Montréal.

Eder J., Koncilia C., 2001 Changes of dimension in Temporal Data, 3rd International Conference on Data Warehousing and Knowledge Discovery - Munich, Sept 3-7

Espil M, Vaisman 2001 A. Efficient Intentional Redefinition of Aggregation Hierarchies in Multidimensional Databases, DOLAP

Han J., Stefanovic N., and Koperski K., 1998 Selective Materialization: An Efficient Method for Spatial Data Cube Construction, Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining Melbourne, Australia, pp. 144-158.

Han, J., Kamber M., 2001 Data Mining: Concepts and Techniques, Morgan Kaufmann Pub

Hurtado C., Mendelzon A., 2001 Reasoning about Summarizability in Heterogeneous Multidimensional Schemas In Proc. ICDT'01

Kimball R., 1996 The Data Warehouse Toolkit. J. Wiley and Sons

Marchand, P., Bédard, Y. and Edwards, G., 2001. A hypercube-based method for spatio-temporal exploration and analysis. GeoInformatica. Submitted, in correction.

Papadias D., Kalnis P., Zhang J., Tao Y., 2001 Efficient OLAP Operations in Spatial Data Warehouses, Symposium on Spatial and Temporal Databases (SSTD), pp 443-459

Pedersen T.B., Jensen C., Dyreson C., 2001 A foundation for capturing and querying complex multidimensional data, Information Systems 26 383-423

Rebout, C., 1998. Adaptation d'une base de données pour une application SOLAP pour l'aide à l'aménagement intégré des ressources forestières, Université Joseph Fourier, Grenoble, France.

Rivest, S., Bédard, Y. & Marchand P., 2001, Towards better support for spatial decision-making: Defining the characteristics of Spatial On-Line Analytical Processing (SOLAP), Geomatica, the journal of the Canadian Institute of Geomatics.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support of the GEOIDE Network of Centers of Excellence via the DEC#2 project: Designing the technological foundations of geospatial decision-making with the World Wide Web. Maryvonne Miquel thanks the Center of Research in Geomatics (CRG), Laval University and especially Yvan Bedard and his team for welcome during one-year sabbatical.

INTEGRATED DYNAMIC MODEL FOR MULTI-RESOLUTION DATA IN THREE-DIMENSIONAL GIS

WENZHONG SHI[a], BISHENG YANG[a,b], QINGQUAN LI[b]

[a]Advanced Research Center for Spatial Information Technology, Department of Land Surveying and Geo-informatics,The Hong Kong Polytechinc University, Hung Hom, Kowloon, Hong Kong. lswzshi@polyu.edu.hk
[b]Research and Development Center for Spatial Information and Network Communication Wuhan Unviversity, Wuhan, Hubei, China 430079. bsyang@263.net

Working Group IV/1

KEY WORDS: Multi-Resolution, Integrated Model, Level-of-Detail, 3D GIS

ABSTRACT:

With the development of digital earth and cyber city, there is an increasing equipment on management and vitalization of integrated 3D model data and image data. How to improve the speed and visualization effect of three-dimesional (3D) model is a major research issue in 3D GIS. In order to improve the speed of 3D visualization, such as flying/walking through, many algorithms have been developed to reduce the data volume of 3D model, for instance, Level_of_Detail (Lod) algorithm. However, image data is also another important factor effecting the operation speed and reality of 3D model. Image data as the texture of surface of 3D model can enhance the reality of 3D model. For the reason that different graphic cards may have different restriction in the size of texture, the texture has to be managed very carefully. This paper proposes a new method to manage image data when a image is used as the texture for a 3D GIS environment. First, the data structure of multi-resolution texture model, which is developed to manage the large texture, is presented. Secondly, the algorithm that creates the multi-resolution texture model is developed, which is of vital importance for the texture mapping. Thirdly, the method of determining a proper texture resolution to map texture on the surface of 3D model based on viewer conditions is discussed. Finally, several experiments are illustrated.

1. INTRODUCTION

Image that is used as texture in a 3D GIS environment is of vital importance, the procedure that map image on the surface of model is texture mapping. Texture mapping is an efficient method to enhance the visualization effect of 3D model scene without adding the geometrical data of 3D vector model (Gahegan, 1999). Usually, the hardware of computer has a certain limitation for a certain texture size. For example, the OPENGL rendering engine limits the maximum texture size, which does not exceed 2048X2048. In 3D GIS system, the field of scene is usually very large, for instance, the terrain surface model can be of 4096X4096 grid size, and the terrain surface can be mapped with texture of 8000X8000 pixels, which exceeds the limitation of graphics hardware. In order to map the texture on the surface of 3D model, many techniques such as Mip-mapping have been developed to map the texture on the surface. It usually generates a serial of textures from the original texture, which called Mip-map pyramid (Williams, 1983), different layer of pyramid has different resolution. But Mip-map usually could not acquire better visualization effect when the texture size exceeds the limited size of graphics hardware, because the large texture has to be formatted in order to meet the requirements, many information of the texture is lost. In order to overcome the drawback of the limit of texture size, many graphics rendering systems such as (SGI□IRIS Performer 2.1,1999) developed Clip-Map method to overcome the drawbacks, Clip-map uses clip-size to clip the amount of texture data. By the clip-size, it can decide which layer texture data to be loaded from pyramid layers (Tanner, 1998). So when mapping texture on the model, only part of texture data is loaded into memory compared with Mip-map method. The difference between Clip-map and Mip-map is illustrated as figure. 1.

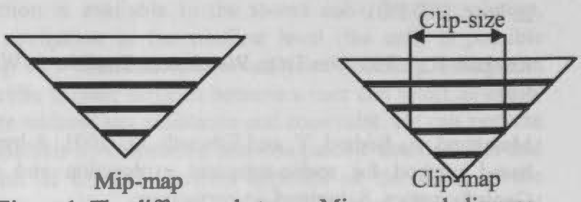


Figure 1. The difference between Mip-map and clip-map

In 3D GIS environment, a 3D city model for example, the terrain surface is large, and the texture of the terrain also exceeds the limit size of texture, moreover, there are so many buildings textures in 3D city model, So how to acquire better visualization effect when mapped the texture on the 3D model is a problem because of the limitation of graphics hardware and system memory. Because the scope of 3D scene is relatively large, only a fraction of the whole scene can be seen when flying/walking through the 3D scene. Moreover, it is evident that different parts of the 3D scene occupy different scope on the screen. For a terrain, supposing the terrain is divided into many tiles, the farther tiles occupy less area on the screen, for building, the farther buildings also occupy less area on the screen. Supposing the same resolution texture is used to map the texture on the different tiles and buildings, it is possible that the resolution of texture is lower for the near tiles of terrain and buildings and higher for the farther tiles of terrain and buildings. Therefore the visualization effect is worse. Supposing the higher resolution texture is used for the near tiles of terrain and building, and the lower resolution texture is used for the farther tiles of terrain and buildings. The visualization effect is thus better. It is evident that the different tiles, which are different distance to the viewer, are with different resolution of texture data. Thus we called the texture multi-resolution texture model. we proposed in this paper, that (1) multi-resolution texture model is developed to manage the large texture, which is useful for large texture mapping, (2) the method to decide the proper resolution of texture in different tiles based on viewer conditions. The frame of multi-resolution texture model is figure 1.

The rest of the paper is arranged as the following. Section 2 presented the data structure of multi-resolution texture model and the algorithm of creating multi-resolution texture, Section 3 propose the method of determining the resolution of texture, Section 4 illustrates several experimental results.

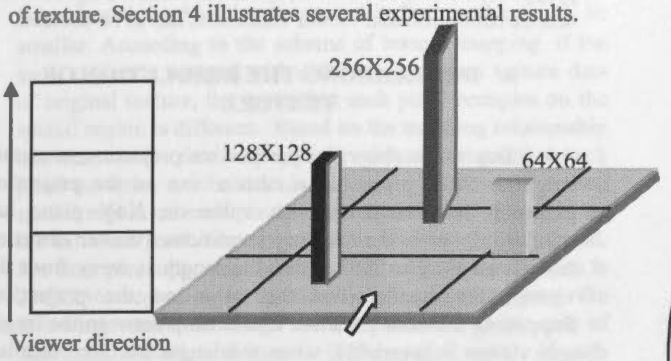


Figure 1. The multi-resolution texture model based on the viewing conditions

2. MULTI-RESOLUTION TEXTURE DATA MODEL

Based on the introduction of section 1, in order to obtain a better visualization effect and map arbitrary size texture on the surface of 3D model, the original texture has to be divided into many small tiles, moreover, so as to overcome the limited size of texture because of graphics hardware, the tile size does not exceed the limited size of graphic card. The regulations about how to divide large texture into many tiles are explained as the following.

Supposing I is the original texture, and $I_1, I_2, \ldots, I_n$ is the sub-area of original texture region. The following conditions are the constraint conditions when divide the original texture data into different small tiles□

1. $I = \bigcup_{i=0}^{n} I_i$
2. $I_m \cap I_n = \phi, m \neq n$
3. $I_m \subset I, m < n$
4. $I_m \subset \{x1, y1, x2, y2\}$

Where the condition 1 demonstrates that the original texture can be divided into n titles, condition 2 demonstrate that supposing that any two titles do not intersect with each other , condition 3 demonstrates all the small tiles consist of the original texture, the title is an element of, condition 4 illustrates that each title can be a given size spatial region, where $(x1, y1)$ is the left-down corner of the spatial region, and $(x2, y2)$ is the right-up corner of the spatial region. So , based on the above conditions, the original texture can be divided into many titles. Moreover, each title has a certain spatial region, which can be used to decide proper texture mapping. In order to mange the texture data efficiently, viable data structure has to be developed to mange them. This is the foundation of texture mapping and texture data management.

2.1 Data structure of multi-resolution texture model

Based on the above analysis, the texture data includes two types, one is the texture data of the terrain surface, the other is the texture data of building surface. Because of the fact that the texture size of building surface is relatively small, it need not to be further divided. The different resolution layers texture of building surface textures can be generated from original texture. However, the large texture of terrain surface has to be divided into many titles. In order to ensure efficiency of texture mapping, the viable data structure has to be developed to manage the texture data of building surface and terrain surface. Table.1 illustrates the data structure of texture data, which is the foundation of retrieval of texture data and texture mapping.

Table 1. Data structure of multi-resolution texture

Image$_1$	Image$_2$	....	Image$_n$

type	space region	layers	texture size	texture data
1	x1,y1,x2,y2	0	1024X1024	01010101.
1	x1,y1,x2,y2	1	512X512	01010101.
... ...	....		....	....
0	x1,y1,x2,y2	3	64X64	01010101.

Table. 1 illustrates the data structure of texture data, where 'type' represents the texture type, 0 is building texture data and 1 is terrain texture data. 'Space region' represents the surface scope that the texture data will be mapped on. 'Layers' represents the texture layers of original texture data generates. 'Texture size' represents the width and height of original texture data. 'Texture data' is the data of original texture data. 'Image$_i$' represents the texture index of each tile or building surface, which is used to retrieve the texture data of different tiles of terrain and building textures.

2.2 The Algorithm of creating multi-resolution texture data

So as to creating the multi-resolution texture data, the size of each tile has to be decided, once the size of each tile is decided, the multi-resolution texture data can be created from the original texture data. The main step of creating multi-resolution texture data is presented as below.

2.2.1 Steps of creating multi-resolution texture data

The method of dividing of large texture is the foundation of multi-resolution texture mapping. Based on the data structure developed, the following algorithm is developed to divide the large texture into tiles.

Step1 Read the original texture data and judge the size of original texture (width and height of original texture)

Step2 According to the minimum size of tile, calculate the number of tiles,
Step3 Based on the criterion of partition, get the texture data of each title and store them into the data structure,
Step4 Create the pyramid model of each tile,
Step5 Storing the texture data of each title or building surface according to data structure□
Step6 End.

Figure 2 illustrates the results of original texture, which is divided based on the algorithm. After the original texture is divided into many small tiles, the index of each tile can be created according to the spatial region of each tile is mapped. When mapping them on the surface of 3D model, according to the distance and direction with the viewer, the proper resolution can be decided based on the direction and distance with the viewer, the criterion about how to decide the resolution of texture is proposed in Section 3.

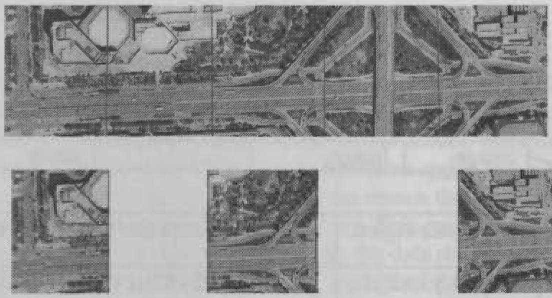


Figure 2. The dividing result of original texture

2.2.2 Mapped Texture on the surface model

Texture mapping is a procedure that mapping the color region to geometrical region, which can be represented as $\xi : C \rightarrow R$, so each vertex of the 3D model has a texture coordinate. The procedure of texture mapping includes two steps (Hill, 2001)□

- Loading the texture data of spatial region,
- Determine the relationship between color space and geometrical space (X,Y,Z) and (U,V) □so each vertex can has a texture coordinate (u,v) □

Based on the analysis above, in order to map the texture on the surface of 3D model, the following four steps have to be finished, illustrated as figure 3.

- Based on the spatial index of each tile and viewer spatial position, decide the resolution of texture data of each tile,
- Load the texture data of each title into memory,
- Calculate the relationship between texture data and surface region of model,
- Mapping texture on the surface according to the relationship.

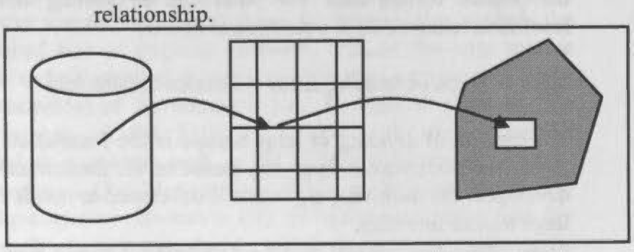


Figure 3. The procedure of mapping texture on the surface of model

For the reason that the limitation of system memory of computer, it is impossible that load all the original texture data of each tile into memory. In order to ensure the efficiency of mapping texture on the surface and the speed of 3D visualization, it is necessary to decide the proper resolution of texture data in each tile and different building surface textures. According to the analysis in Section 1, the tiles of terrain and the surface of building are far away from the viewer, they can be mapped with lower resolution texture data, and the tiles of terrain and the surfaces of building are close to viewer, they can be mapped with higher resolution texture data. According to the facts above, the texture resolution of different tiles and surfaces of buildings are related with viewer conditions, so in order to decide the proper resolution of different tiles, the relationship between texture resolution and viewer conditions has to be decided.

3. DETERMINING THE RESOLUTION OF TEXTURE

According to the theory of perspective projection, a certain length in XoY plane has a certain size on the projection plane. It is evident that the value on XoY plane has relationship with the viewing conditions (viewer direction and distance). The farther that the length is away from the viewer, the smaller that the value on the projection. Supposing the transformation relationship between the length and viewer is invariable, when the length becomes smaller, it's projection value on the projection plane also will become smaller. If the changes of the projection value does not exceed a limited value (for example, the minimum resolution of human eyes), the smaller length can be used to replace the larger length. The same principle also applies to texture mapping. For example, when we mapped texture on a certain surface, different resolution texture data can be generated from the original data. If different resolution texture data of original texture is selected as the texture data to be used as the texture of the same tile or surface, the changes in color space will be generated. According to the principle above, if the changes in color space do not exceed a limited value, the lower resolution texture data can be used as the texture of the surface or the tile. Compared with higher resolution texture data, the lower resolution texture data has less data amount, it needs less memory usage and also costs less time in texture mapping. So it is necessary to explore the relationship between the resolution of texture and viewer conditions (direction and distance). Next, the criterion of determining the resolution of texture data based on viewing conditions will be proposed.

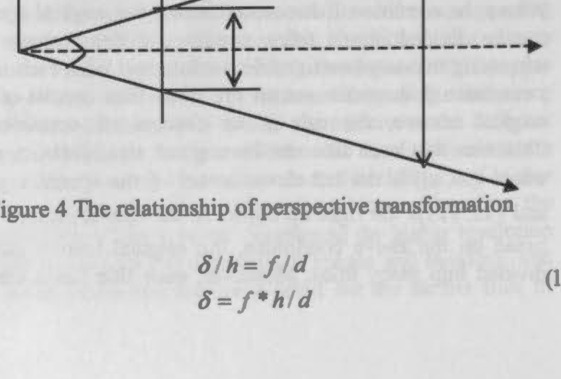


Figure 4 The relationship of perspective transformation

$$\delta / h = f / d \qquad (1)$$
$$\delta = f * h / d$$

where f is the distance from the viewer to the projection plane, and d is the distance between the viewer and projection length, h is the projection value on XoY plane.

Figure.4 illustrates the relationship between the size in XoY and the value on the projection plane based on transformation of perspective projection. According to the theory of perspective projection, when the direction of the value in XOY is parallel with projection plane, the value on the projection plane is maximum.

Equ.1 illustrates the relationship between viewer conditions and projection value. If the scale between projection plane and screen is 1:1, the δ is the pixel numbers on the screen. It is evident that when the distance d is farther, the δ will be smaller. According to the scheme of texture mapping, if the same region is mapped with different resolution texture data of original texture, the space that each pixel occupies on the spatial region is different. Based on the mapping relationship between texture data and spatial region, the space that each pixel occupies can be calculated, so the difference between spatial region of each pixel occupies, which is generated by the different resolution texture data, also can be calculated. For example, supposing the size of the original texture data is 256X256, according to the scheme of texture mapping, the following resolution image can be generated as the size of 128X128, 64X64, and so on. If the spatial region of each pixel occupies is ξ_0 when the region is mapped with original texture data (256X256), the spatial region of each pixel occupies is $2\xi_0$ when the same region is mapped with the resolution texture data (128X128). Thus the difference between the spatial region is ξ_0. According to the Equ.1, the projection value of ξ_0 on the projection plane can be calculated. The calculation result is the changes in the color space, which will have an impact on the 3D visualization effect. So we can think the impact on the visualization is brought by using different resolution texture data on the same region.

Based on the above analysis, the changes, which is generated by using different resolution texture data on the surface, can be calculated by Equ.1. Supposing the changes do not exceed a limited value, the lower resolution texture can be used to replace the higher resolution one. According to the data structure of multi-resolution texture model, each tile or building surface perhaps has several different resolution textures. For each tile, supposing there are $\{\zeta_1, \zeta_{2,} \bullet\bullet\bullet, \zeta_n\}$ different resolution texture data. And the threshold of limited value is τ, the resolution of current texture data on the tile is ζ_i. Thus so as to decide which resolution of texture data is proper for the spatial region, the following algorithm is developed to decide the proper resolution of texture data for the tile or surface.

Algorithm Decide_Texture_Resolution_Of_Tile( )
Step1 According to the relationship between spatial region and different resolution texture data, the length in space that each pixel of different resolution texture occupy can be calculated. Supposing the length sets of each pixel of different resolution occupies is $\{h_1, h_{2,} \bullet\bullet\bullet, h_n\}$,
Step2 According to the relationship between texture data and spatial region and the resolution of texture data ζ_i, the length in space that each pixel occupies (h_i) can be calculated, Based on Equ.1, δ_i (the projection value of h_i on the projection plane) can be calculated. $\delta_i = f * hi / d$,
Step3 Based on the threshold of limited value τ, so the optional value (δ) on the projection plane meet the condition. $\delta_i - \tau \le \delta \le \delta_i + \tau$, so length in space that each pixel occupies also can be calculated, the value of the length meet the condition. $(\delta_i - \tau) * d / f \le h \le (\delta_i + \tau) * d / f$,
Step4 Based on the result of **step 3** and length sets, the proper resolution can be selected from length sets $\{h_1, h_{2,} \bullet\bullet\bullet, h_n\}$. The criterion of selection optional resolution texture data is $|h - hi| = \min, (1 \le i \le n)$,
Step5 Based on Step 2,3,and 4, the resolution of texture data of each tile or surface can be decided,
Step6 End.

Based on the algorithm, the resolution of texture data of each tile and surface of building surfaces can be determined. It is evident that the resolution of texture is lower for the farther surface and is higher for the closer surface. The better visualization effect can be acquired based on the scheme of multi-resolution texture model.

4. EXPERIMENTS AND ANALYSIS

Based on the multi-resolution texture model and algorithm we proposed in this paper, the multi-resolution texture model is implemented in our system Spaceinfo and many experiments are finished in SpaceInfo, and many experimental results are as follow. Figure 5 is the result of mapping texture on the terrain surface based on pyramid texture model, and Figure 6 is the result of multi-resolution texture model. The size of the original texture data is 4096X4096 and the tile size is 256X256. Because of the limitation of graphics hardware, the original texture has to formatted as 2048X2048 resolution based on pyramid texture model.

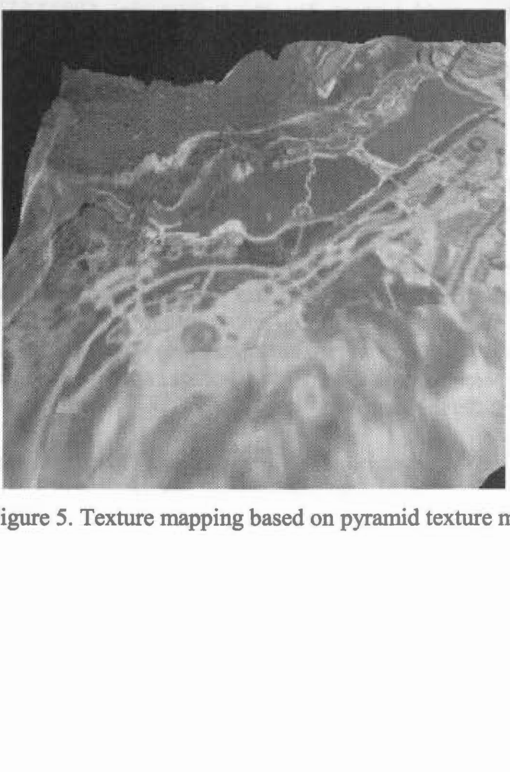


Figure 5. Texture mapping based on pyramid texture model

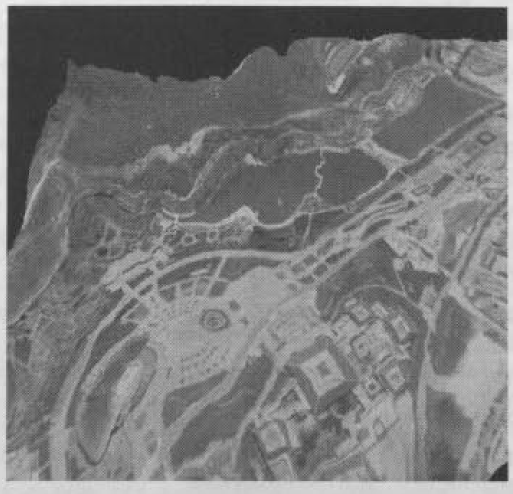


Figure 6. Texture mapping based on multi-resolution texture model

Based on figure 5 and 6, it is easy to see that the multi-resolution texture model can acquire better visualization effect than that of pyramid texture model. In order to test the performance of different tile size.Table2 lists the time performance of different tile size and pyramid texture model.

Table 2 Time performance of different tile size and standard resolution

Texture Size	512X 512	256X 256	128X 128	64X 64	Pyramid texture
6992X5328	19.09s	18.24s	17.18s	18.56s	19.72s
2048X2048	2.68s	2.78s	2.65s	2.64s	4.16s
1024X1024	0.86s	0.82s	0.81s	0.92s	0.70s

Table 2 lists the time of texture mapping of different tile size and pyramid texture. For the texture size 6992X5328, in order to use pyramid texture model to map texture, it has to be formatted so as to meet the requirements of graphics hard. The time span is from texture loading to map the texture on the surface. According to the result of Table 2, it is easy to see that the tile size is 128X128, the best time performance can be acquired. Thus we use the 128X128 tile size as the minimum unit when dividing a large texture.

Pyramid texture model has some advantages such as the simple data management, easy texture mapping and so on, but the size of texture is limited by the graphics hardware. Multi-resolution texture model can map arbitrary size texture, it overcome the limitation of graphics hardware. Table 3 lists the difference between them.

Table.3 Comparison between pyramid texture model and multi-resolution texture model

	Pyramid Model	Multi-resolution model
Data management	Simple	Complicated
Hardware limitation	Has limit	No limit
Texture coordinate	Calculated one time	Each tile is calculated
Edge handle	No extra handling	Extra handling
Visualize Effect	Worse	Better
Programming	Simple	Much difficult

5. CONCLUSIONS

Texture data is of vital importance for a 3D GIS. It can enhance the reality of 3D model without adding geometry data of model. The method of texture mapping and the management of texture data are important for a 3D GIS system. In this paper, we propose multi-resolution texture method to map large texture on the surface of 3D model, such as terrain model. A viable data structure for the management of texture data is presented, and the algorithm is developed to decide the proper resolution texture data. Moreover, the method is implemented in our software SpaceInfo and acquire many better results. Compared with the method of pyramid texture model, multi-resolution texture method has several advantages.
Firstly, it can map arbitrary size texture on the 3D model. Secondly, A new method for management texture data and map texture is proposed. Thirdly, The system memory of this method occupies is less that that of pyramid texture model. Fourthly, the fast speed of texture mapping, the resolution of texture data is related with viewer conditions, so the visualization effect of 3D model is much improved.

Although multi-resolution texture can acquire better visualization effect, how to improve the speed of 3D roaming further needs to be explored in the next step. For the reason that the resolution of texture data is related to viewer conditions, the texture data in the memory has to be exchanged timely when flying/walking through 3D model. Thus a high efficiency exchange schemes has to be developed, which is the next research.

ACKNOWLEDGEMENTS

The work described in this paper was substantially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No.: PolyU 5050, 3-ZB40) and The Hong Kong Polytechnic University (1.34.9709), and partially supported by a grant from the China National Natural Science Foundation (No.69833010).

REFERENCES

Blow, J., 1998. Implementing a Texture Caching System. *Game Developer*. http://www.gdmag.com

Gahegan, M., 1999. Four barriers to development of effective exploratory visualization tools for geosciences, *International*

Journal of Geographical Information Science, 13, pp.289-309.

Hill, Francis S, 2001. *Computer graphics: using OpenGL.* Upper Saddle River, N.J.: Prentice Hall.

IRIS Perform 2.1 API references, 1999. http://reality.sgi.com/performer/perf-99-07/0103.html.

Tanner, C., Migdal, C.J., 1998. The Clipmap: A virtual mipmap. In: Computer graphics (*SIGGRAPG'98*).

Williams, L., 1983. Pyramidal parametrics. In: computer graphics (*SIGGRAPG'83*). pp 1-11.

PREDICTABILITY OF INDIVIDUAL MASS MOVEMENTS FROM SPATIAL DATABASES FOR HAZARD MAPPING

Andrea G. Fabbri[a], Chang-Jo F. Chung[b]

[a] ITC, Hengelosestraat 99, 7500 AA Enschede, The Netherlands
E-mail: fabbri@itc.nl
[b] Geological Survey of Canada, 601 Booth Street, Ottawa, Canada K1A 0E8
E-mail: chung@gsc.nrcan.gc.ca

Commission IV, WG IV/1

ABSTRACT

Predictive models of landslides hazard have been developed from databases that provide the spatial information to characterize the typical geomorphologic settings of the trigger areas of specific types of mass movements. The integration of several layers of spatial data such as DEM derivates, lithology, land use and the distribution of groups of landslide trigger zones belonging to particular dynamic types and the known time interval allows to apply analytical strategies to validate the results of predictions. Two ways of evaluating prediction patterns are: (i) the analysis of prediction-rate curves, and (ii) the stability-sensitivity characterization of the prediction images.

Because the problem of generating a prediction is less critical than the one of understanding its significance, the following strategy is applied in this contribution. First a separation of "good" predictor layers is obtained by an evaluation technique based on empirical validation. Then, the separation of "good" predictor mass movements is generated by iteratively excluding one mass movement from the prediction model. This is to validate its predictability by the remainder of the mass movements used to generate the prediction. The result of this strategy is a measure of how much each mass movement contributes to a prediction. A study of the "bad" predictors provides criteria to understand the information content of the database. Such a process represents a form of data mining that allows not only to extract information from very large databases but also to understand the significance and usability of what we have as available data.

Two case studies in central Italy and Portugal are using hazard mapping databases of different resolution and geological complexity. These application areas are used to discuss the interpretation of predictability of the mass movements and the consequent hazard susceptibility representation.

WG IV / 2:

FEDERATED DATABASES AND INTEROPERABILITY

PREDICTABILITY OF INDIVIDUAL MASS MOVEMENTS FROM SPATIAL DATABASES FOR HAZARD MAPPING

Andrea G. Fabbri[a], Chang-Jo F. Chung[b]

[a] ITC, Hengelosestraat 99, 7500 AA Enschede, The Netherlands
E-mail: fabbri@itc.nl
[b] Geological Survey of Canada, 601 Booth Street, Ottawa, Canada K1A 0E8
E-mail: chung@gsc.nrcan.gc.ca

Commission IV, WG IV/2

ABSTRACT

Prediction models of landslides [illegible] generated from databases that provide the spatial information to characterize the typical geomorphological settings in the [illegible] areas of [illegible] [illegible] specific spatial [illegible] to apply statistical [illegible] to validate the results of predictions. Two ways of [illegible] (i) [illegible] and (ii) the stability [illegible] of the prediction patterns.

Whereas the mechanics of generating a prediction is less complex than the task of understanding its significance, the following strategy is applied in this contribution. Each comparison of "good" predictive maps is generated by [illegible] a technique based on empirical likelihoods. Then, the separation of [illegible] mass movements is generated by [illegible] including one mass movement from the prediction model. This is to validate its predictability by the remainder of the mass movements used to generate the prediction. The [illegible] how much [illegible] prediction. A [illegible] criteria to [illegible] of the database. Such a process represents a [illegible] that allows not only to extract information from very large databases but also to understand the significance and usability of what we have as available data.

Two study areas in central Italy and Portugal are using hazard mapping databases of different resolution and geological complexity. These applications [illegible] are used to discuss the [illegible] of predictability of the mass movements and the subsequent hazard susceptibility representations.

INTEGRATION OF SPATIAL DATA WITHIN A GENERIC PLATFORM FOR LOCATION-BASED APPLICATIONS

Steffen Volz, Jan-Martin Bofinger

University of Stuttgart, Institute for Photogrammetry, Geschwister-Scholl-Str. 24D, 70174 Stuttgart -
(steffen.volz, jan-martin.bofinger)@ifp.uni-stuttgart.de

Commission IV, WG IV/2

KEY WORDS: Federated Infrastructure, Location-Based Services, Multiple Representations, Spatial Data Integration, XML/GML, Conflation, Matching Conflicts

ABSTRACT:

Within the last few years, a new kind of applications has evolved: location-based services are on their way to become one of the leading powers within the field of information technology. At the University of Stuttgart, a research project called NEXUS has been initiated to develop an open and global platform supporting all possible types of mobile, location-based information systems. A federated infrastructure will be built up that manages a detailed model of the real world, which is called the Augmented World Model. Thus, an integrated view on the distributed data sources and services that have to cooperate within NEXUS as well as a standardized interface can be offered to the applications. In order to realize the generic approach of NEXUS, two crucial prerequisites have to be fulfilled: First of all, different data providers have to be able to integrate their data into the Augmented World Model. For this reason, a schema integration takes place that maps the object classes of existing data models onto the classes of the Augmented World Schema. Second, corresponding object instances within the data of the Augmented World Model must be matched. Thus, adjoining or overlapping data sets can be merged in order to derive one complete and semantically enriched spatial representation meeting the needs of different location-based applications. Therefore, appropriate conflation algorithms are being developed. The paper briefly introduces the idea of NEXUS and describes the architecture and the data model of the platform. Then, the approaches for spatial data integration are explained in detail.

1. INTRODUCTION

Location-based services can be defined as applications which know the positions of their users and provide information and services accordingly on a mobile computer device. Thus, the main criterion for selecting information is the spatial context a user is in. Today, only isolated applications like car navigation systems or mobile tourist guides are offered by different providers. They are restricted to certain data formats, certain mobile communication techniques and certain sensor technologies. In order to provide an open and global platform on top of which various types of location-based applications can be developed, NEXUS aims at integrating all the different components.

The quality of a global location-based information system like NEXUS especially depends on the availability of detailed and manifold spatial data. For this reason, heterogeneous data sources that have been acquired by different providers in different underlying data models and scales must be integrated into the NEXUS platform. In order to achieve this goal, a global data model called the Augmented World Model (AWM) has been developed that contains all objects which are relevant for location-based applications. For the integration of heterogeneous spatial data within the NEXUS system, object classes of existing data models have to be mapped onto the classes of the AWM. Thus, a uniform view on the data as well as a standardized interface can be offered to the applications. As a prerequisite for a further usage of heterogeneous data sources within NEXUS, multiple digital instances of one real world object must be identified. Therefore, matching algorithms are being applied in order to merge overlapping or combine adjoining data sets.

This paper comprises an overview of the NEXUS platform in section 2. Section 3 describes how heterogeneous road data sources can be integrated into the Augmented World Model. In section 4 the conflation approaches within NEXUS are presented.

2. THE NEXUS PLATFORM

The NEXUS platform can be described as an open and distributed environment for location based applications. It consists of a federated architecture which manages services and information resources and provides a consistent view for the applications: the Augmented World Model.

2.1 Architecture of the Infrastructure

The components of the NEXUS infrastructure are organized in three tiers (see Figure 1): applications, federation and services. Basically, the architecture of NEXUS can be compared to the one of the World Wide Web. Just like Web servers, NEXUS data servers can be provided by anyone and only have to be registered in a certain directory that stores the address and the properties of an information resource. NEXUS applications are similar to Web applications like browsers or applets. However, the intermediate tier of the NEXUS architecture, the federation, has no counterpart in the WWW. The federation organizes the response to client queries within the distributed NEXUS architecture.

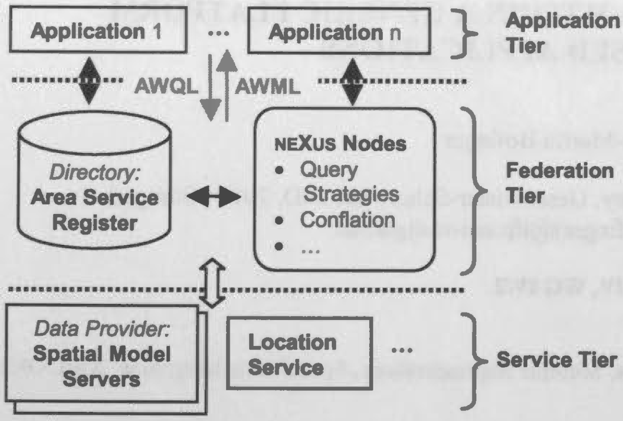


Figure 1: Architecture of the NEXUS platform

2.1.1 Application Tier: NEXUS applications are typically running on small mobile devices which are able to determine their position via appropriate sensors. When they are performing queries, they contact the NEXUS nodes of the federation layer using wireless communication technologies like WLAN or UMTS. Queries are formulated in the XML-based Augmented World Query Language (AWQL). AWQL contains simple spatial and attributive predicates. Usually, the clients receive the results of their queries in the Augmented World Modeling Language (AWML). AWML is also defined in XML and allows to serialize the objects of the AWM.

2.1.2 Federation and Service Tiers: Any application request is directed to a *NEXUS node*. On each node, a federation component is running. The NEXUS nodes are responsible for developing a strategy to respond to application requests. They distribute the queries on the appropriate components of the Service Tier, namely the Spatial Model Servers and the Location Service. A *Spatial Model Server* stores and manages the static spatial data (like buildings, streets, etc.) of one or more geographical areas. It also provides geoprocessing functionalities. The *Location Service* deals with the management of the user positions. Since huge amounts of location data is updated frequently, the properties of standard databases are inadequate. Thus, we have developed a main memory based solution for efficiently storing and retrieving the position data of mobile users (A. Leonhardi & K. Rothermel, 2001).

In order to create a strategy to answer application requests, a NEXUS node must first identify, which spatial data have to be used for query processing. Therefore, it addresses the *Area Service Register* which can be understood as a metadata repository storing general information about spatial data sets (like spatial extent, level of detail, stored object classes etc.) and the addresses of the Spatial Model Servers where they are located. After having queried the appropriate Spatial Model Servers and the Location Service, the NEXUS nodes conflate the data that is returned from the Service Layer and provide additional services like navigation or map production on their own. Eventually, the results are propagated to the application.

The NEXUS infrastructure contains further services which are for example described in (Volz, 2001).

2.2 The Augmented World Model

The Augmented World Model (AWM) is an object-oriented information model for location-based applications (Nicklas & Mitschang, 2001). It allows a unified view for the applications and enables the federation of spatial data from different providers. The object classes of the AWM are defined in the Augmented World Schema (AWS). The AWS comprises the standard class schema, a generic set of all object classes that might be of relevance for location-based applications. To achieve a reasonable semantic, each standard class contains an extensive set of attributes. However, most of the attributes are declared *optional*, i.e. a NEXUS service does not have to make the effort to provide all of the data, but if it wants to, the name and type of the attributes are already defined. In case that an application still needs additional object classes that further specify the existing ones, they can be inherited from the classes of the standard schema, forming a so-called extended class schema (see Figure 2).

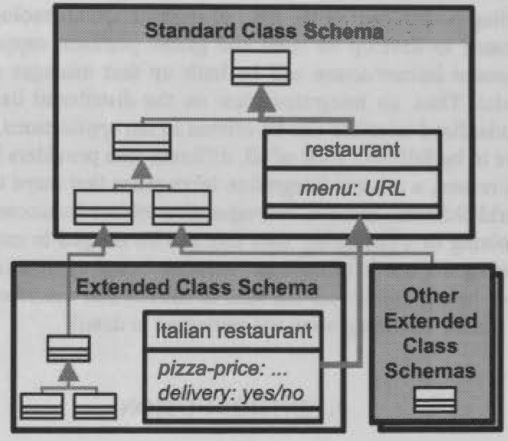


Figure 2: The AWS standard class schema can be extended according to the needs of applications

3. INTEGRATION OF HETEROGENEOUS DATA SOURCES INTO THE AWM

Since the NEXUS platform is aimed at providing an extensive and generic data basis for different kinds of location-based applications, data sources from all possible application fields should be integrated in the AWM. Therefore, existing data models have to be mapped onto the Augmented World Schema. This allows on the one hand a further use of already available data. On the other hand, data from multiple sources can be federated since they are available in a common data format. In the following sections, the process of mapping road data from different sources onto the AWS will be described.

3.1 Existing Data Models for Road Data

The growing popularity of car navigation applications in recent years has lead to a multiple acquisition of road data in different formats. In the following sections the two most common data models for road data in Germany will be introduced. The Geographic Data File (GDF) is used by the major digital road data suppliers, while ATKIS is a topographic and cartographic database. Both formats are acquired in an approximate scale of 1:25 000.

3.1.1 The GDF Data Model:

GDF 3.0 (CEN TC 278, 1995) is a European standard that is used to describe and transfer road networks and road related data. It is much more than a GIS data format, because GDF regulates how data should be captured and how features and their relationships have to be defined. Features correspond to "real world objects" like streets and railways or administrative areas. The feature category defines the type of representation of a feature. Objects represented by a point, line or area are called *simple features*, while *complex features* are composed of a group of simple features. Features contain geometrical as well as topological information. Every feature belongs to a specific feature class. The following feature classes are important for road traffic:

- *Junction* (point): feature bounding a Road Element
- *Road Element* (line): smallest independent unit of the road network having a Junction at each end
- *Road* (complex): contains different Road Elements
- *Intersection* (complex): composed by Junctions and Road Elements

Figure 3 illustrates the relations between simple and complex road traffic features.

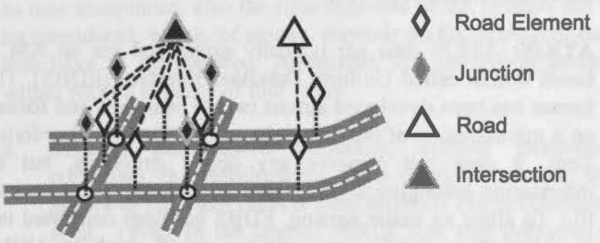


Figure 3: Simple GDF features can be aggregated to complex ones (after Walter, 1997)

Features can also share relationships. Relationships are used to express associations between two or more features like, for example, a prohibited manoeuvre from one road to another.

3.1.2 The ATKIS Data Model:

The data model of ATKIS (Authoritative Topographic Cartographic Information System) has been developed by survey offices of the federal states of Germany (AdV, 1998). It is structured into seven functional classes like e.g. settlement, vegetation and traffic. Each functional class is again subdivided into object groups. Concerning traffic, its object groups are road traffic, railway traffic, air traffic, etc. An object group includes several object types. Object types specify the geometrical representation and the attributes of real world objects. In case of road traffic, defined object types are for example roads, squares or ways. Instances of the object types are the ATKIS objects which are built up by object parts. Object parts contain the geometry of ATKIS objects as vector elements. Therefore, each ATKIS object comprises at least one object part. In case of roads, further object parts are established, if the topology (in case of crossings) or the attributes (e.g. the road width) of an object change. ATKIS demands to build complex objects in case that a road has two or more physically seperated lanes.

3.2 Schema Integration: Mapping existing Road Data Models onto the AWS

Concerning the mapping of existing road data models onto the object classes of the Augmented World Schema, it is important to preserve the original properties of features as far as possible. First of all, a suitable amount of object classes has to be defined in the AWS and thereupon the relations between them and the corresponding object classes of GDF and ATKIS must be determined. Afterwards, the XML schema definition of the Augmented World Modeling Language (AWML) can be extended to road traffic objects and thus the original GDF and ATKIS formats can be converted into AWML.

3.2.1 Defining the object classes:

On the basis of the two underlying data models GDF and ATKIS, a generic set of object classes has been established in the AWS that allows to represent all object classes of the existing data models (see Figure 4). They are all derived from the abstract class *StaticObject* which is the superclass of all static spatial objects (like buildings, streets, etc.) within the AWS. To each object class of the AWS, necessary attributes have been assigned. Only some most basic of them were declared mandatory.

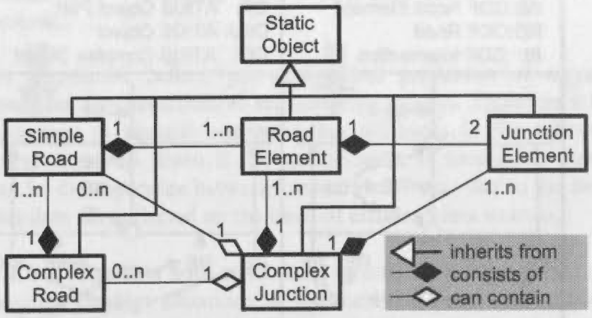


Figure 4: UML representation of road objects in the AWS including cardinalities

Basically, there are 1:1 relations between the object classes of GDF and AWS. However in GDF, Road Elements which are sharing the same street name do not need to be combined to form a road feature. But within the AWS, *RoadElements* carrying the same name have to be aggregated to a *SimpleRoad* object. Additionally - in case of a multi-lane road - GDF road features that have the same name are merged into an AWS *ComplexRoad* (see Table 1).

GDF	Relation	AWS
Junction	1:1	*JunctionElement*
Road Element	1:1	*RoadElement*
Road Element	*n*:1	*SimpleRoad*
Road	1:1	*SimpleRoad*
Road	*n*:1	*ComplexRoad*
Intersection	1:1	*ComplexJunction*

Table 1: Relations between object classes of GDF and AWS

ATKIS	Relation	AWS
Object part	1:2	*JunctionElement*
Object part	1:1	*RoadElement*
Object	1:1	*SimpleRoad*
Complex Object	1:1	*ComplexRoad*
Not existing	–	*ComplexJunction*

Table 2: Relations between object classes of ATKIS and AWS

Regarding the relationships between ATKIS and AWS object classes (see Table 2), it can be recognized that each object part corresponds to exactly one *RoadElement* being bounded by two *JunctionElements*. Except the AWS-*ComplexJunction* class which is not existing in ATKIS, all other object classes share 1:1 relations.

Figure 5 illustrates the mapping of data from GDF and ATKIS onto the Augmented World Schema.

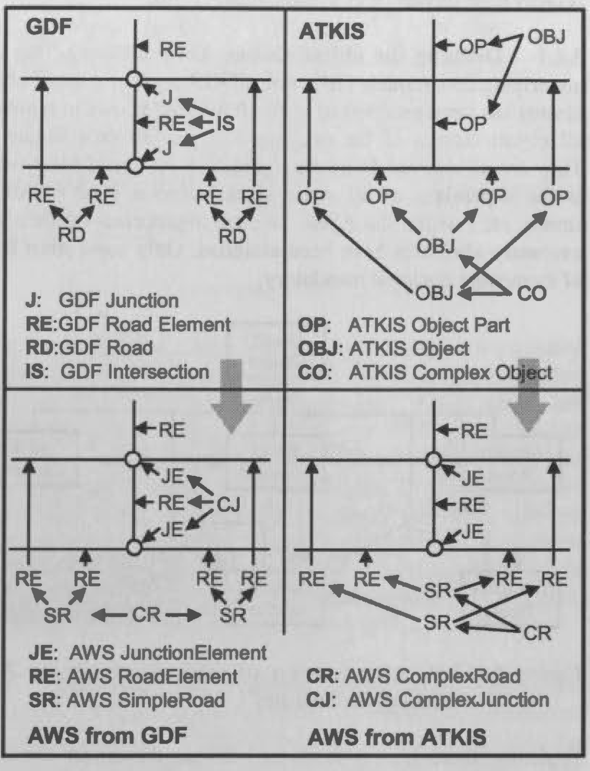


Figure 5: Representation of GDF and ATKIS in the AWS

3.2.2 Developing the AWML Schema: After having built up the object classes for road data within the Augmented World Schema, the definition of the Augmented World Modeling Language (AWML) took place applying the XML schema language (W3C, 2002). Since NEXUS is aimed at providing an open platform, standards are adapted wherever possible. For the geometrical representation of the Augmented World objects, AWML implements the geometry schema definition of the Geography Markup Language developed by the OpenGIS Consortium (OpenGIS, 2001).

Each AWML file consists of a header and a data section. The header carries meta information about the file like its spatial extent, contained object classes, information about the origin of the data, etc. The data section includes the XML description of the AWS objects. Objects of superior classes are referencing subordinate instances by the *href* attribute defined in the XLinks specification (W3C, 2002). For example, a *RoadElement* contains XPointers to the IDs of its From- and To-*JunctionElements* within its *href* attribute. XPointers enable referencing individual elements within an XML document.

3.2.3 Converting GDF and ATKIS into AWML: This section describes the conversion of existing road data into AWML files. Both GDF and ATKIS data sources are stored in a sequential interchange file format requiring special parsers to extract the necessary information. Implementing such parsers can be a tedious task, especially if you are not familiar with the file format. Since AWML is an XML language, the generation of AWML is relatively easy to conduct, because many software libraries support handling XML data with standard interfaces like DOM (Document Object Model). The following paragraphs will give a brief description of the conversion process in our work.

GDF: GDF data are stored in ASCII-based interchange files. In order to extract data of a specific region out of a GDF file, a preprocessing step is performed. It cuts out any feature contained in or intersecting the minimum bounding rectangle of the area in focus. After this preprocessing the conversion tool starts building road objects by parsing the extracted GDF records and inserts them into the Document Object Model (DOM) of an AWML document. As mentioned in section 3.2.1 the tool has to construct *SimpleRoads* out of GDF Road Elements having the same street name and/or number, because in AWS the official name and number of a road are stored in the *SimpleRoad* and not in the *RoadElement*.

The conversion tool is written in C++ and uses Xerces from the Apache XML Project (Apache, 2002) to generate the DOM of the AWML file. The DOM is automatically validated against the XML schema of AWML.

ATKIS: ATKIS data are typically exchanged via an ASCII-based format called Uniform DataBase Interface (EDBS). This format has been developed almost twenty years ago and focuses on a minimization of the data volume by a line-based ordering. Thus, it does not preserve any object structures, but the information belonging to one object is scattered throughout the file. To allow an easier parsing, EDBS has been converted into an object-structured file representation out of which the AWML file is generated.

The parser itself has been written in Java. It reads the ATKIS objects and stores them in appropriate containers according to the object class they belong to. The containers are then written into an AWML file using the JDOM library (JDOM, 2001). JDOM contains classes for generating and parsing DOM instances.

4. CONFLATION OF SPATIAL DATA IN NEXUS

In order to realize a co-processing of heterogeneous spatial data from different origin within NEXUS, it is not only necessary to integrate them in the Augmented World Model. Rather, multiple digital representations of one and the same real world object have to be identified to enable the conflation of adjoining and overlapping spatial data sets. Thus, coherent and semantically enriched data can be generated that best meet the needs of different types of location-based applications.

Generally, a conflation process is the more successful, the less discrepancies between two spatial representations in terms of scale and data model exist. In NEXUS, not only data with a high degree of similarity must be matched, but also correspondencies between heavily differing objects must be detected. The following sections describe the approaches for the conflation of spatial data within the NEXUS platform.

4.1 The Conflation Process

Each time a client query that implies a usage of heterogeneous spatial data sources is directed to a NEXUS node, a conflation process has to be performed. Then, further services can operate on the conflated data (see Figure 6).

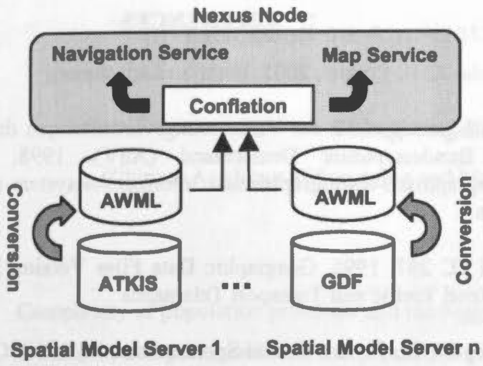


Figure 6: The role of conflation within NEXUS

Since the conflation process must be carried out fully automatically, the algorithms have to be robust and fault-tolerant in order to guarantee an acceptable quality of service for the user. Therefore, errors of the conflation procedure have to be detected by the system itself. This can be realized by introducing quality measures expressing the success of a matching operation between two spatial features. With respect to user acceptance, also the time response of the process has to be considered, which, of course, depends on the amount of data that have to be matched. Altogether, the conflation process within NEXUS consists of four steps:

a. **Schema integration:** Generally, schema integration comprises the development of a database schema that allows to represent objects from different origin (Devogele et al., 1998). In the case of NEXUS, this step conforms to the mapping of heterogeneous data sources within the AWS that was described in section 3.2.
b. **Analysis of meta data:** Each AWML document contains meta information about its data. For example, the coordinate system, the data origin, the spatial extent, etc. are specified in the header section. This information can be exploited to determine, which operations have to be carried out in the conflation process, e.g. if a coordinate transformation has to be done and so on.
c. **Determination of the conflation category:** The conflation category is defined by the type of query the user performs. Either, any data of two AWML files are conflated in overlapping areas to generate one semantically enriched file or two AWML files are merged at their boundaries to receive one coherent data set.
d. **Matching of object instances:** Using semantical, geometrical and topological indicators, object instances are compared in order to determine their degree of similarity. The matching process generally involves many conflicts (see section 4.2) which not only come up due to different data models or scales, but also because of different times of acquisition or erroneous data capturing. Therefore, in some cases a matching process is ambiguous or doesn't yield any result at all.

4.2 Potential Matching Conflicts

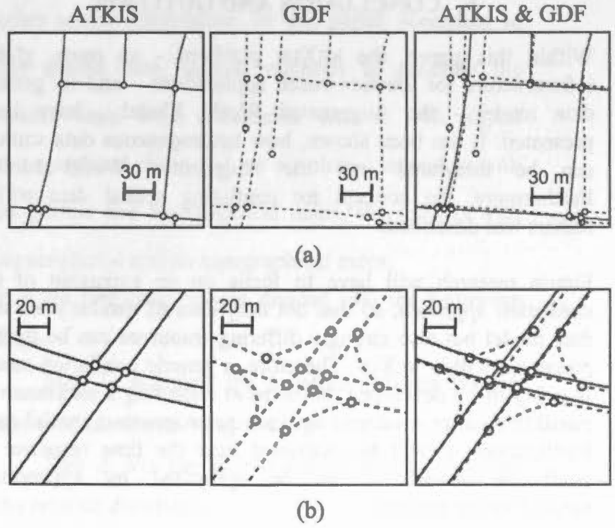
In the following sections, some reasons for matching conflicts are outlined, especially considering road data.

4.2.1 General conflict situations: Depending on the diversity of spatial data sets, different conflicts in the matching process arise. In particular, problems occur when data sets which are represented in different data models have to be merged. For example, if corresponding objects are described by lines in one data set and polygons in another data set, it is not possible to conduct the matching without preprocessing. Therefore, in order to work on common geometrical elements, polygons would have to be reduced to lines by a skeleton algorithm accepting some loss of information.

Another problem is the different levels of detail in which real world objects are stored in data sets. The generalization process takes place during the acquisition of data by an operator and can also be done in order to adapt objects to according scales. Considering this fact, the cardinality of matching objects is not limited to one to one, even many to many relations are possible, considerably increasing the complexity of the conflation process.

As mentioned before, non-geometrical properties of object instances also contribute to the matching process. Therefore it is necessary to identify corresponding attributes in the schema integration step. Even if the identification is successful, there can be discrepancies between common attributes due to the fact that data are collected on the basis of different data sources.

4.2.2 Conflicts with road data: In order to become aware of possible conflict situations in the matching process, both the AWML representations of ATKIS and GDF road networks have to be superimposed and visualized. Since AWML is an XML-based language, it is straightforward to convert AWML files into SVG files (Scalable Vector Graphics) using XSL Transformations (W3C, 2002). In addition, interactive features for further examinations like the selection of objects have been implemented within a GUI using JavaScript.



(a)

(b)

Figure 7: Conflict situations while matching road data of GDF and ATKIS (after Walter, 1997), visualized in SVG

Figure 7 shows typical conflict situations in the matching of road network elements. Because GDF is explicitly meant for car navigation, it is sometimes more detailed than ATKIS. For example, a road represented by one lane in ATKIS data can be represented by a multiply digitised road in GDF (see (a)). Also, complex intersections are captured with a higher accuracy in GDF than in ATKIS (see (b)).

4.3 Matching Approach for Road Data

Objects within geographical information systems have a semantical, a geometrical and a topological component. The semantical and geometrical properties are used as indicators when comparing two digital representations of the same real world object during a feature-based matching process. In addition to the feature-based approach, the relational matching also takes into account the relations of an object to other objects within the same data set. For example, topological relations like "is contained in" or "has *n* neighbours" are considered.

In a matching process, a set of objects A is mapped onto a set of objects B by an appropriate function. Walter (1997) has presented a relational matching solution for ATKIS and GDF road data. First of all, a list of potential matching pairs is generated that can have cardinalities from 1:0 up to n:m. Then, the pairs are rated by feature-based and relational indicators. Thus, the list can be continuously refined by eliminating improbable assignments. In order to achieve a quality measure of the matching process, the result of the automatic procedure is compared to a manually derived reference list. In principle, the same approach can be adopted for the matching of road data within NEXUS.

Since GDF and ATKIS can be represented in AWML, XSL Transformation (XSLT) (W3C, 2002) will be applied for the implementation of basic matching algorithms like the comparison of line lengths or object angles. Still, the capabilities of XSLT concerning more sophisticated conflation operations have to be analyzed. Then, it can be decided which functionalities need to be implemented using external programming languages like Java.

5. CONCLUSION AND OUTLOOK

Within this paper, the NEXUS platform - an open, global infrastructure for location-based applications - and its generic data model - the Augmented World Model - have been presented. It has been shown, how heterogeneous data sources can be integrated into the Augmented World Model. Furthermore, the concept for conflating spatial data within NEXUS was described.

Future research will have to focus on an extension of the conflation approach, so that not only data of similar scale and data model but also strongly differing resources can be further processed within NEXUS. Therefore, a generic conflation model is going to be developed that aims at reflecting a maximum of possible conflict situations between heterogeneous spatial data. Furthermore, it will be examined how the time response of conflation operations can be optimized by appropriate parallelization methods.

6. ACKNOWLEDGEMENTS

The NEXUS project is supported by the Deutsche Forschungsgemeinschaft (DFG - German Research Foundation) under grant FOR 372/1.

The test data have kindly been provided by the NavTech company (GDF) and the state survey office of the federal state of Baden-Wuerttemberg (ATKIS).

7. REFERENCES

Apache XML Project, 2002. http://xml.apache.org

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), 1998. Amtlich Topographisch-Kartographisches Informationssystem (ATKIS). Bonn.

CEN TC 287, 1995. Geographic Data Files Version 3.0, GDF for Road Traffic and Transport Telematics.

Devogele, T., Parent, C. and Spaccapietra, S., 1998. On spatial database integration. *Int. J. Geographical Information Science*, 12 (4), pp. 335-352.

JDOM, 2001. http://www.jdom.org/

Leonhardi, A., Rothermel, K., 2001. Architecture of a Large-scale Location Service. Technical Report 2001/01. Department of Computer Science. University of Stuttgart. 1-17.

Nicklas, D., Mitschang, B., 2001. The NEXUS Augmented World Model: An Extensible Approach for Mobile, Spatially-Aware Applications. *Proc. of the 7th Int. Conf. on Object-Oriented Information Systems*, Calgary.

OpenGIS Consortium (OGC), 2001. Geography Markup Language (GML) 2.0 Implementation Specification. http://www.opengis.net/gml/01-029/GML2.html (accessed 11-17-01)

Volz, S. 2001. Information Management for Location Aware Applications, in: D. Fritsch & R. Spiller, eds, *'Photogrammetric Week '01'*, Heidelberg, pp. 5–12.

Walter, V. and Fritsch, D., 1999. Matching Spatial Data Sets: a Statistical Approach, *International Journal of Geographical Information Science* 13(5), 445–473.

Walter, V., 1997. Zuordnung von raumbezogenen Daten – am Beispiel der Datenmodelle ATKIS und GDF. Dissertation, Deutsche Geodätische Kommission (DGK), Reihe C, Heft Nr. 480.

World Wide Web Consortium (W3C), 2002. http://www.w3.org

Investigation on Application of GIS in Population Management

Chengming LI Zhongjian LIN Jie YIN

(Chinese Academy of Surveying and Mapping, No. 16, Road Beitaiping, Beijing, China, 100039)

cmli@casm.ac.cn

Abstract

Complexity of population problems and the biggest population radix make population management much difficult. Traditional manual record card and population information system lagged the development of economy and society. Geographical information system (GIS) is an effective tool to manage spatial data and attribute data, it has been used widely in many fields such as land, environment, urban planning. However, a few of papers consider the applications of GIS in population management.

In this paper, the doorplate number is used to integrate the population data and spatial data, on the basis of the integrated data, the population geographical information system (PGIS) was constructed. It includes the design of scheme, integration of population data and spatial data, functions of system.

Keywords: Population problem, Population Geographical Information System, Population Data, Design of Scheme, Doorplate number

1. Introduction

China is a large country; it has the biggest population radix. At present, to assure the social stable is the key factor to keep the economy to develop much quickly. So, the work of public safety is very important, in order to do this, population management plays an important role. If traditional mode of population management is still in use, this is not appropriate compared with 20 years ago, the flow of population is so large that present simple system of population information isn't enough. That is to say, in China with improvement of social and scientific technology, with the increasing much quickly of population flow, to integrate the population information and spatial information is much necessary.

Geographical information system is a powerful tool to manage spatial data and database. Although it is widely used in many fields, but due to the shortage of population information, it leads the limitation of GIS applications. Both geographical information and population information are basic information, although they can be used lonely, the integration of them will exert the more important effects.

In this paper, following the introduction part two will discuss the construction of spatial basis. Part three proposed the method link the spatial data and population database. Part four listed the functions of PGIS. At last, the concluded remarks were given.

2. Spatial Basis

Spatial basis is a basis to integrate the population and other social information. In this paper, according to the spatial data, several methods to generate the spatial basis were considered. One is with digital topographical maps and no air-photo, second is with air-photos and topographical maps, the third is with no air-photos and no topographical maps.

For the first case, it was divided into the following steps:

Step 1: scanning the negative airborne photos, and then the digital image were gotten;

Step 2: Via the known control points, finishing the relative direction;

Step 3: collecting the digital elevation model;

Step 4: digital correction and digital mosaic;

Step 5: fusion of remote sensing image and black-white airborne image;

Step 6: According the demands of population management, generating the basic geographical units.

For the third case, it was divided into the following steps:

Step 1: scanning the negative airborne photos, and then the digital image were gotten;

Step 2: Via the known control points and reference data of photo, finishing the backside direction;

Step 3: Using the known digital elevation model, correcting the image and digital mosaic;

Step 4: fusion of remote sensing image and black-white airborne image;

Step 5: According the demands of population management, generating the basic geographical units.

For the second case, it was divided into the following steps:

Step 1: scanning the paper topographical map, and then the digital image were gotten;

Step 2: Via the known control points, finishing the correction of DRG (Digital Raster Map);

Step 3: Digital image processing, and finishing the geometrical correction;

Step 4: fusion of remote sensing image and black-white DRG;

Step 5: According the demands of population management, generating the basic geographical units.

3. Integration of Population and Spatial information

Every family has only one doorplate number. The doorplate number is a keyword to link the house in the spatial basis to the existed population database. Because on the spatial basis the image of can be distinguished, at the same time in the population database the doorplate number is one item. The relation between the image of buildings and the doorplate number is 1:n. Via the relation, the population information and spatial information can be integrated (See Figure 1).

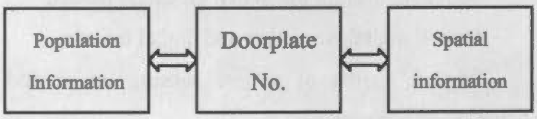


Fig.1 Contact between of Spatial information and Population information

The process of integration of population and spatial information can be divided into several steps as followings:

Step1: to get the population information from the census data;

Step 2: to get the information from resident of police department;

Step 3: to make the unified geographical unit via the spatial information (see figure 2)

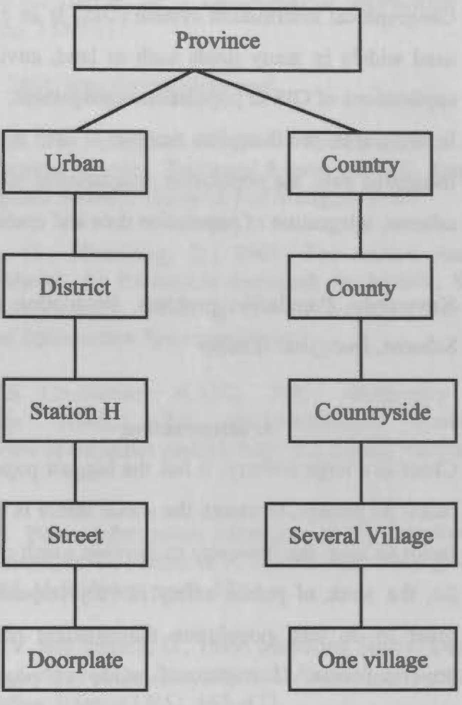


Figure 2. Geographical units of population information

Step 4: to integrate the census data and residence registration data by using the unified geographical units;

Step 5: to link the spatial information database and integrated population information database by the doorplate number in urban area and by the name of village in rural area.

3. PGIS-An Example

In the Hebei province, the city Langfang, which is very near to Beijing, was selected as test area. The 1-meter RS image is used as the spatial basis. A PGIS system was designed and finished. It includes three kinds of functions in the system, first one is to find the place according the population information, second is to find the population information according the spatial information, and the last is population information spatial analysis (See figure 3,4,5).

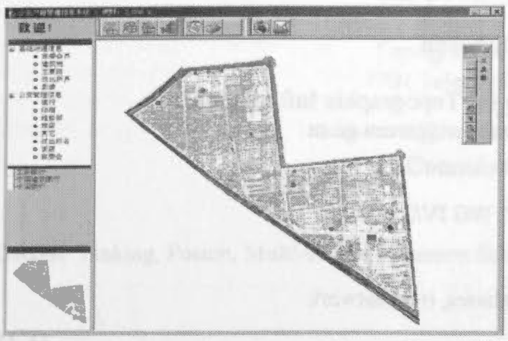
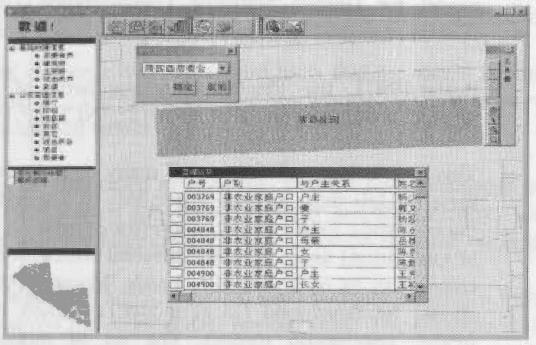


Figure 3 Integration of population and spatial information



户号	户别	与户主关系
003769	非农业家庭户口	户主
003769	非农业家庭户口	妻
003769	非农业家庭户口	子
004848	非农业家庭户口	户主
004848	非农业家庭户口	岳母
004848	非农业家庭户口	女
004848	非农业家庭户口	子
004900	非农业家庭户口	户主
004900	非农业家庭户口	长女

Figure 4. Find the place using population information

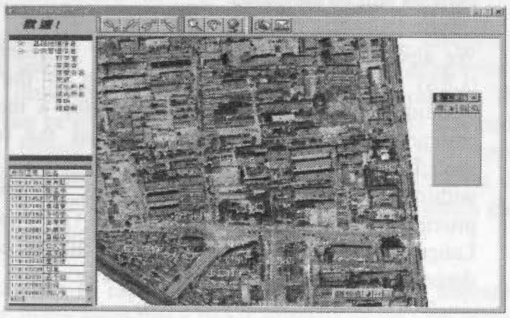


Figure 5. Find the population using place information

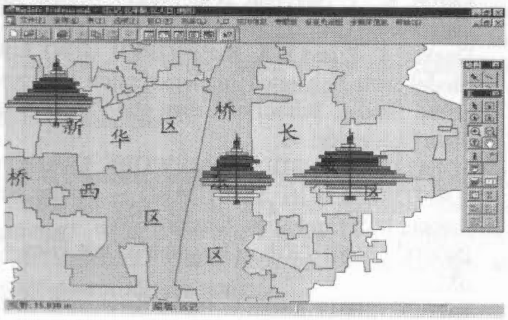


Figure 6. Population spatial analysis.

4. Concluded Remarks

From this paper, several remarks can be drawn: firstly, PGIS is very necessary in china. Secondly, analysis and experiments state that the design and construct method of PGIS proposed in this paper is operative. Thirdly, it is very helpful to population management.

Reference

1. Chengming LI, Jie YIN and Zhigang HONG, 2001, Design and Construction of Population Geographical System. Journal of land and resource. Vol. 4. (In Chinese)

1. Chengming LI et al, 2000, Scheme of Yiwu urban Geographical Information System. Research Report.

2. Hakima Kadri-Dahmani, 2001, Updating Data in GIS: Towards a more generic approach. Proceeding 3 of The 21th International Cartographic Conference. August 6-10,2001, in Beijing. Pp. 1463-1470.

3. Uitermark. H, Oosterom .P, Mars. P, Molenaar. M, 1998, Propagating Updates Corresonding Objects in a Multi-source Environment. Proceeding 8th International Symposium on Spatial Data Handling, pp. 202-213, 1998.

4. H. Samet, Neighbor finding in images represented by octrees ,Computer Vision[J] ,Graphics, and Image processing, 1989, 46, 367~386

5. Egenhofer M and Herring J. Categorizing binary topological relationships between regions, lines and points in geographic databases. In: A Framework for the Definition of Topological Relationships and an Approach to Spatial Reasoning within this Framework, edited by Egenhofer M, Herring J, et.al. Santa Barbara, CA: 1991. 1-28.

UPDATE OF THE CANADIAN ROAD NETWORK IN A DISTRIBUTED GEOSPATIAL DATA FRAMEWORK

D. Boutin, M.Sabourin

Natural Resources Canada, Centre for Topographic Information
Dboutin@nrcan.gc.ca, Msabouri@nrcan.gc.ca

Commission IV, WG IV/2

KEY WORDS: Distributed databases, data modeling, federated databases, road network

ABSTRACT:

For some time now, the Centre for Topographic Information (CTI), part of Natural Resources Canada (NRCan), has wanted to implement a network of partners for capturing and managing geospatial data. These partner organizations, selected for their specific interests (for example, the road network) or for their ease in offering adequate, up-to-date representations of particular phenomena, would make it possible to better serve Canadian community with respect to geospatial data. This new vision has met with the approval of provincial partners and represents a growing trend in the geospatial data infrastructure in most industrialized countries.

These data must be the product of a homogenous, standardized view of the entire Canadian landmass. The approach sought must, to the greatest degree possible, allow each organization to pursue its activities and develop according to its respective needs. Many problems stand in the way of bringing about this theoretical vision of managing spatial data. In recent years, a number of projects have attempted to resolve similar problems, most of which have only been partially successful. One of the main success indicators of cooperative initiatives relates directly to the level of involvement of the participating organizations and individuals. The levels of participation and interest across the board have never been higher. Managing and updating the Canadian Road Network is the first topographic information theme targeted by this new environment for distributed geospatial data.

Sharing geospatial data has been and remains a topic of research and of many large-scale projects. The issue of data heterogeneity in a distributed environment has two distinct but complementary aspects: system architecture and data modeling. [Bishr,1997]. Many private- and public-sector organizations have attempted, through the *Open GIS Consortium* (OGC), to resolve the issues related to system interoperability in particular. This article deals with the issue of data modeling.

Many different models have been developed to manage road network geospatial data. These models attempt to resolve specific aspects such as vehicle navigation, addressing, and 911 emergencies. The main models that attracted our attention are ISO-TC204 GDF [ISO/TC,2000], the data model for Unified Network Transportation (UNETRANS) [Curtin *et al.*,2001], Enterprise GIS-T [Dueker and Butler,1997], the identification standards of the FGDC [FGDC,2000], and NCHRP 20-27(2). [NCHRP,1997].

This article looks at the use of many data models. Each of these models must provide a response to a specific problem. Each specific field of application (transportation, addressing, or navigation) uses a slightly different abstraction in order to respond to its own particular needs. Many levels of models should be developed (local model, common or federated model, application model). The set of models must be organized and put in hierarchical order. The use of the federated model for establishing data homogeneity will be analyzed. The advantages and disadvantages of the associated variants will be discussed.

We also discuss the use of the Linear Referencing System (LRS) to convey the descriptive part of the data. This technique will make it easier to develop a more compact common model. We will also discuss the relevance of specifications of ISO TC-204's Geographic Data Files (GDF) within the purview of this architecture. Moreover, we will provide a brief analysis of the use of Geographic Markup Language (GML) [Lake,2001]. The article ends with a description of the state of work underway.

References

Bishr, Y. (1997). Semantic Aspects of Interoperable GIS.
Curtin, K., V. Noronha, et al. (2001). ArcGIS Transportation Data Model (Draft), UNETRANS Data Model: 37.
Dueker, K. J. and J. A. Butler (1997). GIS-T Enterprise Data Model with Suggested Implementation Choices, Center for Urban Studies School of Urban and Public Affairs Portland State University: 46.
FGDC (2000). NSDI FRAMEWORK TRANSPORTATION IDENTIFICATION STANDARD -- Public Review Draft. Reston, U.S. Geological Survey: 126.
ISO/TC (2000). GDF - Geographic Data Files - Version 4.0: 651.
Lake, R. (2001). Location-Based Services & GML: 8.
NCHRP (1997). A Generic Data Model for Linear Referencing System. Washington, National Cooperative Highway Research Program: 28.

INTEGRATION AND ACCESS OF MULTI-SOURCE VECTOR DATA

Dan Edwards and Justin Simpson

Topographic Engineering Center[1]
7701 Telegraph Road, Alexandria, VA
22315 USA

Commission IV, WG IV/2

KEY WORDS: Linking, Fusion, Multi-source, Common Schema, Best Map.

ABSTRACT:

When vector representations of the terrain from different sources are overlaid, the representations do not agree. The disparity can be due to scale, resolution, compilation standards, source accuracy, registration, sensor characteristics, currency, temporality, or errors. Whatever the case, the data do not agree and the user is unsure which data best represents the terrain.

This paper concerns two managerial tools that are useful in integrating and validating data from multiple sources. The first tool, called "linking", integrates the data automatically, while preserving disparities. A second tool, called "Best Map", assembles the data into a unified whole.

Linking is an automated process that forms a correspondence between feature elements in two overlapping vector data sets; these feature elements must correspond spatially and thematically. Linking simply identifies two representations of the same element on the terrain without judging which representation is better. A fundamental tenet of linking is that we do not know the truth.

The Best Map is an application that automatically assembles linked data sets into a unified representation. The Best Map application neither destroys the original data nor the links. The schemas of the two data sets are combined into one Common Schema that makes full use of the separate attribution.

1. INTRODUCTION

People, wishing to integrate vector information, discover a disturbing reality: multi-source data over the same geographical area are disparate, both spatially and thematically. The disparity can be due to scale, resolution, compilation standards, operator license, source accuracy, registration, sensor characteristics, currency, temporality, or errors. Whatever the case, the data do not agree and the user is unsure which data best represents the terrain.

It is reasonable to expect that a vector data set is internally consistent, in the sense that a line is suitably placed with respect to other lines in the data set, given the limitations mentioned above. Normally there is spatial inconsistency between multi-source data sets over the same geographic area, which raises concerns about spatial validity. Thematic validity comes into question when objects, which seem to be the same, have different attribute values or even different attributes, due to differences in the schemas.

Concerns about consistency and validity lead to other questions. To what extent do overlapping data sets represent the same physical entities? What is the true location of a spatial object? What are its attributes? How does one store, manage and integrate multi-source information?

Multi-source data needs special tools to manage, automatically, the disparities that are endemic to such data. This paper concerns two managerial tools that are useful in the integration and validation data from multiple sources. We shall discuss some of the problems that arise and give our automated solution, called "linking", where disparities can be preserved yet integrated. We shall discuss a second tool, called "Best Map" in which linked data can be viewed as a unified whole. Linking is built around a fundamental tenet: we do not know the truth, without further information or source imagery.

Nevertheless, this linking process generates new information when linked data elements reinforce, augment, or contradict each other. Linked data elements represent differing analysts' views of the same entity. When viewed in the context of the collection specifications, independent interpretations can provide a spectrum of information, ranging from mutually reinforcing to contradictory. Linking is a comparison of data and that comparison can strengthen or weaken our confidence in the underlying data, providing a more realistic corroboration in either case. Map information is inherently pluralistic – we do not know the truth, but we can alter our certainty through linking.

2. WHAT IS LINKING?

When vector data from two sources are overlaid, the vector representations of the terrain do not agree. Linking is an automated process that forms a correspondence between feature elements in two overlapping vector data sets; these feature elements must correspond both spatially and thematically. Linking does not attempt to eliminate differences. Linking simply identifies two representations of the same element on the terrain without judging which representation is better. A fundamental tenet is that we do not know the truth, in the absence of further evidence. See figure 1.

---

[1] The work described here reflects work over a number of years and across several programs. Intergraph Corporation has written prototype linking software. Intergraph and Swiftsure Spatial Systems Inc. have written prototype best map software. These software developments have helped us test and refine the ideas presented here.

For example, higher resolution data cannot automatically be assumed to be "better data" than lower resolution information. Figures 2 and 3 illustrate this point. Figure 2 contains a collection of features loaded from different sources. In this example, an urban area is covered by road feature data from two sources, one of a higher resolution and detail than the other. Some of the information available on the 1:50,000 scale source (black) is unique and not available on the 1:250,000 scale source (gray). But the reverse is also true in a nearby region of these same collective data sets (Figure 3). The assumption that one data set is better than the other is obviously incorrect. The discrepancies may be due to different collection specifications, temporal disparity, sensor differences, or obscuration in the source imagery.

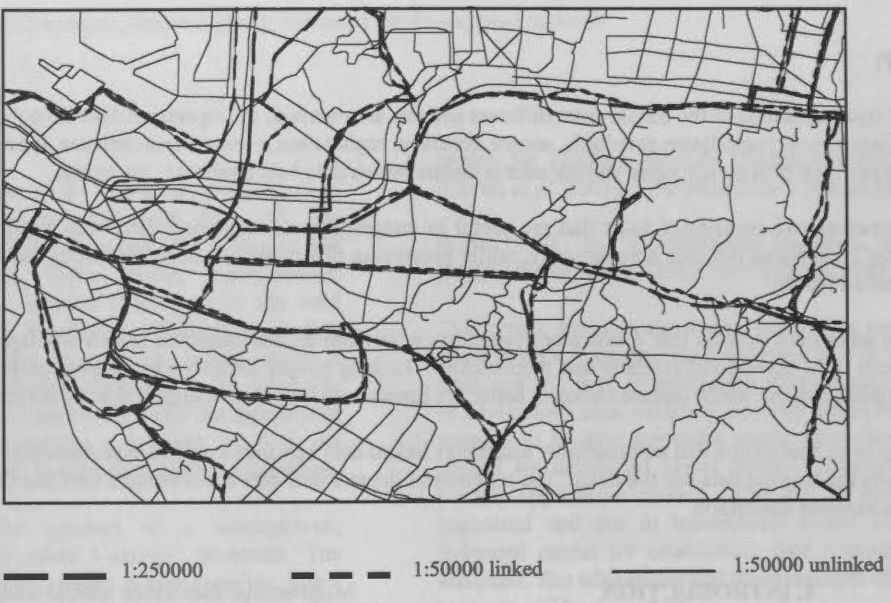


Figure 1

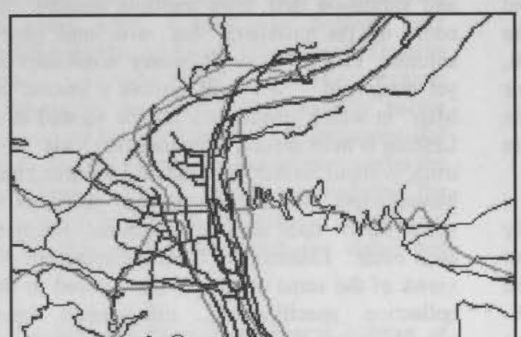


Figure 2

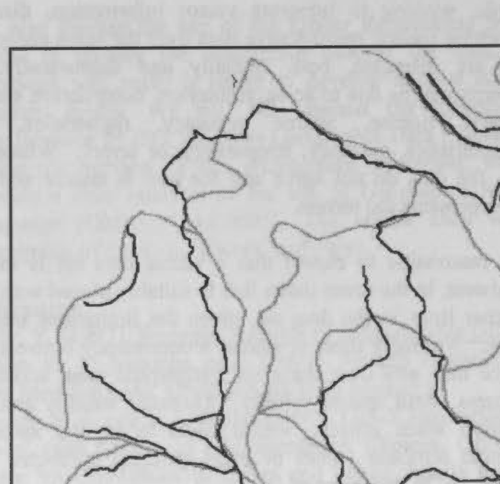


Figure 3

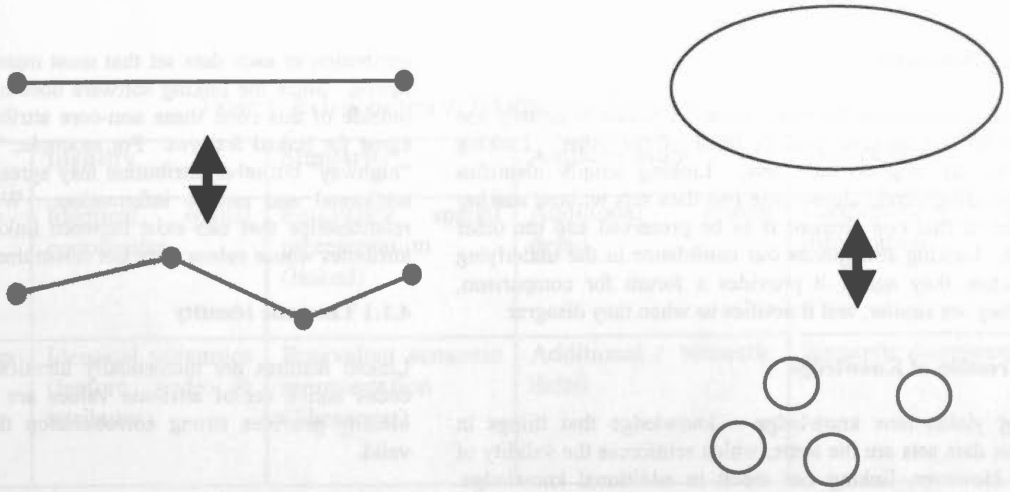


Figure 4. Linking between feature elements

Comprehensive "truth" is not attained from a single source. The linking took, within a pluralistic system, integrates different descriptions of the same geographic area. Data are stored "as-is", as received from the provider, preserving separate spatial and semantic detail. From a data perspective, this resulting integrated view has many topographic feature elements. Each feature represents the "best" version of the spatial component available in the pluralistic database. User applications can then be run on these integrated "best" elements.

Feature elements can represent points, linear entities, or areal entities, and the linking correspondence is not constrained by dimensionality. For example, an areal river in one source can be linked to a linear river in another source. A point bridge can be linked to a linear or areal bridge. An areal city can be linked to a point city. The entities in one source that correspond to entities in another source can be single feature elements or groups of feature elements.

Nevertheless, linking generates new information when linked data elements reinforce, augment, or contradict each other. Linked data elements represent two analysts' views of the same entity. When viewed in the context of the collection specifications, two independent interpretations can provide a spectrum of information, ranging from mutually reinforcing to contradictory. The new information provides a more realistic confidence in the feature elements. (See Section 3.5.)

Although linking is a non-destructive process that preserves the original data sets, linking facilitates the assembly of a unified view of the terrain that represents the intelligent fusion of the disparate information. (See Section 5.) Since any number of data sets can be linked, one can assemble an integrated view of all available data over an area of interest.

3. BENEFITS FROM LINKING

The previous section on the reasons to link provides a background against which we can appreciate the benefits of linking. In this section we shall discuss various benefits that result from linking.

3.1 Potential to Use All Available Information

As discussed above, no single source contains the perfect representation of the terrain for all applications. This pluralistic reality suggests that a goal of mapping is to integrate data from various sources in a way that captures the unique advantages of each source. Commercial data, used outside of the traditional mapping community, may have rich thematic content that can enhance data of a higher spatial fidelity. As the commercial importance of geospatial data increases, all of tomorrow's data sets, both commercial and governmental, could become a rich collective source of information for the mapping community.

3.2 Automated Use of Future Information

New vector data can be linked to existing data – automatically integrating new vector information with the old. Minor discrepancies are resolved and only significant differences are highlighted so that human intervention is invoked only when problems are indeed major. This integration of new data through linking avoids conflation and its inherent pitfalls. (See Section 3.4.) Moreover, linking greatly reduces the need for human editing and the resulting destruction of information.

3.3 Dealing With an Uncertain Reality – We Do Not Know the Truth

We do not know the truth – at least when comparing disparate maps. We have seen examples where low-resolution data contained roads that were not present in higher resolution data. Lower resolution data can be collected with a richer set of attribution than higher resolution data. When maps differ, massive human intervention is needed to explain the differences, and sometimes even human intervention does not suffice. Even with the aid of source imagery, it can be difficult to resolve differences between maps.

Such issues as scale, resolution, spatial accuracy, currency, and thematic accuracy are important aids to be considered when resolving differences between maps. However, it is risky to decide that one feature is correct while another is incorrect based upon automated criteria.

3.4 Not Destructive

Since we do not know the truth, we are reluctant to destroy one of two conflicting data sets, in favor of the other. Linking preserves the original data sets. Linking simply identifies corresponding terrain elements in two data sets without making a judgment that one element is to be preserved and the other deleted. Linking strengthens our confidence in the underlying data, when they agree; it provides a forum for comparison, when they are similar; and it notifies us when they disagree.

3.5 Creation of Knowledge

Linking yields new knowledge – knowledge that things in different data sets are the same, which reinforces the validity of both. However, linking can result in additional knowledge. Linking data sets can result in a synthesis that is greater than the sum of the component parts.

For example, a set of high-resolution roads that are linked to a low-resolution road is endowed with an additional attribute: that each high-resolution road is a candidate for a major road. Attributes can be transferred from one road to its linked counterpart. As we shall see in the discussion of Best Map, linking two road networks provides additional knowledge of how to integrate the networks – of how to pass from one network to the other.

3.6 Opportunity to Compare Sources – Strength in Numbers

Linking is a comparison of data and that comparison can strengthen or weaken our confidence in the underlying data. Linking is a confirmation that separate analysts agree on a representation of the terrain, increasing out confidence that each representation is correct. When linking indicates that two data sets are contradictory, linking each data set to a third might give support to one of the conflicting views. Multi-source map information is pluralistic – we do not know the truth, but we can alter our certainty through linking.

4. PLURALISM[2]

Overlapping vector data provides independent views of the same underlying physical reality – views that are sometimes incongruous or sometimes reinforcing. Among these multiple views, we do not know which view, if any, is correct or even best. The nature of multi-source map data is pluralistic.

Linked feature elements are constrained by the linking algorithm to be similar both in space and in a core set of thematic attribute values. In this section, we shall examine how linking can alter our perception of this pluralistic state of affairs, both inside and outside of the core areas of agreement. The following table suggests the kinds of relationships that can arise in linked information.

4.1 Thematic Pluralism

In this section we compare the thematic content of objects that are spatially similar or identical, as established through linking. Before data sets are linked the user defines core sets of attribution in each data set that must match before features are linked. Since the linking software does not consider attributes outside of this core, these non-core attributes may or may not agree for linked features. For example, "road" should link to "highway" but other attribution may agree, disagree, or provide additional and unique information. We shall examine the relationships that can exist between linked features and their attributes whose values were not constrained by linking.

4.1.1 Thematic Identity

Linked features are thematically identical when their feature codes and a set of attribute values are identical. Thematic identity provides strong corroboration that the attribution is valid.

It is useful to recall a difference between thematic and spatial agreement. One would expect spatial differences between linked features to be small, with the discrepancies being due to scale, resolution, generalization, registration and the like. The technology to derive accurate spatial information is mature. In contrast, supplying attributes to objects in imagery is an imprecise craft, making identical thematic agreement a happy event.

4.1.2 Thematic Similarity

Terrain elements are linked when they are spatially similar and are similar in a core set of attribute values. In this section we are considering linked terrain elements that have thematic similarity outside of this core set.

Thematic similarity increases confidence when, for example, a road, whose transportation use is "road" in the schema of Source 1, links to a road, whose transportation use is "highway" in the schema of Source 2.

Comparing similar attributes within a schema to similar attributes in different schema can increase or decrease uncertainty, depending upon the context. Translation between schemas can be so imprecise that a meaningful comparison of features is difficult. For example, a vegetation feature in Source1 can have the attribute "without trees" while the same feature in Source 2 could be labeled "paddy". These two descriptions are hardly an inspiring confirmation; possibly, the attributes are as close as the two specifications allow. This is a case in which one feature could inherit the attribution of the other, in the sense that the feature without trees is possibly a paddy and that the paddy is possibly without trees.

An example of decreased certainty arises when a forest with predominant tree height of 10-15 meters is linked to a forest with predominant tree height of 15-20 meters. In the absence of other evidence, such as temporality or accuracy, the certainty or each attribute decreases while we become more certain that the predominant height is between 10 and 20 meters.

4.1.3 Thematic Augmentation

If linked features have some identical or similar attribute values, then the unique attributes of one can be transferred as possible attributes of the other, making the fullest use of information from both sources. This is particularly useful when one source contains much richer thematic detail than the other.

---

[2] We thank Ms. Gail Kucera for helping us develop our ideas on pluralism and Best Map. Ms. Kucera suggested using "pluralism" in this context.

Table 1: Relationships in Information Pluralism

	Identity	Similarity	Augmentation	Contradiction
Spatial Pluralism	Identical spatial coordinates	Equivalent spatial representation (linked)	Additional spatial detail	Conflicting spatial information
Thematic Pluralism	Identical semantics (feature code & attributes)	Equivalent semantic representation (Thesaurus)	Additional semantic detail	Semantic disagreement

Linking also provides the opportunity to augment the data with new knowledge – knowledge not explicitly present in either source. For example, if a 1:50000 road is linked to a 1:10000 road, then the higher resolution road is endowed with the additional attribute that is possibly a major road. Using this new information is discussed below in Section 6, Best Path.

4.1.4 Thematic Contradiction

The eyebrow of uncertainty is raised when linked features have attributes that are contradictory. Linking has identified a conflict that needs resolution – either by a human or by another linked source. In practice, thematic contradiction uncovered by linking has proved to be an important tool in automated error detection. If timely resolution is impossible at least the user is given a warning to use the data with care.

4.2 Spatial Pluralism

Linked objects are constrained by the linking algorithm to be spatially similar. Moreover, in the basic model, we expect that a base feature is linked to secondary feature elements. Nevertheless, comparing the spatial content of linked objects can provide new information.

4.2.1 Spatial identity

Given the difficulty that a single operator experiences in placing the cursor twice on the same location, it is unlikely that two feature elements have identical vertices. If two feature elements have identical vertices, the coincidence would be so striking to suggest that the feature elements have a common digital origin. If all the vertices of one feature element are present in a second linked feature element then there is a strong possibility that the first feature element is a generalization of the second.

Spatial identity, coupled with agreement of core attributes is a strong indication that the feature elements represent the same entity and have a common genesis.

4.2.2 Spatial similarity

This is the case where objects are spatially similar while agreeing on a core set of thematic attributes. This is the sense of agreement that the linking software is designed to uncover. To say that linked objects are spatially similar is simply to say that they are linked.

The linking of two objects provides a confirmation that the positioning and core attributes are correct, within the constraints of both the product accuracies and of their respective schemas. In addition, new information can arise from attribute transfer where attributes from one object are transferred to the other.

4.2.3 Spatial augmentation

Transferring spatial information from one source to another can violate the internal consistency of the recipient data set. Ideally, spatial transfer of data is best done manually using editing tools in a photogrammetric environment. However, such spatial transfer represents a temptation that linking is designed to diminish, not increase. We link so that we do not have to manually adjudicate differences between data sets.

While spatial augmentation has its perils, it can have benefits as well. For example, consider a low-resolution road network that is linked to a high-resolution road network. The linked high-resolution roads represent a view of the low-resolution roads but at a higher resolution. The linked high-resolution roads can be generalized to obtain another view of the low-resolution network. It is quite possible that the generalized roads violate the internal consistency of the high-resolution data.

4.2.4 Spatial contradiction

A spatial contradiction arises when a base feature element can not be linked. This is a strong indication that something is wrong, either in our expectation, in the data, or in the linking algorithm. Human intervention is needed to resolve this contradiction – possibly with the aid of additional information or imagery.

Figures 5 and 6 illustrate spatial contradictions that were uncovered by linking. In Figure 5, the gray high-resolution roads should link to low-resolution roads, but linking fails in the center area due to a gap in the high-resolution data. Figure 6 shows that the gap in the high-resolution data is a bridge.

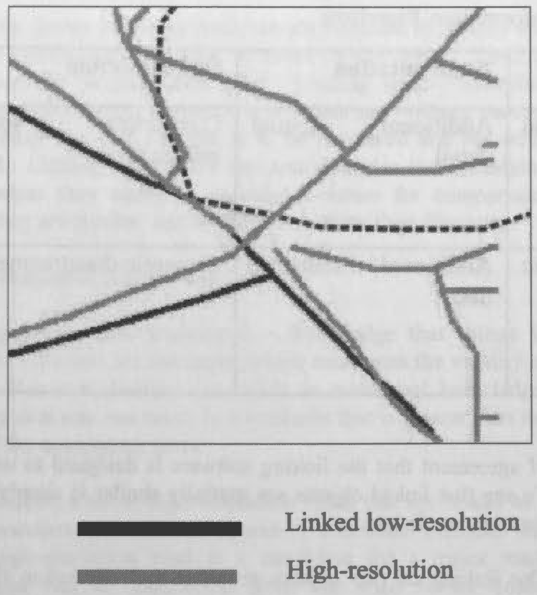


Figure 5

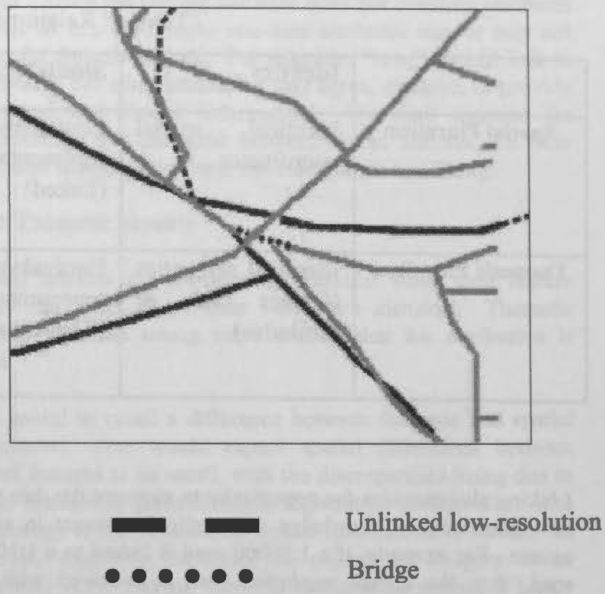


Figure 6

Roads in the low-resolution data are represented as both roads and bridges in the high-resolution data. Linking roads to roads produces a linking failure, while linking roads to roads plus bridges succeeds.

The unlinked low-resolution road at the top of Figures 5 and 6 is another example of a spatial contradiction. It seems that this road should have linked to a high-resolution road. Perhaps this failure is due to a topological problem in the high-resolution data, or to the features being too far apart, or to a flaw in the linking software.

The high-resolution roads exhibit curious behavior near the center of Figure 6, where the unlinked low-resolution road (black dashed) joins the linked low-resolution road (black). Which resolution, high or low, better represents reality? Without further information or source data, we do not know.

Figure 7 illustrates a similar problem. The road network contains a gap because a road has been mislabeled as a runway. Here linking would fail since one does not expect that roads in one source will be represented as runways in a second source that is shown above.

5. BEST MAP

Linking establishes a correspondence between feature elements across two data sets. The Best Map is an application that enables one to view linked data sets, assembling them in an intelligent manner into a unified representation, capable of supporting digital spatial analysis. As we shall see, the Best Map application destroys neither the original data nor the links. The output of the Best Map software can be either stored or deleted, as the user wishes, without affecting the integrity of the source data.

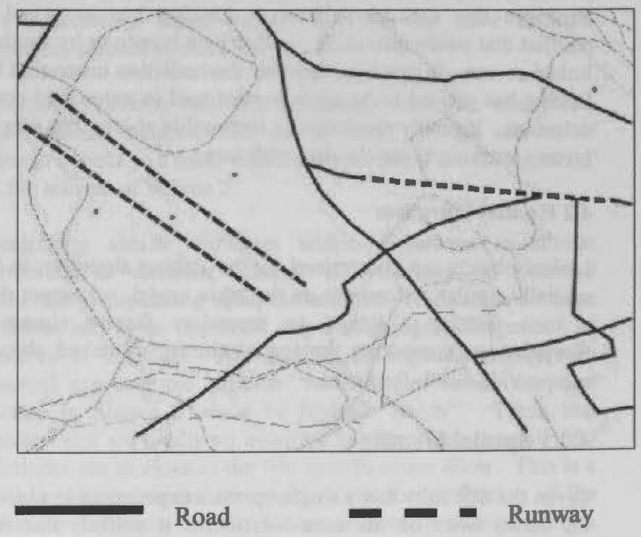


Figure 7.

5.1 Current implementation of linear Best Map

Since the larger issues are still under investigation, this discussion of Best Map centers on software developed independently by Swiftsure and Intergraph, under the direction of TEC. These applications are designed for the case of a low-resolution road network linking to a high-resolution road network. We are still investigating the representation of area, point, or other linear data sets. The rules of similar-scale Best Map are still under investigation, as well.

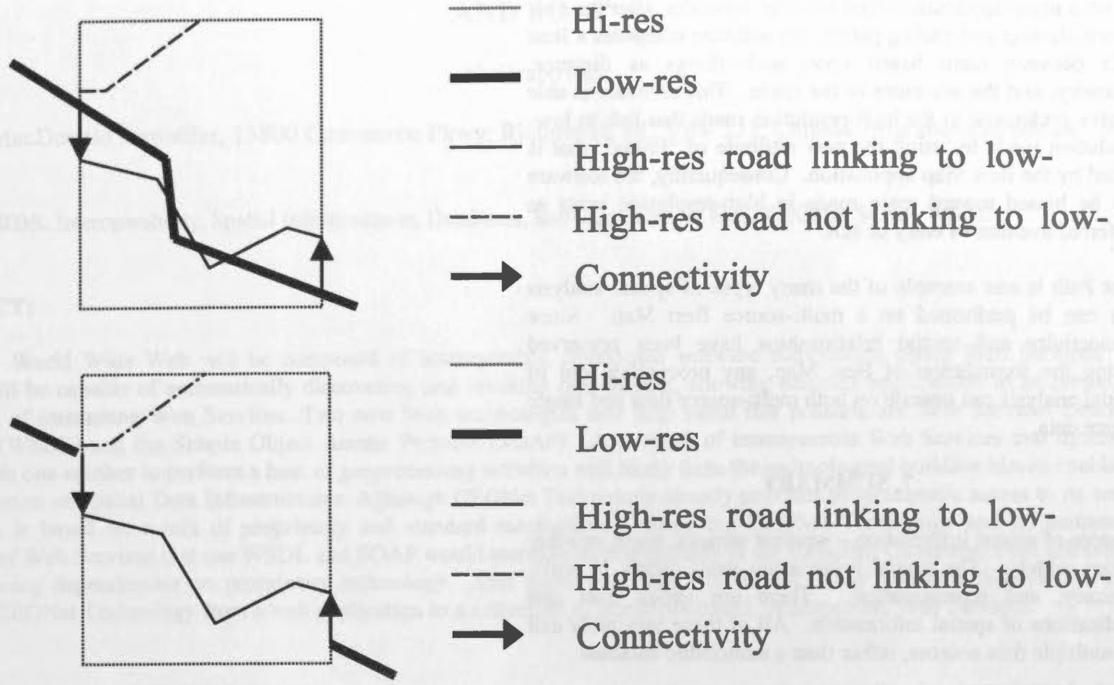

Hi-res
Low-res
High-res road linking to low-
High-res road not linking to low-
Connectivity
Hi-res
Low-res
High-res road linking to low-
High-res road not linking to low-
Connectivity

Figure 8 Construction of Best Map

5.1.1 Spatial rules

In constructing the best representation of the data this software uses spatial and thematic rules. The current spatial rules for creating the Best Map are summarized:

1 in high-resolution areas, the high-resolution data take precedence

2 in outside high-resolution areas, the low-resolution data are the best

3 place connectivity vectors at the boundary

4 retain unlinked low-resolution features that penetrate the high-resolution area.

Include all high-resolution roads in the Best Map. Outside the area of high-resolution, the low-resolution data are best, by default. Along the boundary between these data sets, place connectivity vectors to join a low-resolution road to the high-resolution road to which it links. These connectivity vectors join base and secondary roads that are linked in the interior of the high-resolution area – not simply joining roads that are closest at the secondary boundary. This ensures connectivity of major roads. Linking is based upon the behavior of the base and secondary roads within the secondary area, rather than upon the nearest neighbor at the boundary. See Figure 8.

5.1.2 Thematic rules – Common Schema Formation

Quite possibly, no one schema will encompass the attributes of two linked data sets. Mapping these schemas into an existing schema usually results in a loss of information. Creating a new schema results in non-standard information. The schema of the Best Map is, notionally, the union of the schemas of the source data sets, called the common schema. This non-standard schema is unique to the Best Map, but it has the advantage that no information is lost. This is based on the principles, described above, of thematic similarity and augmentation.

In forming the common schema, a base feature class serves as a core feature class in the common schema. The common schema feature class has all the attributes of the core feature class plus the attributes of all the secondary features classes that are permitted to link to the core feature class. Common attributes, if any, among these core and secondary features classes are used only once in the common schema.

In addition, there are new attributes in the common schema feature class. There is a new attribute indicating that the Best Map feature coincides spatially with a secondary feature element that is linked to a base feature element. Another new attribute indicates that a low-resolution road is unlinked. The connectivity vectors form a feature class in the common schema.

6 BEST PATH

Best Path is an example of a practical application of linking and Best Map. A person might want to travel from a rural area, mapped at a low resolution, to an urban area, mapped at a high resolution. These maps can be linked and a unified Best Map created, enabling a Best Path to be calculated.

The first step is to obtain one continuous, unifying map that combines differing scales, differing attributes and differing spatial representations. This was discussed in the section on Best Map. The second step is to traverse the Best Map, making use of its unique qualities. This is the Best Path application whose first version was developed by Dr. Dianne Richardson of the Canada Centre for Remote Sensing, under a contract with Swiftsure Spatial Systems Inc. Helonics Inc. and Intergraph developed later versions of this software.

As with many applications that traverse networks, after the user selects starting and ending points, the software computes a Best Path between them based upon such things as distance, geometry, and the attributes of the roads. This software is able to give preference to the high-resolution roads that link to low-resolution roads by using the new attribute of "linked" that is added by the Best Map application. Consequently, the software can be biased toward main roads in high-resolution areas as preferred avenues of entry or exit.

Best Path is one example of the many types of spatial analysis that can be performed on a multi-source Best Map. Since connectivity and spatial relationships have been preserved during the formulation of Best Map, any processing tool of spatial analysis can operate on both multi-source data and single source data.

7. SUMMARY

Sources of spatial information – whether sensors, maps, or other – vary widely. The spatial information itself differs in scale, accuracy, and representation. There are varied uses and applications of spatial information. All of these inevitably call for multiple data sources, rather than a monolithic database.

Multi-source data are disparate. We believe that this disparity need not increase uncertainty. Rather, multi-source data gives us the opportunity to increase our confidence by allowing pluralistic representations to become mutually reinforcing, within the context of their respective collection specifications. Differences, which cannot be resolved, can identify possible errors.

We believe that the solution to maintaining a national core geospatial database is to exploit the similarity, duplication, and uniqueness in multiple data sets in order to augment current data and to derive new information. Research in feature linking involves developing a set of tools that enable these capabilities. Research in Best Map involves developing a set of tools that can quickly extract the best data from multiple data sets and assemble this into a coherent picture for a given application and for a given instant in time. To date, our results are exploratory and tentative. Further research is needed.

ACKNOWLEDGEMENTS

Intergraph Corporation has written linking software. Intergraph and a team of contractors, headed by Swiftsure Spatial Systems Inc., have written Best Map software, under the direction of TEC. These software developments have helped us test and refine the ideas presented here. The authors wish to thank the contractors and their colleagues at TEC for the many lively conversations that have contributed to this work.

INCREASING INTEROPERABILITY WITH CEONET TECHNOLOGY USING WSDL AND SOAP

N. Grabovac

MacDonald Dettwiler, 13800 Commerce Pkwy, Richmond, BC V6V 2J3, Canada – ngrabova@mda.ca

KEY WORDS: Interoperability, Spatial Infrastructures, Databases, Software, Internet/Web, CEONet, SOAP, WSDL

ABSTRACT:

The future World Wide Web will be composed of interoperable, distributed software components called Web Services. These services will be capable of automatically discovering and invoking one another, allowing complex applications to be created from collections of interacting Web Services. Two new Web technologies that help make this possible are Web Services Description Language (WSDL) and the Simple Object Access Protocol (SOAP). A network of interoperable Web Services that dynamically interact with one another to perform a host of geoprocessing activities will likely form the technological building blocks enabling the next generation of Spatial Data Infrastructures. Although CEONet Technology already provides programmatic access to its services, this access is based on a mix of proprietary and standard mechanisms. Recasting CEONet Technology and its partners as a collection of Web Services that use WSDL and SOAP would increase interoperability in the Canadian Geospatial Data Infrastructure while reducing dependencies on proprietary technology. This paper describes WSDL and SOAP and how they can be used to transform CEONet Technology from a web application to a collection of standards-based, interoperable Web Services.

1. INTRODUCTION

The first step toward building a Spatial Data Infrastructure (SDI) involved placing databases of geospatial data online and making them accessible via standard protocols. Once there were many such databases, a service was required to find them. Under the direction of the GeoConnections program, and in conjunction with the Canada Centre for Remote Sensing (CCRS), the Canadian Earth Observation Network (CEONet) was created to serve as a directory of these geospatial data sets. Over time, it has evolved into the technology that drives the Discovery Portal for the Canadian Geospatial Data Infrastructure (CGDI), enabling registration, discovery and access to a variety of geospatial resources (Smith and Fraser, 1999; Fraser and Smith, 1998; Smith, 1998).

In addition to access via a Web browser, CEONet Technology (CT) also supports programmatic access through an application programming interface (API) and a set of Re-Usable Components (RUC). The development of CT has, as much as possible, been based on existing and emerging technologies and standards (Smith and Fraser, 1999; Fraser and Smith, 1998). The API and RUCs allow software developers to create web applications that use CT services (Katic, 1998), which is the first step toward supporting interoperability with other Spatial Data Infrastructures (SDI). As the capabilities of CT have expanded, so too has the demand to provide greater programmatic access to these capabilities. If a network of cooperating and interacting SDIs is to become a reality, the ability of applications to interoperate must increase dramatically.

The World Wide Web is evolving. It is no longer simply a massively distributed information repository. The Web is now also a platform for internet-based applications, such as the Discovery Portal, and will soon support large numbers of distributed software components, or Web Services. These Web Services are "self-contained, self-describing, modular applications that can be published, located, and invoked across the Web. Web Services perform functions, which can be anything from simple requests to complicated business processes...Once a Web Service is deployed, other applications (and other Web Services) can discover and invoke the deployed service." (Tidwell, 2000).

This vision of the Web as a collection of distributed interacting software components is one that is shared by the geospatial community (Doyle *et al*, 2001). A network of interoperable Web Services that dynamically interact with one another to perform a host of geoprocessing activities will likely form the technological building blocks enabling the next generation of Spatial Data Infrastructures. These Web Services will be capable of performing a wide variety of tasks, such as discovery of data, coordinate transformations, map and imagery visualization, etc. Powerful geospatial applications could be built by chaining together collections of such services.

The first part of this paper describes two technologies that are helping to make the Web Services vision a reality: Web Services Description Language (WSDL) and the Simple Object Access Protocol (SOAP). WSDL provides the means for a web service to describe its capabilities, location and method of invocation, whereas SOAP is a protocol for invoking Web Services.

The second part of this paper describes how these technologies can be used to transform CT from a largely monolithic internet application into a Web Services-based collection of geospatial services. Doing so would allow CT to support greater interoperability with components of the CGDI, other Spatial Data Infrastructures (SDI), and any Web applications that require the discovery and access services provided by CT.

2. CEONET TECHNOLOGY TODAY

In many ways CEONet Technology already bears some resemblance to a set of interacting Web Services. A typical set of user actions on the Discovery Portal involves: 1) discovering remote data repositories containing geospatial data of interest,

2) finding individual data products within those repositories, followed by 3) accessing that data in some manner – downloading, viewing, ordering, etc. Along the way, the user can make use of a map viewer, postal code look-up tool and place name look-up tool to refine the geographical area of interest for their search. The map viewer can also be used to display the geographic areas covered by the data products that were discovered (Figure 1).

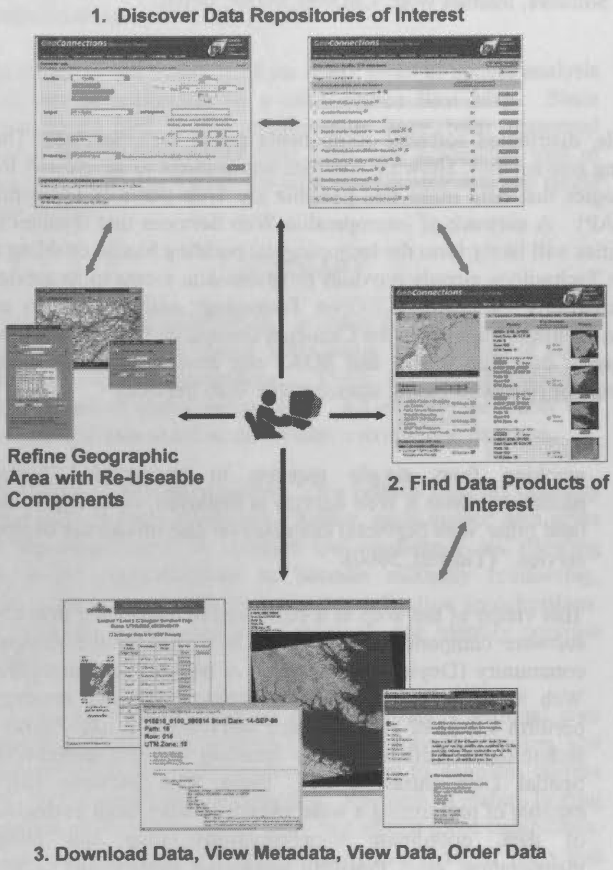

1. Discover Data Repositories of Interest
Refine Geographic Area with Re-Useable Components
2. Find Data Products of Interest
3. Download Data, View Metadata, View Data, Order Data

Figure 1. Discovery and access of geospatial data using CEONet Technology services

We can think of steps 1 to 3 as separate services that are chained together to provide the user with discovery and access services for geospatial information. The discovery services provided by CT are chained with the access services provided by the data repository. This integration of data discovery services with distributed data access services is performed using a mixture of standard and proprietary protocols.

When a data provider registers their information with the Discovery Portal, they provide the location of their data repository and the protocol to use to access their data. These connectivity details are collected and stored in a proprietary format by CT and are used to chain together the CT data discovery service with the data access services supported by the provider. Access types can include the display of browse images, thumbnail images, metadata, downloading of data, online ordering and viewing the data provider's website, all of which may potentially utilize different protocols and return their results in different formats. CT hides this complexity from the user and provides a consistent user interface (Smith and Fraser, 1999).

Web technologies and specifications have advanced to the point where a proprietary solution to chaining together disparate services, such as discovery and access, is – or will soon become – unnecessary. In order to support greater interoperability with other components of the CGDI, as well as other SDIs, these CT discovery and access services can be transformed into Web Services that are described using a standard specification and can be invoked using standard protocols. The next sectión will focus on two of these technologies: Web Services Description Language (WSDL) and the Simple Object Access Protocol (SOAP).

3. WSDL AND SOAP

The Web Services architecture is based on the dynamic discovery and invocation of relevant services. This requires that Web Services are both self-describing and accessible via a standard protocol. Some of the largest software companies in the world, including IBM and Microsoft (in addition to numerous smaller hi-tech companies) have jointly developed a set of specifications for this purpose. Two of these specifications: Web Services Description Language (WSDL) and the Simple Object Access Protocol (SOAP) have gained widespread industry acceptance and vendor support. As a result, they are considered de-facto standards in the Web Services community. Both WSDL and SOAP have been submitted to the World Wide Web Consortium (W3C), an organization concerned with developing standards for the Web.

3.1 Web Services Description Language

When a Web Service goes live, it is registered in a service registry, with a description of the service that includes what it can do and where and how to access it. Web Services can then use these registries to discover the other services they need to collaborate with. The invoking service uses the description of the other service in the registry to determine how to invoke that service and what response to expect in return. (Zaev, 2001). This description is specified in a standard way by using Web Services Description Language (WSDL).

A WSDL document contains sections that describe what a Web Service can do – the operations it can perform – the protocols used to access the service, and where to access the service. These descriptions are written in XML. The operations are defined in an abstract manner, without reference to the access protocol or URL used to access the operation. This allows operations to be bound to different physical access protocols – such as HTTP GET/POST or SOAP (see below) – without changing or duplicating the operational definition. The binding of the abstract operation to a concrete protocol and URL are specified in a different section of the WSDL document. A single operation can be bound to more than one protocol and access point, thus increasing flexibility in how a service can be invoked (Cauldwell, 2001; Christensen *et al*, 2001).

Descriptions written using WSDL allow an application to dynamically determine the protocol and URL for accessing a Web Service, as well as the parameters required by the operation and the format the results will be returned in. This allows Web Services to chain themselves together in real time to perform actions on behalf of a user.

3.2 Simple Object Access Protocol

Once a Web Service has been discovered and its WSDL description used to determine its capabilities, where the service resides and how to access it, the next step is to access, or invoke, the service. Although there are many possible access protocols that could be defined in the WSDL description of the service, the Web Services community has adopted the Simple Object Access Protocol (SOAP) as the standard method for invoking a Web Service.

Like WSDL, SOAP is written in XML. It defines a specification for how XML messages can be passed over a network. Because these messages can contain descriptions of operations and their parameters, SOAP can be used like a remote procedure call to invoke the functionality of a Web Service over the internet (Dix, 2001; Box *et al*, 2000).

Using SOAP as the standard protocol for accessing Web Services does not tie the user to any particular transport method. SOAP can be used over HTTP, SMTP, FTP or any other protocol for which a SOAP binding has been specified and implemented. Current implementations of SOAP use all of the above transport methods. Because SOAP was designed to be extensible, it should be able to bind to any future protocols as well.

A SOAP message consists of a body, an optional header and an envelope. The body is the payload of the message. It contains the XML document to transmit, or the details about the Web Service operation to invoke. The header is used to encapsulate any extensions to the SOAP protocol, such as transaction support, encryption, support for binary attachments, etc. It also contains routing instructions that describe the chain of intermediaries the SOAP message must pass through and what actions those intermediaries should take. The envelope serves as the container of the header and body.

The Simple Object Access Protocol allows for one-way communication, request-response communication as well as arbitrarily complex chains of communication, with multiple intermediaries between the sender of the message and the receiver. At each step in the chain, processing can take place, making SOAP an ideal protocol for constructing Web Service chains.

4. CEONET TECHNOLOGY TOMORROW

Moving from a proprietary method of service chaining to a standards-based approach using WSDL and SOAP could be accomplished as follows:

1. CEONet Technology (CT) services such as data discovery, place name look-up, postal code look-up and map viewing are described using WSDL. These WSDL descriptions are made available to other applications, perhaps using service registry.
2. The CT discovery services are SOAP-enabled, allowing other applications and Web Services to access these services using SOAP.
3. When a data provider registers their information with the Discovery Portal, they also provide a WSDL description that defines the access capabilities of the data provider, how those capabilities should be invoked, and the URL to send requests to.
4. The data provider provides a SOAP interface for invoking these access services.

Imagine the following scenario:

A market researcher needs to produce a map of housing price changes in Vancouver, BC over the last 10 years. Using a custom-built application, she enters the postal code of interest and how much she is willing to pay for the data. The application sends a message to CT asking for its capabilities. CT returns a WSDL document to the application. The application parses the document and discovers the postal code look-up service. It invokes this service to get the Lat/Long coordinates it needs to perform a search for air-photos of Vancouver in that region. It then sends a SOAP request to the CT discovery service (also described in the WSDL document) to search for air photos of Vancouver that cover the geographic area specified by the Lat/Long coordinates provided by the postal code look-up service. A SOAP response is sent back to the application, detailing the data repositories that have air photos with the required coverage, which repositories support online access to the data, how much data access costs and how reliable their access services are. The application selects a data repository based on cost and reliability criteria. At this point, the application could send another SOAP request to CT to download the air-photos (in this case, CT acts as an intermediary between the application and the data repository), or it could request the WSDL description of the data repository of interest and invoke the access service itself. The application then retrieves housing price data for the same postal code from a different Web Service. Finally, the application sends a request to a visualization Web Service to display the price data overlaid on the air photo. The resulting map is displayed to the researcher.

During this scenario, once the research application has a description of CT's capabilities as a WSDL document, it is able to invoke the services it desires without using any proprietary protocols. The application creates a service chain that uses CT Web Services to discover and access the geospatial data of interest, and uses other Web Services to combine this geospatial data with other relevant data for display to the user. A complex chain of Web Services has been created to perform an activity that has value to its user.

5. CONCLUSIONS

Web Services are changing how people and applications will make use of the World Wide Web. The increase in interoperability provided by new Web technologies such as WSDL and SOAP will also be a boon to the geospatial community and Spatial Data Infrastructures. Existing geospatial web applications, such as CEONet Technology, can be recast as Web Services that use WSDL and SOAP. Doing so makes it possible to create complex geoprocessing applications by dynamically chaining together a group of interoperable geospatial Web Services, thus moving a step closer toward implementing some of the goals envisioned the Canadian Geospatial Data Infrastructure and other Spatial Data Infrastructures.

6. REFERENCES

Box, D., Ehnebuske, D., Kakivaya, G., Layman, A., Mendelsohn, N., Nielsen, H. F., Thatte, S., Winer, D., 2000. W3C Note *"Simple Object Access Protocol (SOAP) 1.1"*, DevelopMentor, International Business Machines Corporation, Lotus Development Corporation, Microsoft, UserLand Software. http://www.w3c.org/TR/SOAP (accessed 24 Feb. 2002).

Cauldwell, P., 2001. WSDL: The Web Services Definition Language. In: Cauldwell P, Chawla R, Chopra V, Damschen G, Dix C, Hong T, Northon F, Ogbuji U, Olander G, Richman MA, Saunders K, Zaev Z, *Professional XML Web Services*, Wrox Press, Birmingham, pp. 109-147.

Christensen, E., Curbera, F., Meredith, G., Weerawarana, S., 2001. W3C Note "Web Services Description Language (WSDL) 1.1", Ariba, International Business Machines Corporation, Microsoft. http://www.w3.org/TR/wsdl (accessed 24 Feb. 2002).

Dix, C., 2001. SOAP Basics. In: Cauldwell P, Chawla R, Chopra V, Damschen G, Dix C, Hong T, Northon F, Ogbuji U, Olander G, Richman MA, Saunders K, Zaev Z, *Professional XML Web Services*, Wrox Press, Birmingham, pp. 77-107.

Doyle, A., Reed, C. (eds), Harrison, J., Reichardt, M., 2001. OGC Interoperability Program White Paper, "Introduction to OGC Web Services", Open GIS Consortium. http://ip.opengis.org/ows/010526_OWSWhitepaper.doc (accessed: 24 Feb. 2002).

Katic, N., 1998. Reusable Components In Development of Geo-spatial Software. In: *Proceedings Earth Observation and Geo-Spatial Web and Internet Workshop (EOGEO'98)* 17-19 February, 1998, Salzburg, Austria. http://www.sbg.ac.at/geo/eogeo/ (accessed 24 Feb. 2002).

Smith, S., 1998. Low Cost Distributed Geospatial Searching. In: *Proceedings Earth Observation and Geo-Spatial Web and Internet Workshop (EOGEO'98)* 17-19 February, 1998, Salzburg, Austria. http://www.sbg.ac.at/geo/eogeo/ (accessed 24 Feb. 2002).

Smith, S., Fraser, D., 1999. Rising Above a Quagmire of Search Protocols. In: *Proceedings of the Earth Observation and Geo-Spatial Web and Internet Workshop (EOGEO'99)*, 9-11 February, 1999, Washington D.C., USA. http://webtech.ceos.org/eogeo99/Papers/rolker/rolker_eogeo99.html (accessed 24 Feb. 2002).

Tidwell, D., 2000. IBM devloperWorks Web services tutorial "Web services – the Web's next revolution". http://www-105.ibm.com/developerworks/education.nsf/webservices-onlinecourse-bytitle/BA84142372686CFB862569A400601C18?OpenDocument (accessed 24 Feb. 2002).

Zaev, Z., 2001. Evolution of Web Services. In: Cauldwell P, Chawla R, Chopra V, Damschen G, Dix C, Hong T, Northon F, Ogbuji U, Olander G, Richman MA, Saunders K, Zaev Z, *Professional XML Web Services*, Wrox Press, Birmingham, pp. 11-57.

USING SCHEMA AND DATA INTEGRATION TECHNIQUE TO INTEGRATE SPATIAL AND NON-SPATIAL DATA : DEVELOPING POPULATED PLACES DB OF TURKEY (PPDB_T)

Abdulvahit Torun

General Command of Mapping (GCM), Cartography Department, 06100 Cebeci, Ankara, Turkey – atorun@hgk.mil.tr

KEY WORDS: Spatial, Databases, Integration, Schema Integration

ABSTRACT:

Databases (DBs) are designed from scratch due to application specifications of organizations. Sometimes the same real world object occurs in different semantics in different DBs. DBs may be autonomous on different computers and computers may have different operating systems, DBs may be designed for different purposes and applications and DBs may use different data formats. In order to use legacy data instead of re-collecting the same content, data in a wide variety databases are needed to be integrated. Data integration is done either gathering data without any integration or by integrating several source data partly or as a whole into a single DB. The method, integrating several DBs into one single DB, is implemented while developing Populated Places DB of Turkey (PPDB_T). To implement the integrated DB, information about populated places except population was collected from a plain table, population information are extracted from State Statistics Institute (SSI) DB and administrative boundaries are taken from Digital Chart of Turkey (DCT) at scale 1:1000000. Ladder technique is applied for schema and data integration. Although the data is stored in and managed by a non-spatial DB system, in order to improve geometric accuracy of location information, a subset of PPDB_T is transformed into a spatial DB and location information are corrected with the help of standard topographic map index and administrative boundaries of provinces and districts. This need forces mapping a subset of the non-spatial DB containing location information into spatial world and integrate it with DCT to do corrections and transfer the corrected location and information back to PPDB_T.

1. INTRODUCTION

Organizations have a large variety of databases to conduct their everyday business. Large organizations have their databases on different platforms. Usually, databases are designed from scratch due to application specifications of organizations. Sometimes, the same real world object occurs in different semantics in different DBs. Some users store very needed data independently which are existing in the database of organization (Elmagarmid et.al., 1999).

As organizations have become sophisticated, data management and data sharing become a mess. With the enabling technology for networking, distributed computing and communication, data and process sharing become applicable. Connecting and sharing sources among existing DBs forces to deal with autonomy and heterogeneity. These DBs may be autonomous on different computers and computers may have different operating systems, may be designed for different purposes and applications and may use different data formats.

DBMSs and applications are designed considering incorporate data and process sharing to ease autonomy. Main efforts are spent in three directions; 'schema integration', 'federated DB' and 'multi-database language' (Elmagarmid et.al. 1999, et.al. Ozsu et.al. 1999). Schema integration is defined as integrating component DBs at schema level. Schema integration is translation of a schema from a data model into another data model manually, automatic or with both techniques. Although, federated DB approach allows more autonomy and flexibility in heterogeneity, partial schema integration is done repeatedly on-the-fly and the user needs master help. Multi-DB language approach favors autonomy over heterogeneity such that a common multi-DB language that enables accessing and manipulating disparate DBs (Elmagarmid et.al., 1999).

PPDB_T is a non-spatial relational DB containing information about populated places. The relationships among entities resemble the hierarchical relationship among populated places in the real world. In addition to conventional use of databases in different applications, PPDB_T is usually used for topographic map production by GCM, Turkey. Although, thematic correction is finished, geometric enhancement of populated places is being continued. PPDB_T is being designed as a central DB. Integration with State Statistics Institute (SSI) DB and other spatial DBs are accomplished by developing programs in Application Programming Interface (API) of Microsoft Access, ESRI Arc/Info and ArcView coding languages. In the following chapters, concepts on database integration and DB integration strategy applied for PPDB_T are given.

2. DATABASE INTEGRATION

A distributed DB is developed from scratch by using a top-down methodology. However, in most of the cases the tendency is management of legacy systems and re-use of data in recent years. The DBs, which are distributed geographically, in different models, in different environment and with different semantics are needed to be integrated to use the available data. This problem is called as 'DB integration'. This type of problem yields more difficulty compared to distributed DBs design. DB integration is a bottom-up work contrary to distributed DB design, which is a top down work.

The problems caused by semantic heterogeneity and design autonomy (heterogeneity) should be resolved to achieve DB integration. Semantic heterogeneity is defined as perceiving the same real world phenomena differently with respect to different viewpoints of applications. Design autonomy is a matter of schema and syntax of DBs. Schema defines how the features are expressed within the DB. The discrepancy within the type of DBMS, API used in development and the data model that the DBMS uses to express the features defines syntax autonomy.

Database integration is the process by which information from participating (component) databases are conceptually integrated to form a single cohesive multi-DB -in other words Global Conceptual Schema (GCS)-. Therefore, DB integration is unifying existing data stores into a single framework (Figure 1).

Here are some concepts, which are frequently used in database integration to make the discussion clear. Local Internal Schema (LIS) is the schema of individual internal schema of DB. Local Conceptual Schema (LCS) is logical organization of data at each site. External Schema (ES) is the schema for a user application. Global Conceptual Schema (GCS) is enterprise view of the data-logical structure of the data at all the sites. Federated Conceptual Schema (FCS) is combination of a set of export schemas of individual DBs. Data dictionary is information (or a DB) about the DB schema, relations, attributes, domains of attributes, relationships of a DB. A data dictionary can be implemented as a metadata. GES, LES stand for Global External Schema and Local External Schema respectively (Figure 1). LES and GES are the views for local and external users respectively (Bobak 1996, Ozsu et.al. 1999).

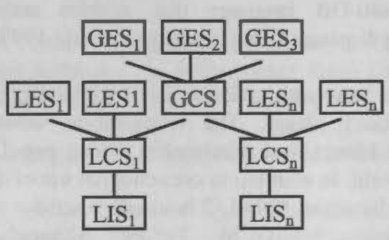


Figure 1. Schema Hierarchy in Database Integration and relationship within local and global schemas

2.1 Database Integration Strategies

DB integration implies uniform and transparent access to data managed by multiple DBs. The mechanism is based on an integrated schema which involves all or part of the schema of component DBs. Heterogeneous distributed DBs are integrated and re-designed as a whole by means of global schema integration, federated DBs and multi DB language approaches.

There are three types of DB integration strategies; firstly integrating the data, secondly schema integration and thirdly developing software (Devogele et.al. 1998).

Data Integration

A very basic approach is providing a global catalogue of accessible information sources to user to allow him to do integration by himself without attempting any integration. Each data source may be described as a metadata containing information about representation model, scale, last update date, data quality etc. Most of the data warehouses on the web are working in this way. The user should search, find, decide and find tools to select, extract, integrate and query the data.

A further step is integrating the source DBs as one single DB by putting the data together. This approach violates the basic rules of distributed DBs such as data distribution, data autonomy (independence). The data and applications should be converted into the integrated DB. Integration is accomplished in two ways; either integrating existing data by hand or extracting relevant data from data sources in suitable form to be integrated centrally.

For the former method, the application is split into small pieces each of which accesses only a single data store. Then, the local data is carried to the site of global application. For the latter method, data is extracted from the data source to make a single DB such as EuroGlobalMap (former MEGRIN project) or SABE of EuroGeographics. DB integration based on data integration is lack of scalebility. Every time DB is enlarged, the same integration processes should be followed from the beginning. Moreover, it yields duplication. The same data resides at its local DB and at the integrated DB. Therefore, maintaining the consistency becomes difficult.

Schema Integration

Source DBs can be integrated by means of schema integration or using standardized data or schema models. Schema integration is a practical method such that the data is integrated as a logical DB with a global schema. With a given set of local DB schemas belonging to an individual DBMS, an integrated schema which subsumes those local schemas is created by synthesizing the schemas. There are difficulties in schema integration, because of different structures and semantics among local DBs. The user is not aware of the underlying systems, the whole system imitates as if one single DB. Those integrated DBs of this type are known as multi-DBs.

Standardization of data and models provides a base for integration. Standardization of data modeling and manipulation features enables exchange of these data and methods among heterogeneous systems.

Data model standards define which abstract modeling concepts that are to be used for modeling the real world. For non-spatial data, there are standards for relational and Object Oriented data models. Several spatial data modeling standards are defined by ISO (TC 211), by CEN(TC 287) and OpenGIS. Each of these standards provides a view for different aspect of data.

There are schema standards consisting of data and process design for specific application areas, which are defined by OpenGIS. These standards define a common target (canonical model) for data conversion and interoperability. But they don't define how to convert existing data into standards and how to integrate data.

These standard data and design models provide a common understanding to squeeze the interoperability – schema, format and semantics- problem into semantic problems.

Software Developing

Software can be developed to support data interoperability among diverse DBs. The software may work in three different ways; as a gateway, as a persistent view and as a tool to construct a Federated DB. A gateway interface connects different DBs and allow accessing data in other systems. Although, defining a persistent view over several DBs allow access to distant data, consistency constraints can not be defined. Connection to distributed data is limited with the definition in the view. Federated DB allows combining and scaleable integration of distributed data. Site autonomy is not violated within a federated DB.

Federated (virtual) DB is defined by creating a schema representing component DBs. However, data instance is at its original local site and is created temporarily on-the-fly when required. Local DB administrators specify a subset of DB for FDB users. Import/export into/from FDB is managed by federated system on the basis of standard data model and manipulation language.

Federated Schema is a less loosely coupled multi-DB through a federated schema. There is no global schema that comprises component schemas. A user defines his own schema based on the export schemas of component DBs that are allowed for sharing. The Component DBMSs may have different models creating heterogeneity, which must be solved during integration. Federated schema is a union of export schemas of Component DBs, $Federated\ Schema\ = \bigcup ES_{Component\ DB}$

Gateway and View approach provide the user with a multi-DB access language (SQL) without any unification of the semantics of data from various sources. Federated DB (FDB) is an integrated view of data which is managed by the FDB. Users access FDB like a centralized DB without worrying the localization of data or without worrying about the type of DBMS.

2.2 Integration Methodology for PPDB_T

A hybrid integration technique is applied in order to develop Populated Places DB of Turkey (PPDB_T) (Torun, 2000). Data integration method -integrating several DBs into one unique DB- and schema integration method are used for integration. Data insertion from the component DBs is done by running the programs written in API of PPDB_T. Schema and Data integration is accomplished by using ladder technique in which first of all, schema of newly designed PPDB_T of GCM is integrated with the data collected for populated places which are stored in a plain table. Then, schema of PPDB_T is extended to integrate it with SSI DB schema.

For integrating different data sets describing the same phenomena, first a correct understanding of the semantics of the existing data should be developed. Schemas are transformed into a common data model. For instance SSI DB schema and DB instance are converted into SQL and dbf respectively. Then, an accurate correlation among structures (schemas) is established to avoid comparing different type of objects within the same category. The inter-schema correspondence at meta level and inter-DB correspondence at data level are identified and described. Finally, integration is described precisely to prevent merging irrelevant data together. The conflicts are solved semi-automatically. Integrated schema is generated on top of contributing data sources (Figure 2). Schema integration is done by using modified 5 level integration architecture for the purpose of single DB generation (Abel et.al. 1998, Ozsu et.al. 1999).

Modified 5 Level Schema Integration Architecture

1. Local Schema: Local schema of plain table, SSI DB and DCT are made available by the Component DBs.

2. Component Schema: Representation of Local schema in canonical (standard) data model is defined by using SQL. Canonical data model is employed for unifying divergent local schemas in a single schema.

3. Export Schema: Export Schemas are defined from component schemas based on integrated global schema of PPDB_T.

4. Schema Integration: Export Schemas are integrated to form a single schema. Since the desired final DB is not a federated one, the integrated schema is not a Federated Schema, either. Firstly, two component schemas are integrated. Then, the third one is added to the integrated schema. This method is called as ladder technique. Data integration follows schema integration (Figure 3).

5. External Schema: Upon the integrated schema different conceptual/external schemas are defined for different purposes and usage.

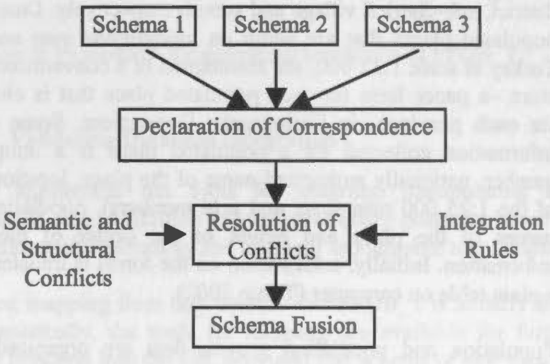


Figure 2: Integration methodology for constructing PPDB_T (After Devogele et.al, 1998)

3. DEVELOPING POPULATED PLACES DATABASE OF TURKEY (PPDB_T)

Information about populated places is used in topographic map production for typography and feature placement in GCM. Population information and relevant statistics are collected and stored by SSI. In order to maintain population information up-to-date for map production, DB integration is necessary to share the legacy data by SSI instead of re-constructing the same content. Content, definition and thematic granularity and accuracy of both non-spatial DBs are considering the administrative hierarchy of Turkey (Torun 2000).

PPDB_T contains information about inhabited places such as its name, its former names, its administrative status, its hierarchy within administrative system, location information of its center, its population etc.

Data is gathered from three different sources. Initially, the information about populated places except population was collected and stored in a plain table on computer. Population information is extracted from State Statistics Institute (SSI) DB.

Finally, administrative boundaries are taken from Digital Chart of Turkey (DCT) at scale 1:1000000.

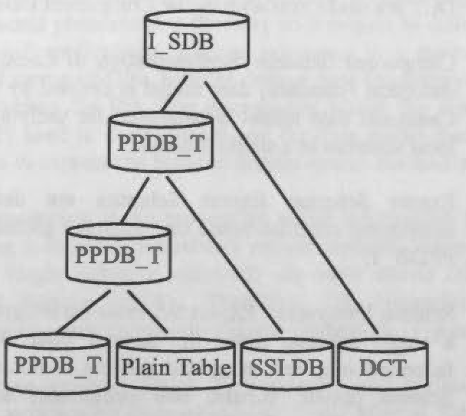


Figure 3: PPDB_T is developed by using ladder method with three steps

3.1 Defining Associations

Turkey has a hierarchical administrative system, which looks like a balanced tree structure. The sequence of residential entities (populated places) from top to bottom is province, district, sub-district, village and suburb respectively. Data about populated places that are being on topographic map series of Turkey at scale 1:25 000, are maintained in a conventional data store –a paper form for each populated place that is clustered for each province- in Cartography Department. Some of the information collected for a populated place is a unique ID number, nationally authorized name of the place, location (title of the 1:25 000 map sheet and grid numbers), population, old names of the place and height of the center of the town information. Initially, information on the forms is transferred in a plain table on computer (Torun 2002).

Population and population growth data are organized as a spreadsheet by SSI. The provinces are ordered due to plate number sequence. Rest of the populated places are ordered considering alphabetic order at each level within its parent unit. Cumulative data are added into the list representing for higher level populated places. Suburbs, which are not considered as a unit, are not existing in SSI DB.

3.2 Removing Schematic and Semantic Conflicts

Organization and hierarchy of populated places in data stores of both GCM and SSI are almost the same. Main difference comes out of schema and format. SSI DB is modified and DB schema is enhanced in order to prepare the data for integration. Data is transferred into a common schema and format (dbf format) for which a middleware is developed to transfer population data (Figure 4).

Semantic heterogeneity is resolved by using single semantics and preservation of data consistency and integration is provided by resolution of naming conflicts. With schema integration, the problems of name conflicts, schematic differences, missing data are resolved. A multi-word name in one DB might be a combined word in the other or different names –old and new names or multiple names for a place- are assigned for the same place. These are partially resolved with computer programs. A populated place might be clustered in a false class in one of the DBs. After resolving those conflicts given above, two data sources are integrated considering integrated PPDB_T schema.

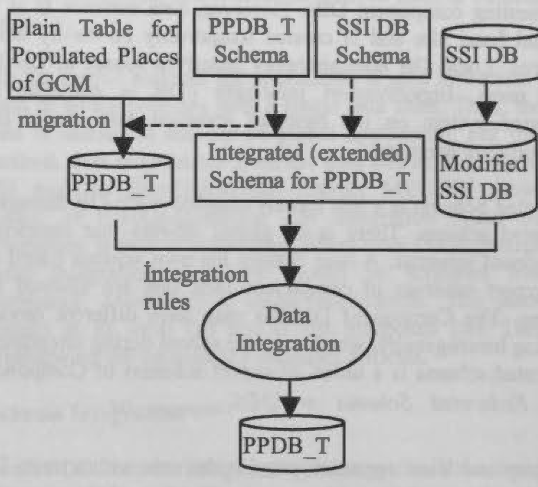


Figure 4: Integration procedures to construct PPDB_T

3.3 Database Integration to build PPDB_T

Schema Integration

First of all, database schema of PPDB_T is designed by means of actual and future needs. DB schema of PPDB_T is designed considering the administrative binding of populated places. Then, relevant attributes of SSI DB are added to the PPDB_T schema in addition to a foreign key, which provides PPDB_T to connect SSI DB.

A subset of PPDB_T is exported to generate a spatial DB consisting of points of populated places. This export schema is integrated with DCT schema to do spatial analysis and to do corrections on spatial and non-spatial attributes of populated places. When correction is finished, information in the spatial DB is transferred back to PPDB_T for improvement. There happens no schema integration.

Schemas of two DBs are different. DB schema of GCM for Populated Places is designed by using relational model. However, Population DB of SSI is a table containing all kind of quantitative values and statistics. DB schema of SSI is migrated into schema of PPDB_T of GCM. This process prevents scalebility of the resultant schema and DB instance. Each time there happens a change, the integrated schema and integrated DB should be updated.

Data for spatial DB is extracted from PPDB_T in a common model for both DBMS and GIS software -ESRI Arc/Info- to transfer data back and forth. These are done by developing a tiny software which extracts data from PPDB_T into a common model, transforms the common model into a spatial DB and visa-versa. Different languages are employed for spatial and non-spatial definitions and manipulations.

Data Integration

Data integration process is done in four major steps. Data integration is accomplished by applying ladder technique. Firstly, DB schema of PPDB_T is designed considering the available non-spatial digital information sources and further needs. PPDB_T is a non-spatial relational DB containing location information, which is used to produce spatial data. Location of a populated place is represented at the center of gravity of the settlement area. Secondly, PPDB_T is populated with the data from plain table. Therefore, the plain table is mapped into PPDB_T. Thirdly, PPDB_T schema is extended in order to import population information from exported data of SSI DB.

Each populated place in SSI DB is searched in PPDB_T by tracing the path through the hierarchical administrative path by using type of the feature. As the populated place is located, value of population attribute is updated and a unique ID value generated from the traced path for the populated place in SSI DB is stored as a foreign key. If the populated place is not met in DB, user resolves the conflict.

The new spatial DB namely derived Spatial DB (d_SDB) is generated by means of a non-spatial predicate, which has spatial meaning when transformed into spatial world. For instance, a predicate 'those places that are bound to province Ankara' yields a non-spatial set of populated places. However, each place has a location on earth. The coordinates of those places help to transfer the set of tuples into spatial world.

Derived Spatial DB (d_SDB) is based on a common schema for both PPDB_T and spatial data processor -for the time being ESRI-ArcInfo- that will import the data. Export schema is created by means of a non-spatial predicate that cuts the DB both vertically -a subset of attributes- and horizontally -a subset of tuples-. The exported non-spatial data is mapped into spatial format to generate d_SDB (Figure 5).

However, primary key is repeated in every fragment in vertical fragmentation. Thus, disjointness is valid only on non-primary key attributes in vertical fragmentation.

DCT contains a set of spatial data classes such as administrative boundaries, hydrology, transportation, elevation, populated places (only provinces and sub-provinces), physiography.

The relationship among PPDB_T and d_SDB is preserved by keeping the same primary key -Populated Place ID- in corresponding relations of both DBs.

The first level analysis of geometric correction is to search for 'Are the populated places bound to province Ankara being inside the administrative boundary of it' or 'Are the populated places located within the map sheet K being inside the boundary of mapsheet'

The final step is to design an integrated spatial DB (I_SDB) with the three component DBs; PPDB_T, DCT and spatial DB derived from a subset of PPDB_T.

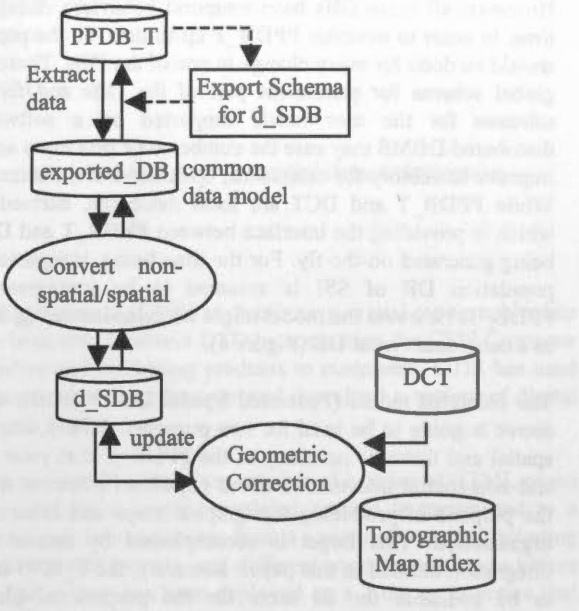


Figure 5: Procedures for geometric correction of PPDB_T.

3.4 Validating Database Integration

The application has tools for entering, manipulating and updating data in addition to intelligent query generator based on administrative hierarchy and standard topographic map indexes.

Since, mapping from data sources into PPDB_T is initially done automatically, the tools for mapping are available for further data injection into the DB, provided that schema of new data set is integrated with schema of PPDB_T. Moreover, there are tools to check consistency considering administrative hierarchy and to compare different versions of PPDB_T.

Problems of Integration

Constructing an integrated DB from existing DBs yield some problems due to lack of interoperability among DBs.

GCM and SSI have different non-spatial data to an extent for the same context. The data differs because of partly semantic but mainly schematic and format (syntax) discrepancy. For instance, in sub-district and village level of administrative hierarchy, a sub-district might be classified as village or visa versa in two non-spatial DBs. Updateness is the main reason for this kind of anomalies. If one of the DBs stores a different name from the current one for a populated place the corresponding populated places can not be matched till the mistake is removed or association is built.

Although, formats of both DBs are different, the format of SSI DB (plain text or spreadsheet) can be transformed into a common format such as dbf that is an open format for API of DBMS of PPDB_T.

Future Work

As PPDB_T is a central DB, communication, conversion, transfer, integration among PPDB_T and other spatial and non-spatial DBs are succeeded by using developed software. However, all these DBs have temporal behaviors changing in time. In order to maintain PPDB_T up-to-date, all the processes should be done for every change in one of the DBs. Therefore, a global schema for permanent part of the DBs and federated schemas for the user needs supported by a software or distributed DBMS may ease the cumbersome processes and may improve scalebility for extensions, updateness and maintenance. While PPDB_T and DCT are local databases, derived SDB, which is providing the interface between PPDB_T and DCT, is being generated on-the-fly. For the time being, integration with population DB of SSI is assumed to be immigrated into PPDB_T. However the model might be extended taking SSI DB as a component local DB (Figure 6).

The federated model (Federated Spatial DB (F_SDB)) defined above is going to be used for two purposes. Firstly, improving spatial and thematic accuracy of the PPDB_T that yield spatial and non-spatial information about populated places is used for the purpose of producing topographic maps and others in the organization. This target is accomplished by means of DB integration defined in this paper. Secondly, the F_SDB is going to be available for all users for the purpose of planning, analysis, thematic and statistical map generation etc.

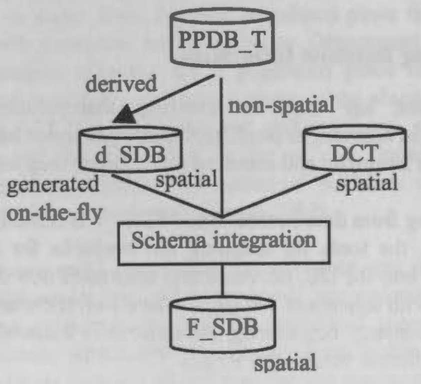


Figure 6: Federated Database model having three components.

4. CONCLUSIONS

In this paper, after introducing database integration strategies proposed in the literature, schema and data integration problems of PPDB_T are expressed. In real world applications, no modular model is fulfilling the needs. Therefore, a hybrid strategy comprising schema integration, data integration and software development in a good balance is defined for integration spatial and non-spatial databases. Schema and data are prepared for integration by transforming them into defined common models. Schema integration is done by hand. However, data integration is mostly done automatically by written software. Data transformations from non-spatial to spatial world and visa versa are accomplished by written programs. Consistency of PPDB_T is maintained considering application specifications and administrative hierarchy of Turkey. PPDB_T is going to be re-engineered to deploy the DB into distributed environment with a federated architecture.

References :

Abel., D.J., B.C.Ooi, K.L.Tan, S.H.Tan, 1998. Towards integrated geographical information Processing, International Journal of Geographic Information Science, 12(4), pp. 353-372.

Bobak, A.R. 1996. Distributed and Multi-Database Systems, Artech House, Boston, pp 121-138

Devogele, T., C.Parent, S.Spaccapietra, 1998. On Spatial Database Integration. International Journal of Geographic Information Science, 12(4), pp. 335-352.

Elmagarmid, A. et.al., 1999. Management of Heterogeneous And Autonomous Database Systems, Morgan Kaufmann Publishers, San Francisco, pp.2-32.

Hepner, P. 1995. Integrating Heterogeneous Databases: An Overview, http://citeseer.nj.nec.com/cs (accessed Dec. 2001)

Ozsu, M.T., P. Valduriez, 1999. Principles of Distributed Database Systems, Prentice Hall, New Jersey, pp. 75-101

Torun, A., 2000. Populated Places DB Project, General Command of Mapping. Internal Report, Cartography Department, General Command of Mapping, Turkey.

Torun, A. 2002. Designing Populated Places Database of Turkey (PPDB_T) by Using Relational Model, Harita Dergisi, 128 (on publish).

Acknowledgements

Populated Places Database of Turkey is designed and developed in Cartography Department of General Command of Mapping, Turkey.

ON-LINE DISTRIBUTION OF DIGITAL TOPOGRAPHIC DATA

Michel Gilbert

Natural Resources Canada, Centre for Topographic Information

mgilbert@rncan.gc.ca

Commission IV, WG IV/2

KEY WORDS: Topographic data, e-commerce, geographic information system, distribution of geospatial data, web mapping

ABSTRACT:

One of the main activities of the Centre for Topographic Information in Sherbrooke (CTIS) is distributing digital topographic data (DTD) of the Canadian landmass. For a number of years, customers have been able to obtain DTD by contacting the CTIS Customer Support Group. The team has established a procedure for processing orders and delivering products to customers. CTIS has used new Internet technologies to develop a system that enables users to search, consult, order, pay and download a variety of digital products quickly, efficiently, and fully automatically 24 hours a day, 7 days a week.

Users are able to access the data distribution system using their Internet browsers. The interface consists of a series of HTML pages that guides the customer in selecting the data that correspond to his needs. The user's completed request is transmitted to a transaction server, which extracts, converts, and delivers the data. The extracted data may come from a variety of sources, including spatial databases and files. Moreover, they can be converted into different transfer formats, and different coordinate systems before being delivered to the user's individual directory on the CTIS FTP site. The system has been deployed on a 3-tier architecture that ensures autonomy between the various system components, secure access to data sources, minimal processing on the client side, and easy system expandability in response to demand.

In conclusion, CTIS's future plans with respect to the distribution of topographic data will be presented. This discussion will include new formats and products that will be integrated into the current site, as well as new technologies that bring the development towards services to display and access geospatial information.

RÉSUMÉ:

Une des principales activités du Centre d'information topographique de Sherbrooke (CITS) consiste à distribuer des données topographiques numériques (DTN) du territoire canadien. Depuis plusieurs années, les clients peuvent se procurer des DTN en contactant l'équipe responsable du service à la clientèle du CITS. Cette équipe a mis en place une procédure pour assurer le traitement de la commande et la livraison des produits aux clients. Profitant des possibilités offertes des nouvelles technologies associées à Internet, le CITS a développé un système qui permet aux utilisateurs de rechercher, consulter, commander, payer et télécharger les différents produits numériques à distance de façon plus rapide et efficace et automatique 24 heures par jour, 7 jours par semaine.

Ce nouveau système de distribution des données est accessible à partir de navigateur Internet de l'utilisateur. L'interface entre le fournisseur et l'utilisateur consiste en une série de pages html permettant de guider le client dans la sélection des données qui répondent à ses besoins. Une fois complétée, sa requête est transmise à un serveur de transactions qui effectue l'extraction, la conversion et la livraison des données. L'extraction des données peut provenir de différentes sources de données incluant des fichiers et des bases de données spatiales et les données extraites peuvent être converties dans les différents formats de transfert et dans différents systèmes de coordonnées. La livraison des données s'effectue sur le site FTP du CITS dans un répertoire réservé à l'utilisateur. Le déploiement du système est basé sur une architecture trois tiers. Ce déploiement assure une indépendance entre les différentes composantes du système, une sécurité d'accès aux sources de données, une charge de traitement minimale du côté client et une extension facile du système en fonction de la demande.

Pour conclure, les plans futurs du CIT-S qui concernent la distribution des données topographiques seront présentés. Ce dernier volet abordera les nouveaux formats et les nouveaux produits qui seront éventuellement intégrés au site actuel, ainsi que le développement de services d'affichage et d'accès de l'information géospatiale basés sur les nouvelles technologies.

1. INTRODUCTION

The Centre for Topographic Information in Sherbrooke (CTIS) is responsible for the distribution of digital topographic data (DTD) of the Canadian landmass that it has produced. The increasing use of the Internet, the e-commerce, the use of databases to store geospatial data, and the recent development in Web mapping technologies have contributed to leading us to review our traditional ways of distributing our DTD products. Taking into account this new environment, the new CTIS data distribution system enables users to access geospatial information more quickly, safely and efficiently.

The CTIS online data distribution system has two components. The first is geared to provide continuous DTD access to customers on a subscription basis, while the second enables users to purchase DTD using e-commerce. The section below entitled "Online Purchasing and Subscriber System" deals with the various components and features of this system. The paper ends with a section on Data Access Services that describes CTIS's future plans with respect to the distribution of topographic data that focus on the access to the data using mapping services (Web Mapping Server) and feature services (Web Feature Server) based on the implementation specifications of the Open GIS Consortium (OGC).

2. PURCHASING AND SUBSCRIBER SYSTEM DESCRIPTION

The system is designed to support Online Purchasing and Subscriber data order and delivery. Online Purchasing is designed for low data volume consumers of CTIS. This mode enables users to buy DTD products via e-commerce, using a credit card or another mode of payment. Basically it's a payment per transaction approach, including lifetime data usage rights.

The Subscriber section is designed to support DTD data order and delivery for users who have previously acquired a subscription license. This license provides broader and more economical access to DTD products over a predefined region. Data usage rights, however, are restricted to the life of the subscription. The Subscriber section requires authorization (login) and licensee validation (region); no payment task is involved here since a subscription is paid only once a year (through an upfront payment). The rest of the paper will focus on system description. There will be no further reference to the Subscriber component since both subsystems are very similar.

The system gives access to various CTIS products. The user can order any of the available CTIS Digital Topographic Data (DTD), namely the National Topographic Data Base (NTDB), the Canadian Digital Elevation Data (CDED), the Updated Road Network (URN), Administrative Boundaries, and CANIMAGE. All characteristics of our products are included in the online distribution system in terms of content, data formats, metadata, and coordinate systems.

2.1 Framework Architecture

The system was developed in cooperation with the private sector. The contracting firm participated in system design and setting the look and feel of the Web site, while ensuring that Treasury Board Web interface requirements were complied with. The firm was also involved in the Java programming required to implement the different functionalities of the Web site. CTIS own contribution to this new system and its Web interface focused mainly on providing the content and the data conversion environment required for vector data format translation as well as the Webmaster, direct customer support, and system maintenance.

The system was developed entirely in Java programming language based on a Java framework developed by the private sector. The framework gives CTIS common tools for software development and system architecture that required communication between systems and software components (database, Web clients, and servers). The next diagram (Figure 1) illustrates the major framework components.

2.2 Online Purchasing Web Site Architecture

The user accesses the CTIS Web site through its own Internet browser (NetScape, Internet Explorer). The application is a thin client, accessed from HTTP and HTTPS communication protocols. The system has been deployed using a 3-tier architecture that ensures autonomy between the various system components, secure access to data sources, minimal client processing, and easy system expandability in response to demand.

The presentation level (1st tier) presents information, receives requests, and controls the user interface. The software applications used to communicate with users at this level are the Web server (Apache) and the servlet runner (Tomcat). Java servlets communicate with applications services developed as Enterprise Java Beans (EJB) components using the Internet Inter-ORB Protocol (IIOP). Servlets are Java classes used for user requests and responses.

The logic level (2nd tier) implements the business rules and is available to the presentation level. This level constitutes the central key of the system and it protects the data from direct user access and it protects the presentation level for data model changes. On the application server side, queries received from Java servlets are processed by EJB components, which access different databases at the data level through JDBC (Java DataBase Connectivity).

The data level (3rd tier) is responsible for data storage, which includes two major databases. One contains information related to our products and clients; the other contains the geospatial data available for distribution.

From a data-flow perspective (Figure 2), the user accesses the CTIS Web server using its own Internet browser. Through the use of Java servlets and EJB components, the application server interacts with CTIS clients and products metadata database. The user completes an order session by submitting the transaction. Once the data order is ready for processing, the data delivery system takes over. This subsystem extracts the required Digital Topographic Data (DTD) from the geospatial database or disks and processes them according to user requirements and products selected. The resulting data are then copied to CTIS FTP site for the user to download. Processing is generally completed within minutes, but could be longer depending on the number of files ordered. The user is advised via email that processing has been completed.

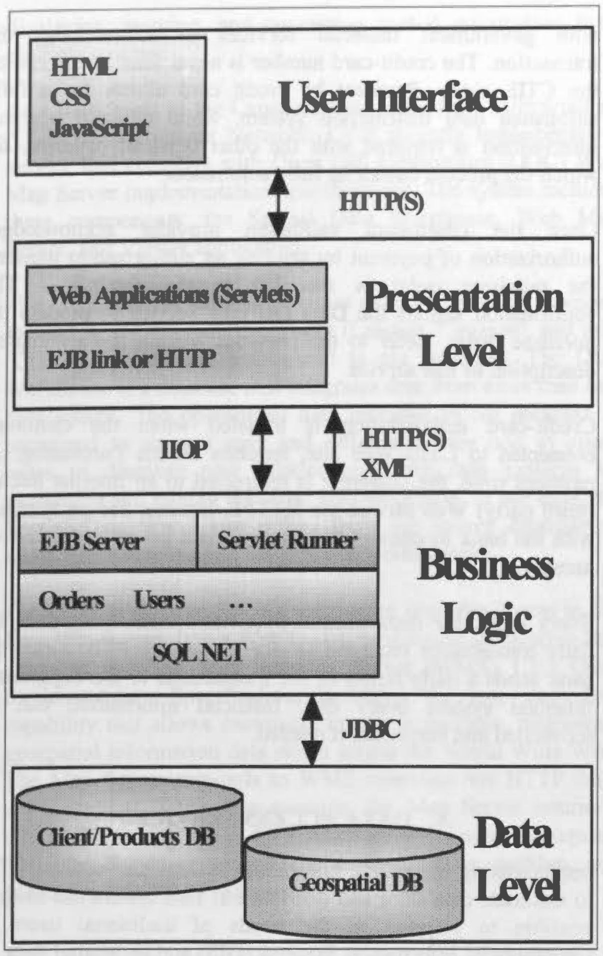

HTML
CSS
JavaScript
User Interface
HTTP(S)
Web Applications (Servlets)
EJB link or HTTP
Presentation Level
IIOP
HTTP(S)
XML
EJB Server
Servlet Runner
Orders
Users
...
SQL NET
Business Logic
JDBC
Client/Products DB
Geospatial DB
Data Level

Figure 1. 3-tier Architecture Diagram

From a data-flow perspective (Figure 2), the user accesses the CTIS Web server using its own Internet browser. Through the use of Java servlets and EJB components, the application server interacts with CTIS clients and products metadata database. The user completes an order session by submitting the transaction. Once the data order is ready for processing, the data delivery system takes over. This subsystem extracts the required Digital Topographic Data (DTD) from the geospatial database or disks and processes them according to user requirements and products selected. The resulting data are then copied to CTIS FTP site for the user to download. Processing is generally completed within minutes, but could be longer depending on the number of files ordered. The user is advised via email that processing has been completed.

2.3 Data Delivery Service

Data Delivery includes extracting the geospatial data from the database or disks, converting the data into the required GIS format, and transferring the generated files to the user's FTP directory as shown. The Data Delivery Service first receives an order ID number. It then looks for it and extracts the corresponding information from the clients and products metadata database and prepares the data conversion.

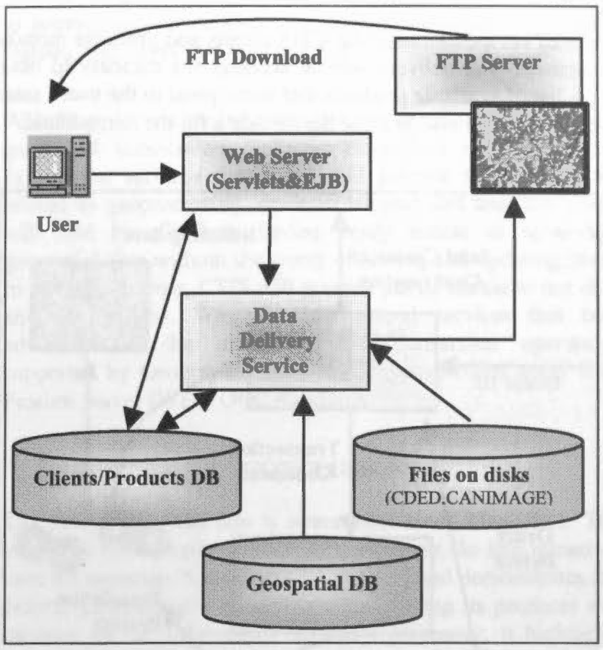

FTP Download
FTP Server
User
Web Server (Servlets&EJB)
Data Delivery Service
Clients/Products DB
Files on disks (CDED.CANIMAGE)
Geospatial DB

Figure 2. Data Flow Overview

2.4 Data Delivery Service

Data Delivery includes extracting the geospatial data from the database or disks, converting the data into the required GIS format, and transferring the generated files to the user's FTP directory as shown. The Data Delivery Service first receives an order ID number. It then looks for it and extracts the corresponding information from the clients and products metadata database and prepares the data conversion.

A transaction command (one order item) is then sent to the conversion process queue, which looks for an available conversion process to send the transaction command to. The first available conversion process will then perform the DTD conversion and the resulting data files will be created. The FME (Feature Manipulation Engine) software is used to convert vector spatial data. This subsystem can easily be expanded by adding FME licenses on additional CPUs. For the CANIMAGE product, PCI software is used to create and enhance the ortho-image.

The transaction services are controlled by a queue manager that handles how processing is shared and distributed among the available computer resources. Once data conversion is completed, the Data Delivery Service sends the user a data-delivery confirmation by e-mail.

2.5 Web Site Components

The search engine, shopping cart, identification forms, payment, and data translation are the system's five major components.

The search component allows users to search for products that are available. A user can search for a specific product using textual criteria or by browsing graphical indexes of regions of Canada. The textual criteria that can be specified are the National Topographic System (NTS) map sheet number and the date (new since). The user can use the indexes to build his or her own list of data sets by clicking directly on the map. The search component uses the CTIS clients and products metadata database. The delivery service accesses the database to obtain the list of available products that correspond to the user's search criteria. It can also provide the metadata for the items found.

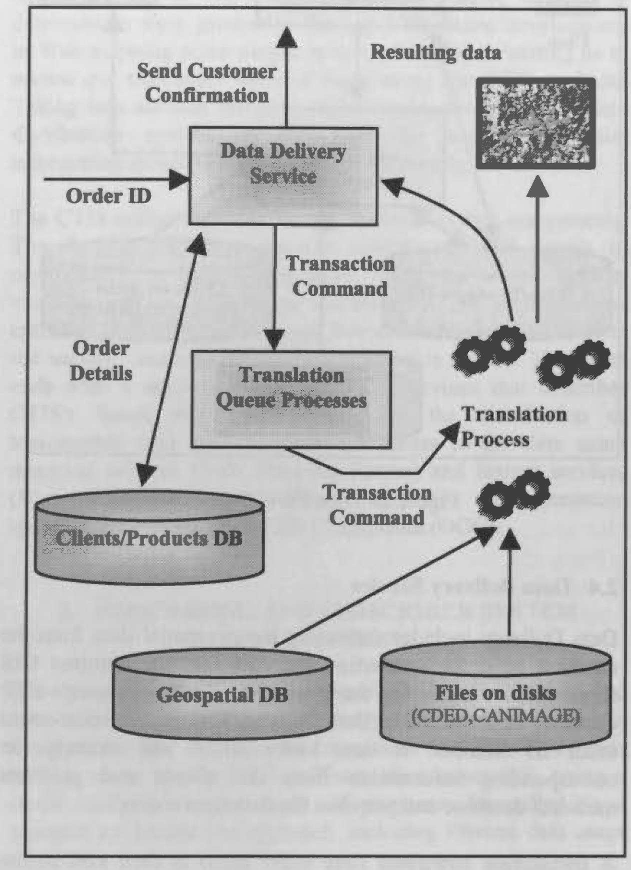


Figure 3. Data Delivery Service

The shopping cart allows users to build a DTD order, keeping track of the items selected from the different searches carried out. Each item in the cart includes the product identifier, type, version, edition, options (data transfer format, themes, etc), and price. The cart has a view function that enables users to display the geospatial data as images for most of products. Shopping-cart features also enable the user to remove items, modify an item's options, view product metadata, and see the detailed price. The cart can be saved anytime for checkout at a later date.

The identification component provides the means for gathering and recording the customer identification information required to acquire the topographic data. This component also identifies and validates users who already have a user name and password, then extracts the information needed from the CTIS clients and products metadata database to complete the purchase and deliver the data.

Five different modes of payments can be used: credit card, cheque, customer and government accounts, and other. The payment component uses a 128-bit Secure Socket Layer (SSL) and a VeriSign certificate to secure transactions. In addition, credit-card transactions are handled by a specialized electronic payment provider. The provider takes charge of the payment, performs a real-time credit validation, and serves as the link with government financial services to acknowledge the transaction. The credit-card number is never seen nor stored on the CTIS side. Payment by credit card allows for a fully automated data distribution system, while minimal operator intervention is required with the other types of payment, for which the process cannot be fully automated.

Once the credit-card validation provider acknowledges authorization of payment by sending an authorization number, the purchase order is recorded in the database. The confirmation signals the Data Delivery Service to process the purchase order. Refer to the previous section for a complete description of this service.

Credit-card authorization is initiated when the customer, connected to CTIS Web site, requests Online Purchasing. At payment time, the customer is redirected to an Internet Secure (third party) Web service via HTTPS. Internet Secure interacts with the bank to commit the payment and get an authorization number.

Every night, the departmental financial system downloads the daily transactions recorded in our system by FTP. Since the bank sends a daily report of the transactions to the department financial system every day, financial information can be reconciled and corrected, if needed.

3. DATA ACCESS SERVICES

Implementing the online purchasing system has enabled CTIS to automate data sales and distribution. This system has made it possible to respond to the needs of traditional users of Geographical Information Systems (GIS) and geospatial data. In order to broaden our customer base and better serve users, CTIS is now turning to emerging technologies that support standardized interoperable solutions that allow access to Web-based, wireless access to geospatial data.

3.1 Data Access for Visualization

As a first initiative to make geomatics accessible to a broader segment of users, CTIS launched Toporama as a new product in fall 1999. Toporama (http://toporama.cits.rncan.gc.ca) is a Web site that provides access to free topographic images of Canada. It is a raster representation of part of the contents of CTIS's NTDB and ortho-rectified Landsat 7 products. In addition, Toporama offers simple navigation and localization tools to help and guide users through the images.

In light of recent developments for accessing geospatial data, CTIS initiated the establishment of Web Mapping applications to replace the first version of Toporama. In fact, Web Mapping Technology is changing the rules for spatial technology applications and opening new opportunities. Web Mapping is an innovative Web-based system that allows users to define map compositions and deploy them on the Internet. More people are using the Web to get geospatial information. As the Web Map Server base grows, users have an easier time getting and making maps via the Web. Furthermore, it enables providers of spatial information to more readily set up servers that provide maps to users. In addition, the Open GIS Consortium (OGC) has played a role in developing open interface specifications. This has impacted on the interoperability that enables network communication among server applications and access for displaying, merging, and converting spatial information from multiple sources.

As a participant in the Canadian Geospatial Data Infrastructure (CGDI) Development Network, CTIS recently implemented a service that complies with Open GIS Consortium (OGC) Web Map Server implementation specifications. The system includes three components: the Spatial Data Warehouse, Web Map Server, and Viewer application.

The CTIS **Spatial Data Warehouse** contains data sets in vector (NTDB, URN, Boundaries), raster (Landsat 7 images), and grid (CDED) formats that correspond to our products. The Data Warehouse is a database that integrates data from more than one data source. The operational data included in the database is organized to support easy and efficient access and to allow users to discover new relationships and data patterns by navigating and mining the large volume of data. In our case, the Spatial Data Warehouse differs from our source database to allow better viewing and faster access performance.

The CTIS **Web Map Server** component provides access to the Data Warehouse. It is a Web-based, server-side application that complies with the Web Mapping Server (WMS) 1.1.1 OGC Implementation Specification, including the map cascading capability that allows customers to access multiple, distributed geospatial information data stores across the World Wide Web. The Map Server responds to WMS operation via HTTP from any compliant WMS. For example, the Map Server returns a GIF or JPEG map image for a WMS getMap operation request. The Map Server recognizes the getCapabilities, getMap, and getFeatureInfo WMS requests.

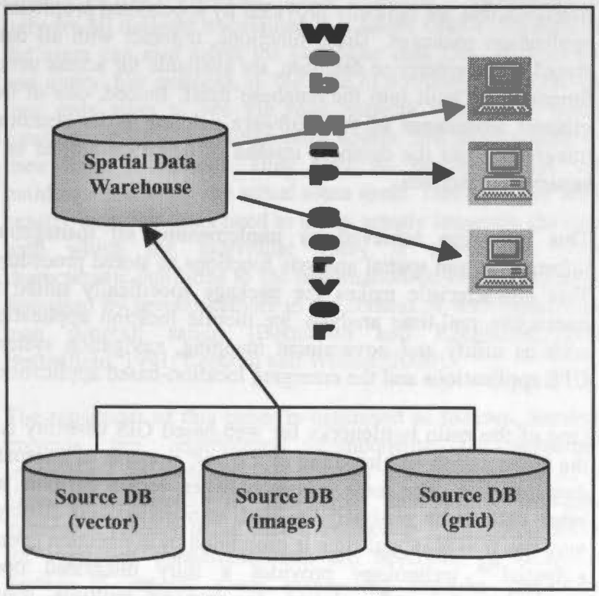


Figure 4. CTIS WMS Architecture

The **Viewer** component is a client-side application that allows users to spatially browse data warehouses served from any Web Map Server that is compliant with the Web Map Server OGC implementation specification. The Viewer can be accessed over the Web using a Web browser. The application is divided into several sections. The View section includes the image of the current map, showing the selected layers, navigation tools, and the legend. The Layer section contains the tools that allow users to interactively create their own maps by selecting data stores, themes, and layers that are available.

Web Mapping technology will help us improve access to our geospatial information. Software developers will be able to exploit the services that CTIS will provide for applications related to geoprocessing, decision support and analysis. Users will also benefit from having ready access to up-to-date geospatial data without the worry of storing and updating them. In the coming year, CTIS will provide public access to our data and applications. We will then extend services that take advantage of the extraction and transaction operations supported by Geographic Mark-up Language (GML) and Web Feature Server (WFS) OGC standards.

4. CONCLUSIONS

The new CTIS Web site is operational since May 2001. The new Web site complies with the Government On-line initiative, uses the common "Canada look and feel," and demonstrates the federal government's commitment to offering its products and services on the information highway. Moreover, it highlights the shift to Web-based commerce for every transaction component, from ordering to payment and delivery.

The online purchasing section is one of the main additions to the new Web site. It enables the Centre to sell topographic data to a wide-ranging public. The system is fully automated, from product search to data delivery, including secure credit-card transactions. Similar to the online purchasing system, the subscriber system automatically delivers topographic products ordered by users who have previously purchased a CTIS Subscription license for Digital Topographic Information.

In the years to come, CTIS will improve its spatial data distribution system by providing new search tools, new products, and new data and metadata formats. Interoperability is becoming a requirement to enable access to nationwide data services. As a national geospatial data custodian, CTIS will contribute to the Canadian Geospatial Data Infrastructure development network by implementing Web Map Service and eventually services to access the data that are compliant with international standard.

5. REFERENCES

CubeWerx, 2001. *CubeStor for Oracle*, User Manual, version 2.10. Canada, September 20, 2001.

CGDI Architecture Working Group, 2001. *Architecture Description, Canadian Geospatial Data Infrastructure*, January 19, 2001.

McKee, L., 2001. *Web Mapping Guide – Technology Trends*, Open GIS Consortium Inc. http://www.geoplace.com/gr/webmapping/technology.asp (accessed 1 March 2002).

Open GIS Consortium Inc., 2001. *Web Map Service Implementation Specification*, Open GIS Consortium Inc., Version 1.1.0, 2001-06-21.

Simoneau, B., 2001. *Online Distribution of Digital Topographic Data*. Digital Earth 2001 Conference, Fredericton, New Brunswick, June 2001.

GEOSPATIAL DATA HANDLING FOR WEB-BASED AND MOBILE APPLICATIONS

M. Bertolotto [a], *, J. D. Carswell [b], L. McGeown [c], P. Thijs [c]

[a] Dept. of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland - michela.bertolotto@ucd.ie
[b] Digital Media Centre, Dublin Institute of Technology, Aungier St., Dublin, Ireland - jcarswell@dit.ie
[c] e-Spatial Solution, Dublin, Ireland - info@e-spatialsolutions.com

Commission IV, WG IV/2

KEY WORDS: Spatial, Database, Web based, GIS, Real-time, Query.

ABSTRACT:

One of the main bottlenecks for web-based GIS usability is in the transmission and handling of the vast amounts of geospatial data involved. The shear volume of raster, vector, attribute, and other data to be analysed, queried, transmitted and displayed over the WWW in real-time is prohibited by connection speeds. In this paper we introduce e-Spatial^{TM} technology, a fully integrated open-standard web-based solution to manage multiple spatial relationships directly in an Oracle Spatial database and to deploy in real-time spatially enabled (or standard) applications completely independent of data volume. Indeed, no application code is ever installed on the client side. With the database generating all requested information on the fly, only the minimum data required is downloaded to the client, which leads to the most economical usage of bandwidth and resources for real-time response. This technology dramatically extends the functionality of Oracle Spatial by allowing real-time spatial data display, collection, editing, manipulation, and query using a standard web browser. Therefore, applications can be deployed on any device that supports the Java Virtual Machine, thus providing full support for wireless, PDAs and other mobile devices.

1. INTRODUCTION

Many GIS applications (such as those developed by ESRI, MapInfo, Autodesk, Intergraph) require the use of proprietary application-specific software (East et al., 2001). Furthermore, they store spatial and non-spatial data attributes separately: typically spatial properties of objects are stored in files that are managed by a file management system while attribute data is stored in a commercial database (e.g., a relational database). This split design presents the difficulty of maintaining data integrity between spatial and attribute data as the two types of data are not managed by the same engine. Oracle Spatial provides the information infrastructure that includes a single database system for managing both types of data, together with a data structure that is independent of the particular application.

From a data management point of view, using such a unified approach has an important advantage: it allows users to access full function spatial information systems based on industry standards, i.e. with an open interface to all data (e.g., SQL). Furthermore, the capability of storing spatial data in existing enterprise-wide database management systems enables the spatial deployment of several more enterprise applications that may have otherwise remained unexploited. Analysis and reporting functionality is also facilitated by the complete integration of management information and spatial data repositories.

e-Spatial^{TM} technology (www.e-Spatial.com) has been developed within the Oracle Spatial unified environment. Its web-enabling application architecture component iSMART^{TM} and its Java plug-in iSIS dramatically extend the functionality of an Oracle Spatial database by allowing real-time spatial data display, collection, editing, manipulation and query using a standard web browser (Bertolotto et al., 2001). By utilising e-Spatial^{TM} technology, Oracle Spatial users can exploit a full range of advanced spatial data handling and management functions that are typically provided by specialised proprietary application packages. These functions, together with all data stored in the enterprise database, are available for access using functionality built into the database itself. Indeed, one of the greatest advantages of this software solution is its seamless integration into the database instead of being developed as a separate application.

This has been achieved by implementing all management information and spatial analysis functions as stored procedures. This characteristic makes the package specifically suited for interactive real-time analysis for mobile location applications such as utility and government mapping, navigation systems, GPS applications and the emerging location-based applications.

One of the main bottlenecks for web-based GIS usability is in the transmission and handling of the vast amounts of geospatial data involved. The shear volume of raster, vector, attribute, and other data to be analysed, queried, transmitted and displayed over the WWW in real-time is prohibited by connection speeds. e-Spatial^{TM} technology provides a fully integrated open-standard web-based solution to manage multiple spatial relationships directly in the database and deploy in real-time spatially enabled (or standard) applications (in a two or three tier architecture) completely independent of data volume. Indeed, no application code is ever installed on the client side resulting in a thin client applet. With the database generating all requested application information on the fly, only the minimum data required is ever downloaded to the client, which leads to the most economical usage of bandwidth and resources for real-time response. Therefore, applications can be deployed using standard browser technology on any device that supports the Java Virtual Machine, thus providing full support for wireless, PDAs and other mobile devices.

---

* Corresponding author.

While in the past spatial information was utilised within specific applications and exclusively by expert high-end users, recently, thanks to the diffusion of desktop GIS and the Internet, its integration within the widest range of information systems is becoming a common requirement. Within this context, a pressing need relates to providing non-expert users with easy-to-use environments where they can visualise, query and manipulate spatial data without requiring any specific programming or database management background. To this purpose, additional tools to build and customise graphic user interfaces that facilitate interaction with the iSMART™+iSIS platform have also been developed.

In this paper we focus on the geospatial data handling characteristics of i-Spatial™ Information Server (iSIS) developed by e-Spatial™ Solutions. The data model utilised is the Oracle Spatial object-relational model. Furthermore, we discuss the problem of guaranteeing the preservation of topological consistency as it represents a critical issue from the point of view of data management (Egenhofer and Franzosa, 1991). We describe how the integrated Oracle 9i topology management functionality completely controls topology within the iSMART™+iSIS platform: if a spatial element is updated, all spatially related elements are automatically changed accordingly to guarantee consistency.

Finally, existing systems have limited ability to switch on/off specific subsets of spatial data. They are restricted to using all the data contained within each data layer. Therefore the data content is fixed. However, using e-Spatial™ technology not only are all datasets maintained in a single Oracle database, but the users can also select the particular data they wish to view and query. For example, in the case of a routing query the user may only wish to see the road network from their starting location, and the buildings only within 500 metres radius of their final destination. They have no requirement to view buildings data along the actual route itself. This ability to define exactly what the users need to query greatly improves the speed with which the individual query is completed. This characteristic conforms to the information-on-demand approach discussed by several authors in the context of web-based vector map generalisation (Bertolotto and Egenhofer, 2001; Buttenfield, 1991; Cecconi and Weibel, 2000).

The remainder of this paper is organised as follows. Section 2 presents the architectural components of e-Spatial™ technology. In section 3 we discuss the management of the topological structure of geo-spatial data within web-based systems developed using such a technology. Additional utilities for improved user-friendliness are described in section 4. Section 5 reports some concluding remarks and future developments.

2. e-SPATIAL™ ARCHITECTURE

In this section we briefly describe the two major components of e-Spatial™ technology, namely the iSMART™ database development technology and the iSpatial™ Information Server (iSIS), together with their architectural characteristics. More details on the e-Spatial™ architecture can be found in (Bertolotto et al., 2001).

iSMART™ is an intelligent Oracle application database development technology that allows database administrators and software engineers to design, build and deploy spatially enabled (as well as standard) Internet applications without requiring any proprietary or application specific source code. It therefore allows to reduce overall system development and deployment costs. All iSMART™ applications are automatically generated on the fly directly from the Oracle database and can be used and modified on-line without having to log-off existing Internet applications or writing any client application code.

The iSpatial™ Information Server is a Java plug-in component of iSMART™. It consists of a database management application program that enhances the spatial functionality of an Oracle Spatial database.

2.1 iSMART™ Architecture

The iSMART™ development environment relies on a three-tier architecture comprising three main layers, namely the Client Layer, the Application Server Layer, and the Database Layer (see Figure 1).

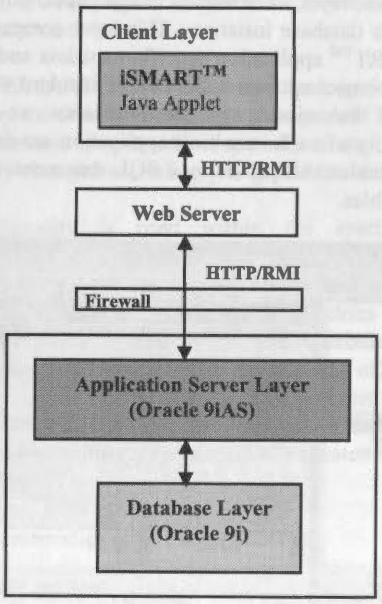


Figure 1. The iSMART™ architecture components.

All communications between the client layer and the database are conducted through the application server layer. The application is executed on the client using an applet that runs in a standard web browser. The applet communicates with the application server using the existing HTTP or RMI networking protocols.

The client layer is a light-weight client machine running a Java Virtual Machine (JVM). The iSMART™ Java applet is a micro thin applet that executes all commands received from a Panel EJB.

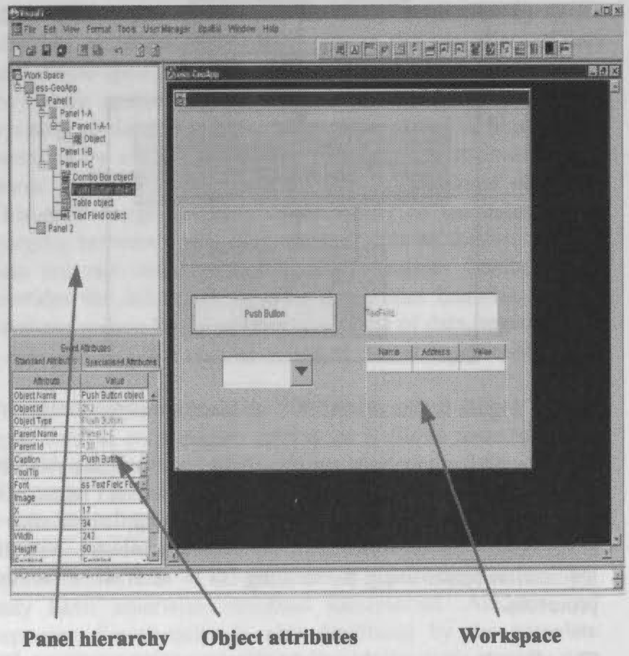
The application server currently used by iSMART™ is Oracle 9iAS. This layer contains several Enterprise Java Beans (EJBs) described in the following. Session EJBs are responsible for all communication with the client. As it is a stateful session EJB, an object of this type is instantiated for each user session. This

EJB validates all data submitted by the iSMART™ Applet before passing it to the entity beans.

Panel EJBs are entity beans that query the iSMART™ Application Specific Metadata for all information about the relevant panels (or forms to be displayed on the client applet) from the database and send information back to the iSMART™ EJB which returns it to the client. Each Panel EJB has a hierarchical structure (i.e., it can contain nested sub-panels and objects). Each panel in the user application corresponds to a Panel EJB in the hierarchy. In Figure 2, an example of panel hierarchy corresponding to a user application is shown as it appears in the Visuali development environment (see Section 4). On the right hand side of the screen the user builds panels in a workspace; on the left hand side the panel hierarchy is displayed.

SQL statements are needed to retrieve from the database the information to be displayed on the client site. Pre-defined SQL commands are stored in the iSMART™ Metadata tables of the database. The entity beans responsible for the execution of SQL statements are called SQL EJBs.

The database layer, as described in this paper, can refer to single or multiple database instances. This layer comprises two parts: the iSMART™ application specific metadata and the user data. The first component is a collection of standard database tables. The core behaviour and characteristics as well as the functionality of each user-built application are defined in these metadata tables. All pre-defined SQL statements are also stored in these tables.


File
Edit
View
Format
Tools
User Manager
Spatial
Window
Help
Work Space
ess-GeoApp
Panel 1
Panel 1-A
Panel 1-A-1
Object
Panel 1-B
Panel 1-C
Combo Box object
Table object
Text Field object
Panel 2
Event Attributes
Standard Attributes
Specialised Attributes
Push Button
TextField
Name
Address
Value
Panel hierarchy
Object attributes
Workspace

Attribute	Value
Object Name	Push Button object
Object Id	[illegible]
Object Type	[illegible]
Parent Name	[illegible]
Parent Id	[illegible]
Caption	Push Button
ToolTip	
Font	ss Text Field Font
Image	
X	17
Y	34
Width	242
Height	50

Figure 2. Screenshot of a user application under construction.

2.2 iSIS architecture

Similar to the iSMART™ architecture, the iSIS architecture presents three main components: the Client Layer, the Application Server Layer, and the Database Layer (see Figure 3).

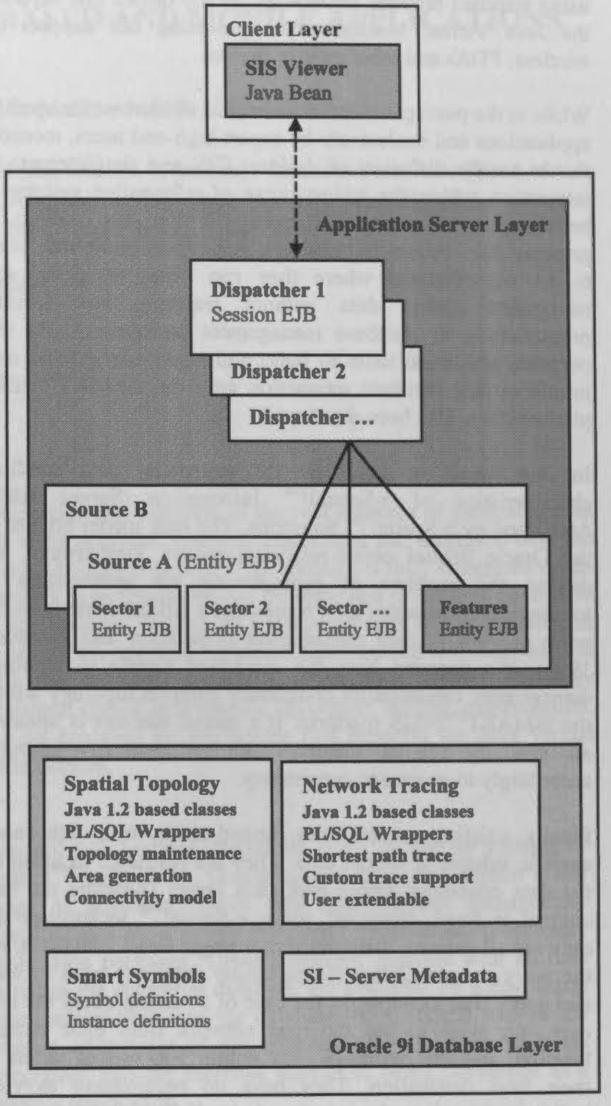

Client Layer
SIS Viewer
Java Bean
Application Server Layer
Dispatcher 1
Session EJB
Dispatcher 2
Dispatcher ...
Source B
Source A (Entity EJB)
Sector 1
Entity EJB
Sector 2
Entity EJB
Sector ...
Entity EJB
Features
Entity EJB
Spatial Topology
Java 1.2 based classes
PL/SQL Wrappers
Topology maintenance
Area generation
Connectivity model
Network Tracing
Java 1.2 based classes
PL/SQL Wrappers
Shortest path trace
Custom trace support
User extendable
Smart Symbols
Symbol definitions
Instance definitions
SI – Server Metadata
Oracle 9i Database Layer

Figure 3. iSIS architecture components.

The iSIS client layer bean is a plug-in that controls the display and manipulation of the vector and raster data within the web browser. The bean requests the specified data from the Dispatcher EJB (on the application server layer described later), which in turn renders all appropriate data for client display using the bean.

The application server layer contains several EJBs. The Dispatcher EJB is a stateless session bean that handles communication with the client and coordinates the display of the different Sector EJBs. These are entity beans that handle the retrieval of the data from the database for a given area (called a sector). Source EJBs are entity beans that control the access to the SI server metadata in the database. This information is used to determine whether the subset of data considered (called source) is a vector or a raster dataset and where the source information can be found. A Feature EJB is an entity bean that controls access to the feature tables (storing the feature symbology) in the database. The feature information is used by the Sectors to display object information.

The iSIS database layer consists of a collection of Oracle database tables and stored procedures written in Java that run inside the Oracle Aurora Virtual Machine. This way, all functions and procedures have fast and efficient access to the information stored inside the database without adding to the network overhead. These functions can be accessed via PL/SQL wrapper code.

The database has four distinct features:

1. *Spatial Topology*: the spatial topology package builds and maintains a topology model in the database for the spatial objects. This topology is built and maintained from the geometry objects in a specified source table (more details on the data model and management of the topological structure are provided in Section 3).

2. *Smart symbols*: they are representations of map symbols that can have an internal connectivity and a number of external ports. The smart symbol definition models the possible states for the internal connections between the ports of the object. The smart symbol instance models the current state of all the internal connections of a given smart symbol. This state information can then be used by the trace modules to conduct a network trace inside the database.

3. *Network Tracing*: the network trace classes are a series of predefined classes that can be extended by the user to perform any type of trace. These classes work with the topology information and the smart symbols to determine connectivity and connection information. The base trace classes use topological connectivity to determine if objects are traceable and connected. Users can easily extend this model to determine if components are traceable by accessing other criteria such as attribute data from secondary tables or other criteria. The definition for the trace algorithms are also stored in a set of metadata tables that can be modified and extended with the use of SQL. This allows the users to create their own trace definition on the fly.

4. *Spatial Information Server Metadata:* since the spatial information server is a generic tool to display spatial information from any table that contains Oracle spatial data columns or Oracle Intermedia images, it requires some information on where it can find these objects and images and how it should access them. This information is stored in the SIS metadata tables.

3. TOPOLOGICAL DATA MODEL

In this paper we describe the spatial data model on which e-Spatial^{TM} technology is based together with the mechanism for guaranteeing the preservation of topological consistency, an essential property for data usability.

The spatial data model used by e-Spatial^{TM} technology is the Oracle Spatial object-relational model. Object-relational models extend classical relational models by introducing object oriented concepts. The advantage of using object-relational systems (instead of object-oriented systems) is that they provide object orientation features while remaining compatible with previous relational applications. Therefore they minimise the need for complex migration of existing data, re-training of personnel and re-writing of application programs.

The basic object features of the Oracle object-relational model are type definition services that extend significantly the domain definition services of relational systems. Oracle Spatial offers the option to use both the relational and the object-relational models. Such a model consists of a schema, a spatial indexing mechanism and a set of operators and functions that facilitate the storage and retrieval of spatial data in an Oracle Database. Spatial queries such as window and range queries as well as spatial joins can be performed using the predefined spatial functions and operators provided.

The spatial attribute of an object (i.e., its geometric representation in a coordinate system) is stored in a single column. A single record (row) is used per geometry instance. Oracle Spatial allows to store vector data in the form of points, lines and polygons together with their topological relationships such as disjoint, overlap, touch, contains, etc. (Egenhofer and Franzosa, 1991). The topological consistency checks (e.g., check if a given object is self intersecting) and topological queries (e.g., are two objects disjoint) are based on pre-defined spatial operators and functions.

Within the iSMART^{TM}+iSIS platform topology is completely controlled by the Oracle 9i topology management functionality: if a spatial element is updated, all spatially related elements are automatically changed accordingly to guarantee consistency.

Such functionality is used within the spatial topology component of the iSIS database layer (described in Section 2): the iSIS database *topology engine* builds and maintains the topological relationships of all the geometries (node, arc, boundary, area, region, etc.) within the database itself. The topology engine supports the automatic building of the topology inside the Oracle Spatial database as data (geometries) are bulk loaded into the database (i.e., as vector data is bulk loaded the topological relationships are automatically created inside Oracle geometry tables).

Specific characteristics of this component are:

- Bulk topology creation from existing tables containing Oracle Spatial objects. This function allows topology to be built (rebuilt) from existing data.

- Topology maintenance functions handle changes in the topology when elements in the base table are created, modified or deleted. To ensure data consistency this is usually handled via database triggers, but other custom approaches are possible, depending on the application's needs. The topology system can be configured to combine multiple sources or parts of different sources into a topology layer. Examples include building topologies from certain features from a combination of source tables.

- Multiple topology layers. The system is not limited to just one topology. The system supports an unlimited number of topologies.

- Object connectivity model. The topology can be used as a connectivity model for linear topologies. When other applications change the source tables, the connectivity is maintained automatically since it is dependant on the topology, which in turn is automatically maintained.

- Client independent. Regardless what desktop software is being used to edit the spatial data, topology and connectivity are maintained automatically inside the database.
- Dynamic generation of areas with support for holes from the boundaries in the topology. Will act differently depending on the signal you trace.

A topology wizard within iSIS allows users to assemble, build and maintain topology. iSMART^{TM} also provides a set of generic tools to extend and customise the topology wizards to each organisations specific requirements.

4. DEVELOPMENT TOOLS

Additional tools are provided for facilitating the building of applications that exploit spatially enabled datasets. These tools are particularly useful for non-expert programmers as they allow the creation of features together with their symbology, the importing of digital imagery and incorporate the drag-and-drop methodology to graphical user interface design.

The *Pyramid Builder* is a utility designed to load scanned aerial/satellite photos and topographic maps into the database. The loaded imagery is then available for heads up digitising. This image loading utility follows the long standing approach to digital image handling within existing dedicated image processing systems by pre-processing the image data into multi-resolution representations (Carswell and VdLaan, 1992), hence the name "Pyramid Builder". The idea being to eliminate the need to resample the image at run time by loading the best fit resolution image to the zoom factor currently active within the web browser.

e-Spatial^{TM} technology advances this trend by further subdividing the individual reduced resolution image files into tiles of a pre-defined pixel size. Thus, when a web browser is viewing a particular geographic location, only those individual tiles at the requested zoom factor will be physically sent over the network, and not the entire image file. This of course is essential to real-time viewing of digital imagery due to their inherent size restrictions: a typical black&white aerial photo scanned at 30um (850dpi) resolution requires approximately 60mb of disk space (uncompressed) (Kern and Carswell, 1994), while colour imagery at the same scanning resolution will require three times as much space, i.e. 60mb for each of the three colour bands; red, green, and blue.

The *Feature builder* is a utility that allows users to define new data sources and the feature classes that the source controls. All feature characteristics (e.g., colour, weight, pattern, etc.) are defined using this utility. Also incorporated within the Feature Builder is the capability to add events to specific feature digitising operations. For example, in a land parcel application, when a "Property Centroid" feature first gets digitised, an event can be associated with this operation that calls a pre-defined SQL query to populate the "Property ID" column in the user attribute "Property" table and subsequently display this attribute value as the centroid text in the graphics. Alternatively, if a Property Centroid gets deleted from the graphics, an SQL query previously attached as an event to this type of an operation (on this type of feature) can delete all the user attributes associated with this centroid feature.

The iSMART^{TM} application builder, called Visual*i*, has been designed to allow any user to build a customised graphic user interface for both standard MIS and spatial applications without writing any source code. This application development incorporates drag-and-drop functionality to add objects, such as buttons, tables, combo boxes, etc. to panels and sub-panels in a hierarchical fashion (see Figure 3 for a screen shot example of a user application under construction).

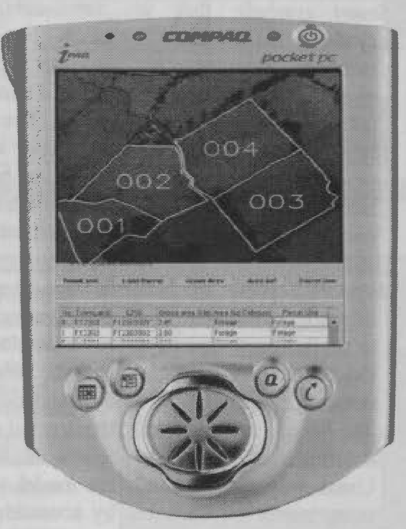


Figure 4. A LIMS application developed using e-Spatial^{TM} technology and viewed on a PDA.

Visual*i* allows the building of "business rules" that are attached as events to the buttons and other GUI objects contained within the application panels. For example, a business rule can be created and attached as a "mouse pressed" event on a "Property Value" button object that, when pressed, subsequently shades all the property polygons according to their "Property Value" attribute.

SQL commands are also created within the Visual*i* application builder through the use of the SQL Command Editor. Using drag-and-drop operations, SQL queries are built where table columns can be dynamically attached to their respective "textfield" objects on the application panels. Thus both retrieving data from the database and inserting/updating data is easily accomplished.

5. CONCLUSIONS

In this paper we presented the advantages of the geospatial data handling approach of e-Spatial^{TM} technology that allows for effective development of web-based and mobile spatial applications within an Oracle database environment. Such a technology utilises the Oracle Spatial object-relational model for on-line storage, retrieval and query of spatial data. This data model presents the great advantage of automatically preserving topological consistency, an essential property for data exchange and usability between remote sites (such as in web-based and mobile applications). Therefore, if a spatial element is updated (for example, by a remote user), all spatially related elements are automatically changed accordingly to guarantee consistency.

From a data management point of view, the greater advantage of using Oracle Spatial is that it handles spatial and non-spatial data in a completely unified manner. Both types of data are stored in the same database: the geometric representation of an object (e.g., in the form of a point, line or polygon) shares the same table with the other attributes of the object. This contrasts typical spatial handling systems (such as commercial GIS) in which spatial properties of objects are stored in files that are managed by a file management system (such as shape files, design files, etc.) while attribute data is stored in a commercial database.

An important concern for the design and development of contemporary and next-generation information systems relates to interoperability, i.e., the capability of autonomous systems to exchange data and to handle processing requests by means of a common understanding of data and requests (Sondheim et al., 1999). Specifications on the conceptualisation of spatial entities have been provided by the OpenGIS Consortium. The Oracle Spatial object-relational model (upon which the e-Spatial™ data model relies) conforms to interoperability and standardisation requirements as it corresponds to the "SQL with Geometry Types" implementation of spatial feature tables described in the OpenGIS ODBC/SQL specification for geospatial features (OGC, 1996).

e-Spatial™ technology has been developed within the Oracle Spatial unified environment. Its web-enabling application architecture component iSMART™ and its Java plug-in iSIS represent a dramatic extension of the functionality inherent to an Oracle Spatial database. Indeed, it allows for real-time spatial data display, collection, editing, manipulation and query simply using a standard web browser. Advanced spatial data handling and management functions, typically provided by specialised proprietary application packages, can now be exploited by Oracle Spatial users without having to install any application-specific source code. This has been achieved by implementing all management information and spatial analysis functions as stored Java procedures. Therefore, applications can be deployed using standard browser technology on any device that supports the Java Virtual Machine, thus providing full support for wireless, PDAs and other mobile devices.

An example of web-based application developed using Spatial™ technology is the on-line Land Information Management System (LIMS) serving 125,000 farmers in the territory of the Republic of Ireland (developed for the Irish Department of Agriculture). The implementation and testing to make such a system available on hand-held mobile devices (Figure 4) are currently being finalised (Bertolotto et al., 2002).

Finally, existing systems have limited ability to switch on/off specific subsets of spatial data. Using e-Spatial™ technology users can select the particular data they wish to view and query. This ability to define exactly what the users need to query greatly improves the speed with which the individual query is completed, a critical issue for mobile applications. This characteristic conforms to the information-on-demand approach discussed by several authors in the context of web-based vector map generali sation (Bertolotto and Egenhofer, 2001; Buttenfield, 1991; Cecconi and Weibel, 2000). The potentialities of the application of e-Spatial™ technology in this regard are currently under consideration.

Furthermore, greater flexibility presents greater opportunities for personalisation: for example, users could build personal profiles and the system could automatically push personalised context-aware information – a topic of future work we are also investigating with this technology.

6. REFERENCES

Bertolotto, M., Carswell, J. D., McGeown, L., McMahon, J., 2002. Web-Based Spatial Information Management Systems. In: *Proceedings 95th Annual CIG Conference*, Ottawa, Canada.

Bertolotto, M., Carswell, J. D., McGeown, L., McMahon, J., 2001. iSMART™+i-Spatial™ Information Server: Deploying Integrated Web-Based Spatial Applications Within an Oracle Database Environment. In: *Proceedings of the International Workshop on Web Geographical Information Systems (WGIS2001)*, IEEE CS Press, Kyoto, Japan.

Bertolotto, M., Egenhofer, M., 2001. Progressive Transmission of Vector Map Data over the World Wide Web. *GeoInformatica*, 5(4), pp. 345-373.

Buttenfield, B., 1999. Progressive Transmission of Vector Data on the Internet: A Cartographic Solution. In: *Proceedings 18th International Cartographic Conference*, Ottawa, Canada.

Carswell, J., VdLaan, F., 1992. Tools for Large Scale Operational Raster Processing. In: *Proceedings EGIS92*, Munich Germany.

Cecconi, A., Weibel, R., 2000. Map Generalization for On-demand Web Mapping. In: *Proceedings GIScience2000*, Savannah, GA.

East, R., Goyal, R, Konovalov, A., Rosso, A. , Tait, M., Theodore, J., 2001. The Architecture of ArcIMS, a Distributed Internet Map Server. In: *Proceedings SSTD'01*, *Lecture Notes in Computer Science n. 2121*, Springer-Verlag, pp. 387-403.

Egenhofer, M., Franzosa, R., 1991. Point-set topological spatial relations. *International Journal of Geographic Information Systems*, 5(2), pp. 161-174.

Kern, P., Carswell, J. D., 1994. An Investigation into the Use of JPEG Image Compression for Digital Photogrammetry: Does the Compression of Images Affect Measurement Accuracy. In: *Proceedings EGIS94*, Paris, France.

OGC (OpenGIS Consortium), 1996. The OpenGIS abstract specification. http://www.opengis.org/public/abstract.html (accessed March 2002)

Sondheim, M., Gardels, K., Buehler, K., 1999. GIS Interoperability. In: *Geographical Information Systems: Principles, Techniques, Management and Applications*, P. Longley, M.F. Goodchild, D.J. Maguire and D.W.Rhind (eds.), Second Edition, Cambridge, GeoInformation International, pp. 347-358.

EXTENDING GEOSPATIAL REPOSITORIES WITH GEOSEMANTIC PROXIMITY FUNCTIONALITIES TO FACILITATE THE INTEROPERABILITY OF GEOSPATIAL DATA

Jean Brodeur[a, b], Yvan Bédard[a]

[a] Centre for Research in Geomatics, Laval University, Cité universitaire, Québec QC Canada G1K 7P4
[b] Natural Resources Canada, Centre for Topographic Information, 2144 King Street West, Sherbrooke QC J1J 2E8
Brodeur@RNCan.gc.ca, Yvan.Bedard@scg.ulaval.ca

Commission IV, WG IV/2

KEY WORDS: Geospatial data interoperability, Geospatial repository, Geosemantic proximity, Communication, Semantic similarity

ABSTRACT:

Today, with the common availability of Internet technologies, interoperability of geospatial data has become a necessity for sharing and integrating geospatial data. More specifically, it is seen as a solution to solve syntactic, structural, semantic, geometric and temporal heterogeneities between data sources. In Canada, we observe such heterogeneities from existing geospatial databases. For example, *Vegetation* ▣, *Trees* ▣, *Wooded area* ⧄, *Wooded area* ⊡⧄, *Milieu boisé* ▣ and *Zone boisée* (unknown geometry), found in different geospatial data specifications, describe the same type of phenomena. Recently, we have proposed a conceptual framework for geospatial data interoperability based on human communication concepts. This framework introduces the idea of *geosemantic proximity*, which provides reasoning capabilities to assess the semantic, geometric, and temporal similarities between geospatial concepts and geospatial conceptual representations. In the present paper, we review the conceptual framework and present an architecture of a system based on this framework. In fact, the architecture uses a geospatial repository, namely *Perceptory*, as a data source's ontology upon which we add *geosemantic proximity* functionalities. These functionalities evaluate the similarity of the information stored in the data source with the information required by another one in order to facilitate the interoperability of geospatial data.

1. INTRODUCTION

Many geospatial data sources are today publicly available to end users, especially by the way of Internet and geospatial data infrastructure (e.g. CGDI in Canada and NSDI in the United States). Typically, these data sources have been produced for specific purposes of people and organizations. For example, in Canada, there is the National Topographic Data Base (NTDB) (Natural Resources Canada, 1996) produced for national mapping and GIS application purposes, the VMap libraries (VMap, 1995) for military purposes, the Street Network Files and the Digital Cartographic Files (Statistics Canada, 1997) for enumeration purposes, and various topographic data sources produced at larger scales by provincial departments (e.g. (OBM, 1996; Québec, 2000)). These geospatial data sources abstract the topographic reality in various ways, which causes problems of sharing and integration when users join data from many sources. To illustrate this, we observe that forest-like phenomena are abstracted as *Vegetation* ▣ in NTDB, *Trees* ▣ in VMap, *Wooded area* ⊡⧄ in Ontario Digital Topographic Data Base, and *Milieu boisé* ▣ in the *Base de données topographiques du Québec* (where the pictograms mean polygonal, linear or point geometry; see (Bédard and Proulx, 2002) for the description of spatial pictograms). This rises up syntactic, structural, and, moreover, semantic, geometric, and temporal heterogeneities between the various data sources (Bishr, 1997; Charron, 1995; Laurini, 1998).

Interoperability of geospatial data has been introduced early in the 1990's as a solution for the sharing and the integration of geospatial data and geoprocessing resources (Kottman, 1999). The Open GIS Consortium Inc., the ISO/TC 211-Geographic Information/Geomatics, governmental organizations, the geographic information industry, and the research community have worked in co-operation to achieve the current foundation of geospatial data interoperability. Major progresses are noted especially for syntactic and structural heterogeneities (Rodriguez, 2000). As such, documents like (ISO/TC 211, 2001a; ISO/TC 211, 2001b; Open GIS Consortium Inc., 1999; Open GIS Consortium Inc., 2001) provide definitions of the content and the structure of geometric data as well as syntactical descriptions of geospatial data. However, interoperability of geospatial data must go beyond this fact to include solutions for semantic heterogeneity (Egenhofer, 1999). Following this line of thought, we have recently proposed a conceptual framework for geospatial data interoperability in order to position a new approach called *geosemantic proximity* for the assessment of semantic interoperability of geospatial data (Brodeur and Bédard, 2001). It consists in a human communication-like process between two agents, which exchange in order to share and integrate geospatial data from each other.

The elaboration of geospatial repositories is recognised as a good practice in the development of geospatial databases. A geospatial repository constitutes a comprehensive source of knowledge that captures the semantics and the structure of the data being stored in a geospatial database (Brodeur et al., 2000). Then, it can be considered as an application *ontology* (Gruber, 1993) to support geospatial data interoperability.

In this paper, we present an architecture of a system that integrates the idea of *geosemantic proximity* with geospatial repositories to assess the semantic interoperability of geospatial data.

The rest of the paper is subdivided as follows. In the next section, we review related notions to geospatial data interoperability. The third section summarizes our conceptual framework of geospatial data interoperability and the idea of *geosemantic proximity*. In section 4, we present a system architecture that implements the idea of *geosemantic proximity* upon *Perceptory*, a typical geospatial repository (Bédard and Proulx, 2002), based on our framework for geospatial data interoperability. Finally, we conclude in section 5 and present future work.

2. INTEROPERABILITY-RELATED NOTIONS

Our conceptual framework for geospatial data interoperability and the *geosemantic proximity* approach are based on a number of fields such as philosophy, human communication, cognition, computer science, and geographic information. More specifically, we have considered works related to ontology, context, semantic proximity, topology, semantic interoperability, and mapping specifications.

As people end up understanding each other when communicating, we think that interoperability of geospatial data obeys a *human communication process* (as described by (Schramm, 1971)). A human communication process consists in an individual transmitting details depicting real-world phenomena he/she has in mind to someone else. Typically, the communication process is made up of a human source, a human destination, physical signals, a communication channel, a source of noise, and a feedback component. Multiple representations of reality are involved in a communication process, namely the source and destination cognitive models and the physical signals that are used to transmit a message between the source and the destination.

The source and the destination cognitive models are built from the direct observation of phenomena and from the observation of intentional semantic signals from someone else. These signals are captured by our sensory systems to form *perceptual states* (Barsalou, 1999). The human selective attention gathers properties of interest and stores them as *perceptual symbols* (hereafter called *concepts*) (Barsalou, 1999) permanently in memory. A concept consists of both cognitive elements–i.e. hidden data-like elements–and a translation function that encapsulates these elements. This translation function recognizes and produces physical signals (hereafter called *conceptual representations*) about that concept (Barsalou, 1999). In memory, concepts are aggregated in clusters, which express a kind of similarity between each others (Krech and Crutchfield, 1971).

A conceptual representation acts as an intermediary between the source and the destination. It transmits a source's concept adapted to a specific context and a particular use towards destination.

To illustrate the communication process in the context of geospatial data interoperability, let's use the following example inspired from (Kottman, 1999). A person interesting to rent a house or the like asks an agent of a local service for information about *houses for rent* in the area of *Sherbrooke*. Once the agent receives the request, he/she interprets it, and provides a list of available *dwellings for rent* in the area of *Sherbrooke* to the person, which answers completely the person's request. In such a communication process, interoperability happens between the person and the agent.

The communication process as a model of interoperability for geospatial data includes multiple representations and descriptions of real-world phenomena. The representations and descriptions of real-world phenomena is still a subject matter studied in ontology and database modeling.

In its philosophical meaning, ontology is concerned with the description of the world in itself, with a model and an abstract theory of the world, and with the science of being, of the type of entities, of properties, of categories, and of relationships (Peuquet et al., 1998; Smith and Mark, 1999). However, in artificial intelligence, it corresponds to "an explicit specification of a conceptualisation" (Gruber, 1993) and "a logical theory accounting for the intended meaning of a formal vocabulary" (Guarino, 1998). As mentioned previously, one phenomenon could be described in multiple ways. Thus, following Gruber's and Guarino's definitions, we assume an ontology to be a formal representation of phenomena supported by a vocabulary and definitions that explicit the intended meaning, and represents phenomena with their interrelationships.

In database modeling, a conceptual model consists of an abstract description of a portion of reality from a data-centered point of view. A conceptual model is a tool to think about, to document, to communicate, and to develop data sources about parts of reality (Bédard, 1999). It captures, structures, and catalogues selected features in general categories, classes, properties, relationships, generalizations, aggregations, roles, constraints, behaviours, geometric properties, temporal properties, and so on, using a given formalism (e.g. UML). A good practice in developing conceptual model is to support it with a data dictionary that specifies the intended semantics of features. The conceptual model along with the data dictionary constitutes the repository's essential components (Brodeur et al., 2000). A repository is defined as a collection of metadata that is structured in such a way to provide the semantics and the structure of the objects stored in a database (Brodeur et al., 2000). It consists of information to assess fitness for use of data, to support data integration from multiple sources, and to support data interoperability. However, databases with their corresponding repositories (conceptual models and data dictionaries) are usually developed to serve specific needs and specific uses by database practitioner of various backgrounds and experiences. Consequently, the same part of reality is abstracted differently from one conceptual model to another. This drives to problems of interoperability when merging conceptual models and geospatial data to elaborate a more comprehensive set of data.

The abstraction of real-world phenomena is basically driven by the situation or the circumstances in which phenomena are perceived and used. This refers to the *context*, which influences the definition of concepts and conceptual representations with specific intrinsic and extrinsic properties (described later in this section). Context is recognized as a basic element for the assessment of semantic interoperability, which provides abstractions with their fundamental semantics (Kashyap and Sheth, 1996). A main requirement for semantic interoperability of geospatial data is to have reasoning methods that take the context into consideration. Semantic proximity in a context-based perspective is seen as such a reasoning method, which expresses qualitatively the semantic similarity between abstractions (e.g. semantic resemblance, semantic relevance, semantic relation, semantic equivalence and semantic incompatibility) (Kashyap and Sheth, 1996).

Conceptual representations are used to transmit concepts within a given context. However, to be interoperable, concepts and conceptual representations have to refer to the same phenomenon and, thus, the same identity of the phenomenon must be recognized from these various abstractions. Therefore, concepts and conceptual representations are not as important as the phenomena they represent. Identity is then considered to be a closely related notion to geospatial data interoperability. It consists in a meta-property from which we distinguish and individualize phenomena (Guarino and Welty, 2000), which allows the recognition of real-world phenomena representations.

Concepts and conceptual representations typically circumscribe a particular set of phenomena. We can imagine they follow a geometric-like metaphor, such as a segment on a semantic axis, with an interior and boundaries. The interior of the segment consists of the set of intrinsic properties and boundaries of the segment, of the set of extrinsic properties. Intrinsic properties are those providing the literal meaning. They describe the essential nature of a phenomenon. They are not dependent of any external factors. Basically, the identity of a phenomenon can be recognized from the intrinsic properties. Identification, attributes, attribute values, geometries, temporalities, and domain are good intrinsic property candidates for geospatial concepts and geospatial conceptual representations. Extrinsic properties are those influenced by external factors. They provide meaning based on their interaction with other concepts or conceptual representations and as such, set the limit of the literal meaning of a concept or a conceptual representation. Behaviours and relationships (semantic, spatial, and temporal) are good extrinsic property candidates for geospatial concepts and geospatial conceptual representations. Thus, extrinsic properties are associated to the notion of boundary. The notion of boundary has been discussed in (Casati et al., 1998); they have identified two types of boundaries: *bona fide* and *fiat* boundaries. *Bona fide* boundaries are associated to genuine or physical demarcation as it is the case for *buildings*, *runways*, and *the body of a person*. *Fiat* boundaries correspond to human driven demarcations, which are more theoretical, mathematical, or virtual and have no direct relationship with physical objects. This is the case for *administrative boundaries* or the boundary between waterbodies such as the *St.Lawrence Gulf* and the *Atlantic Ocean*. Concepts and conceptual representations are essentially defined by humans and, thus, can be associated with *fiat* boundary (Smith and Mark, 1999). According to (Casati et al., 1998), objects of *fiat* boundaries follow a topology that includes an interior and a boundary. This kind of topology has been a subject of interest in geospatial information (Egenhofer et al., 1994). We propose to extend the use of topology for the assessment of *geosemantic proximity*.

3. INTEROPERABILITY OF GEOSPATIAL DATA AND GEOSEMANTIC PROXIMITY

This section reviews our conceptual framework for spatial data interoperability and the idea of *geosemantic proximity*. They constitute the theoretical foundation of the architecture presented in the next section.

3.1 Interoperability of Geospatial Data

As introduced in the previous section, geospatial data interoperability follows a human communication-like process. To illustrate this, let's assume the situation in which a user agent (A_u) of geospatial data wishes to have information about the hydrologic network for flood analysis in the region of Sherbrooke. He/she sends a query to a geospatial data source, called a data provider agent (A_{dp}), to get information about *lakes* and *rivers* within *Sherbrooke*. When A_{dp} receives the A_u's query, it interprets it–i.e. to find and associate concepts it knows with the received conceptual representations (e.g. *watercourses* and *waterbodies* in the proximity of *Sherbrooke*). Once the query has been interpreted, A_{dp} gathers and sends to A_u the information that fulfils totally his/her original request (e.g. *Lac des Nations*, *Magog River*, and *Saint-François River*). In this process, A_u and A_{dp} use their own vocabulary to communicate their respective abstractions of real-world phenomena. Because of their common set of symbols and backgrounds, they can end up understanding each other (Bédard, 1986).

From this situation, we elaborated a conceptual framework for geospatial data interoperability that uses five expressions of reality: ***R***, ***R'***, ***R''***, ***R'''***, and ***R''''*** (Figure 1). These expressions are five separate ontologies that are linked together within a communication process to form what we call the *five ontological phases of geospatial data interoperability*. ***R*** represents the topographic reality as observed by A_u at a given time about which he needs information. ***R*** is beyond description. ***R'*** is the A_u's abstraction of ***R***. It corresponds to the set of properties of ***R*** selected by A_u and arranged into concepts. These properties constitute the A_u's cognitive model. ***R''*** is the set of conceptual representations encoded by A_u, which express relevant properties of ***R'***'s concepts to describe the A_u's specific need (e.g. *Lakes* or *Rivers* within *Sherbrooke*). These conceptual representations are the data used for interoperability that are placed in the communication channel towards destination A_{dp}. ***R'''*** refers to the set of concepts that A_{dp} has in memory. These concepts are used to decode ***R''***'s conceptual representations and to assign them an explicit meaning, for instance *watercourses* ⧄, *waterbodies* ▣ and *Sherbrooke* ▣. ***R''''*** gathers the conceptual representations encoded by ***R'''***'s concepts, which comply with ***R'***–i.e. the A_u's initial request–(e.g. *Lac des Nations* ▣, *Magog River* ⧄▣, and *Saint-François River* ▣). These conceptual representations are finally decoded and validated by A_u to identify if they infer the needed information. We can say that interoperability occurs only once ***R''''*** is validated. As we can observe from this conceptual framework, geospatial data interoperability consists in a bi-directional process with a feedback mechanism in both directions, which ensures that messages get destination and are satisfactorily understood by the recipient.

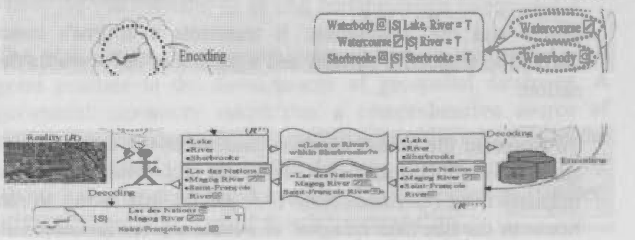


Figure 1: A Framework for Spatial Data Interoperability (Brodeur and Bédard, 2001)

Translation processes (encoding and decoding) are typically viewed as middleware components. However, in this conceptual framework, they are tied to concepts that are found in ***R'*** and ***R'''***. They are used to recognize and produce conceptual representations that fit the concept.

3.2 Geosemantic Proximity

As mentioned, a concept must be able to recognize and to generate conceptual representations. Therefore, concept's reasoning capabilities are considered a key element of the conceptual framework. As such, we suggest the idea of *geosemantic proximity* (*GsP*). It consists in a context-based approach that assesses the semantic similarity between a geospatial concept and a geospatial conceptual representation.

The context is basic in the abstraction process. Even if it is a fictitious and an imaginary notion, it is omnipresent and it guides the abstraction of phenomena. As such, the context makes some properties of phenomena more interesting than others do. Accordingly, the context (C) of a concept or a conceptual representation is described by the set of built-in properties, which are of two types of properties: intrinsic properties ($C°$) and extrinsic properties (∂C). Consequently, we define the context as in equation 1.

$$C_K = C_K° \cup \partial C_K \tag{1}$$

where:

C_K: Context of K

$C_K°$: Intrinsic properties of C_K

∂C_K: Extrinsic properties of C_K

GsP is a component of the concept's translation process. Analogously to human reasoning, it qualifies the likeness of a geospatial concept with a geospatial conceptual representation by comparing their respective intrinsic and extrinsic properties. As such, *GsP* consists of the intersection of the context of the geospatial concept K with the context of the geospatial conceptual representation L (equation 2). It is further expanded to a

$$GsP\,(\mathrm{K,L}) = C_K \cap C_L \tag{2}$$

where:

C_K: Context of the geospatial concept K

C_L: Context of geospatial conceptual representation L

GsP (K,L): *Geosemantic proximity* between K and L

four-intersection matrix (equation 3) to work out the different intersections between intrinsic and extrinsic properties. Each different intersections of the matrix can be tested empty (denoted by Φ or *f*) or non-empty (denoted by ¬Φ or *t*). The sixteen derived predicates (2^4) that are presented in row major form (i.e. row by row) according to the four-intersection matrix characterize the different *geosemantic proximity* cases: *GsP_ffff* (*disjoint*), *GsP_ffft*, *GsP_fftt* (*contains*), *GsP_tfft* (*equal*), *GsP_ftft* (*inside*), *GsP_tftt* (*covers*), *GsP_ttft* (*coveredBy*), *GsP_fttt* (*overlap*), *GsP_tttt*, *GsP_tfff* (*meet*), *GsP_tftf*, *GsP_tttf*, *GsP_ttff*, *GsP_fttf*, *GsP_fftf*, *GsP_ftff*.

$$GsP\,(\mathrm{K,L}) = \begin{bmatrix} \partial C_K \cap \partial C_L & \partial C_K \cap C_L° \\ C_K° \cap \partial C_L & C_K° \cap C_L° \end{bmatrix} \tag{3}$$

To illustrate the *GsP* approach, let's look about the following example. Consider that we have two data sources as different agents like in Figure 1. First, we have the National Topographic Data Base (NTDB) and second, the *Base de données topographiques du Québec* (BDTQ). For some reason (e.g. for update purposes), BDTQ requests to NTDB information about *street* ⧄. When NTDB receives the request, it searches in its content to find a concept that is semantically similar to *street* ⧄. It finds that the concept *road* ⧄ has an attribute called *classification*, which can take the value *street* of similar definition to *street* ⧄. They have as well the same type of geometry. As such, they have common intrinsic properties. Also, as part of the BDTQ's description of *street* ⧄, *street* ⧄ shows relationships with other classes of road that are part of the NTDB's description of *road* ⧄ and, as such, *street* ⧄ extrinsic properties are related to *road* ⧄ intrinsic properties. Consequently, the *geosemantic proximity* of the concept *road* ⧄ from NTDB with the conceptual representation *street* ⧄ requested by BDTQ is *GsP_ttff* or *contains.*

4. GEOSPATIAL REPOSITORIES AND *GsP* FUNCTIONALITIES

To validate the above conceptual framework of geospatial data interoperability and the *GsP* approach, a prototype is currently under development. It uses geospatial repositories that serve as agents' ontologies on top of which *GsP* functionalities are added. This section presents a system architecture, which describes how *GsP* functionalities are integrated to *geospatial repository*.

4.1 Geospatial Repository

A geospatial repository consists of a collection of metadata structured in such a way to provide the semantics and the structure of objects stored in a database. A geospatial repository can then be used as the ontology of the database that it describes.

In the architecture presented below, we are using *Perceptory*, which is a typical geospatial repository. *Perceptory* consists of a UML-based conceptual model building tool and an object database dictionary. It captures and manages representations of user's perceptions, which support the development of geospatial database. Accordingly, *Perceptory* allows definitions of *classes*, characteristics such as *descriptive attributes*, *geometries*, *temporalities*, and *visual information*, relationships including *associations*, *aggregations*, *compositions*, *dependencies*, and *generalizations*, *operations* (i.e. class behaviours), and more (Brodeur et al., 2000). It consists of a graph-like structure where relationships link classes together. However, the addition of *geosemantic proximity* functionalities would enhance *Perceptory* in order to serve for geospatial data interoperability.

4.2 An Architecture

The architecture, described below and illustrated in Figure 2, consists of three distinct components: two agents and a communication channel. They are describes below.

The two agents (A and B) are identical in this architecture. An agent can receive conceptual representations that are transmitted in the communication channel. Each of these conceptual representations is transformed in a data structure called conceptual representation, which takes place in the agent memory. This data structure is like a *perceptual state* of a cognitive agent. In order to be recognized, a conceptual representation is passed to a proxy. This proxy is a process that acts as an intermediary to locate a concept corresponding to the passed conceptual representation. A concept can be located either in the concepts data storage or in *Perceptory*. Concepts is an internal data storage where the most recent concepts used by the agent are placed temporarily. The object structure of a concept is described later in the section. This concepts storage is like the short-term memory of a cognitive agent. *Perceptory* is a direct access storage that includes a description of all classes, relationships, and so on that define the content of a geospatial database. *Perceptory* is like the long-term memory of a cognitive agent. The proxy looks first in the concepts storage to locate a concept that is similar to the conceptual representation. If a concept of this storage recognizes (i.e. is similar to) the conceptual representation, then it is used to answer the other agent. If not, then the proxy gets concepts from *Perceptory*. A graph traversal algorithm is used to navigate in *Perceptory*. It begins with a concept of the concepts storage that appears to be the most related one with the conceptual representation in order to access its associated concepts in *Perceptory*. The accessed concepts are returned in a concept (with no "s") object structure, which is identical to the structure of the concepts placed in the concepts data storage. Each concept obtained from *Perceptory* evaluates its similarity with the conceptual representation. This process is repeated recursively with concepts associated to the previous concepts until a concept that is *GsP_tffi* or *equal* is found or *Perceptory* is traversed completely. If an equal concept is located, then it is used to answer the other agent. Otherwise, concepts showing other kind of similarity with the conceptual representation are sorted by their *GsP* and the most similar one is used to answer the other agent. It might happen that no concept is found similar to the conceptual representation. Finally, the proxy returns the answer in term of encoded conceptual representations, which are sent towards destination the other agent in the communication channel.

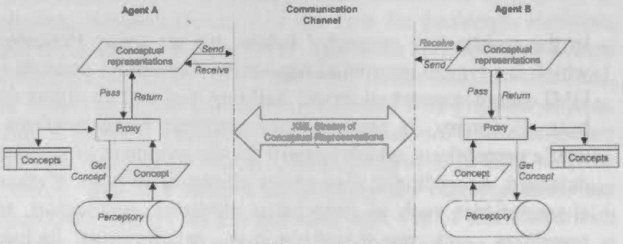


Figure 2: Architecture of the system

A concept is an object structure that consists of data elements that are hidden to other agents (Figure 3). These data elements are composed of a set of intrinsic properties and a set of extrinsic properties. On the one hand, the set of intrinsic properties is made of the definition of a *class*, its *characteristics* including *geometries* and *temporalities*, and its domain that are obtained from *Perceptory*. On the other hand, the set of extrinsic properties is made of the set of *operations* along with the membership of the class to *associations* and *generalizations* that are also obtained from *Perceptory*. As these data elements cannot be directly accessible by other agents, they are encapsulated by three functions: *recognize*, *generate*, and *gspRelate*. The *recognize* function evaluates if this concept can used to assign a meaning to the conceptual representation. The *gspRelate* function supports the *recognize* function by assessing the *geosemantic proximity* of the concept with the conceptual representation. The *generate* function produces conceptual representations that represents the concept within a specific context. Again, the *generate* function uses the *gspRelate* function to ensure that the generated conceptual representation is similar to the concept. Consequently, these three functions add reasoning functionalities on top of *Perceptory*, which provides the data. Thus, *Perceptory* and the *GsP* approach can be used together in order to assess automatically the interoperability of geospatial data.

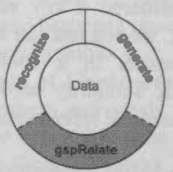


Figure 3: Object structure of a concept

Like a concept, a conceptual representation is formed of a set of intrinsic properties and a set of extrinsic properties, which encode selected intrinsic properties and extrinsic properties of a concept according to a specific context. Conceptual representations are transformed in an XML stream when placed in the communication channel and sent towards destination. An XML stream consists of a set of conceptual representation descriptions according to a predefined structure described by a DTD or an XML Schema. Instead of providing basic information of the concept such as the name of the concept, the attribute names and values, an XML conceptual representation provides also the definition of the concept, of the attributes, and of the attribute values. It provides also the domain, the geometry, and the temporality of the concept as well as the domain of attributes. Encoded likewise, a conceptual representation provides the necessary data to assess the semantic interoperability of geospatial data.

5. CONCLUSION

Supported by works on communication, cognition, ontology, geographic information, context, and semantic similarity, we have developed a conceptual framework for geospatial data interoperability. This framework corresponds to a human communication-like process that takes place between two agents. In this framework, we have differentiated two types of abstraction: *concept* (geospatial) and *conceptual representation* (geospatial). Concepts are stored in the agent memory and conceptual representations are used to communicate information about concepts. Concepts have reasoning capabilities, namely to recognize a conceptual representation and to generate a conceptual representation. Additionally, they are supported by the notion of *geosemantic proximity*, which is basically a context-based approach that qualifies the similarity of a concept with a conceptual representation. *Geosemantic proximity* corresponds to a four-intersection matrix between intrinsic and extrinsic properties of a concept and a conceptual representation. Finally, this conceptual framework serves to develop an architecture of a system that adds *geosemantic proximity* functionalities upon a geospatial repository, namely *Perceptory*. These functionalities extend geospatial repositories to facilitate the interoperability of geospatial data.

A prototype that is based on this architecture is currently under development. It aims at validating the conceptual framework and the *geosemantic proximity* approach. We expect that this research will conduct to important progress for semantic interoperability of geospatial data.

6. REFERENCES

Barsalou, L. W., 1999. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), pp. 577-609.

Bédard, Y., 1986. A Study of the Nature of Data Using a Communication-based Conceptual Framework of Land Information. Ph.D. Dissertation, University of Maine, Orono, Maine.

Bédard, Y., 1999. Visual Modelling of Spatial Database Towards Spatial PVL and UML. *Geomatica*, 53(2), pp. 169-186.

Bédard, Y., and Proulx, M.-J., 2002. Perceptory Web Site, http://sirs.scg.ulaval.ca/Perceptory

Bishr, Y., 1997. Semantics Aspects of Interoperable GIS. Ph.D. Dissertation, ITC Publication.

Brodeur, J., and Bédard, Y., 2001. Geosemantic Proximity, a Component of Spatial Data Interoperability. In: *Proceedings of International Workshop on "Semantics in Enterprise Integration" (OOPSLA 2001)*. Tampa, Florida, pp. 6.

Brodeur, J. et al., 2000. Modelling Geospatial Application Databases using UML-based Repositories Aligned with International Standards in Geomatics. In: *Proceedings of Eighth ACM Symposium on Advances in Geographic Information Systems (ACMGIS)*. Washington, D.C, ACM Press, pp. 39-46.

Casati, R. et al., 1998. Ontological Tools for Geographic Representation. In: *Formal Ontology in Information Systems*, Guarino, N. ed. IOS Press, Amsterdam pp. 77-85.

Charron, J., 1995. Développement d'un processus de sélection des meilleures sources de données cartographiques pour leur intégration à une base de données à référence spatiale. Mémoire de maîtrise, Université Laval.

Egenhofer, M. et al., 1994. The 9-Intersection: Formalism and Its Use for Natural-Language Spatial Predicates, Technical Report, 94-1, NCGIA.

Egenhofer, M. J., 1999. Introduction: Theory and Concepts. In: *Interoperating Geographic Information Systems*, Goodchild, M. et al. eds. Kluwer Academic Publisher pp. 1-4.

Gruber, T. R., 1993. A Translation Approach to Portable Ontology Specification, Technical Report, KSL 92-71, Knowledge Systems Laboratory.

Guarino, N., 1998. Formal Ontology and Information Systems. In: *Proceedings of Formal Ontology in Information Systems (FOIS '98)*. Trento, Italy, IOS Press, pp. 3-15.

Guarino, N., and Welty, C., 2000. A Formal Ontology of Properties (Preliminary Version). In: *Proceedings of ECAI-2000 Workshop on Applications of Ontologies and Problem-Solving Methods*, pp. 15.

ISO/TC 211, 2001a. *ISO/DIS 19107 Geographic Information - Spatial Schema.*

ISO/TC 211, 2001b. *ISO/DIS 19115 Geographic Information - Metadata.*

Kashyap, V., and Sheth, A., 1996. Semantic and Schematic Similarities Between Database Objects: a Context-Based Approach: a Context-Based Approach. *The VLDB Journal*, 5(pp. 276-304.

Kottman, C., 1999. The Open GIS Consortium and Progress Toward Interoperability in GIS. In: *Interoperating Geographic Information Systems*, Goodchild, M. et al. eds. Kluwer Academic Publisher pp. 39-54.

Krech, D., and Crutchfield, R. S., 1971. Perceiving the World. In: *The Process and Effects of Mass Communication*, Schramm, W., and D. F. Robert eds. University of Illinois Press pp. 235-264.

Laurini, R., 1998. Spatial Multi-Database Topological Continuity and Indexing: a Step Towards Seamless GIS Data Interoperability. *IJGIS*, 12(4), pp. 373-402.

Natural Resources Canada, 1996. *National Topographic Data Base - Standards and Specifications*. CTIS, Sherbrooke.

OBM, 1996. *Ontario Digital Topographic Database - 1:10,000, 1:20,000 - A Guide for User.*

Open GIS Consortium Inc., 1999. *OpenGIS Simple Features Specification for SQL*. OpenGIS Consortium Inc., Wayland, Mass.

Open GIS Consortium Inc., 2001. *Geography Markup Language (GML) 2.0*. Open GIS Consortium Inc., Wayland, Mass.

Peuquet, D. et al., 1998. The Ontology of Fields, Report of a specialist meeting held under the auspices of the Varenius Project, Bar Harbor, Maine.

Québec, 2000. Base de données topographiques du Québec (BDTQ) à l'échelle de 1/20 000 - Normes de production (V 1.0)

Rodriguez, M. A., 2000. Assessing Semantic Similarity Among Entity Classes. PhD. Thesis, University of Maine, Orono.

Schramm, W., 1971. How Communication Works. In: *Communication: Concepts and Processes*, DeVito, J. A. ed. Prentice-Hall Inc, New Jersey pp. 12-21.

Smith, B., and Mark, D., 1999. Ontology with Human Subjects Testing: An Empirical Investigation of Geographic Categories. *American Journal of Economics and Sociology*, 58(2), pp. 245-272.

Statistics Canada, 1997. *Digital Boundary File and Digital Cartographic File 1996 Census (Reference Guide)*. Minister of Industry, Ottawa.

VMap, 1995. Vector Map (VMap), Level 1, Military Specifications, Mil-V-89033.

Acknowledgements: The authors wish to acknowledge the contribution of Natural Resources Canada – Centre for Topographic Information in supporting the first author for this research and of the GEOIDE Network of Centres of Excellence in Geomatics (project DEC#2).

FACILITATING INTERDISCIPLINARY SCIENCES BY THE INTEGRATION OF A CLOSi-BASED DATABASE WITH BIO-METADATA

J. L. Campos dos Santos [1,3], R. A. de By [1], P. M. G. Apers [2] and C. Magalhães [3]

1. International Institute for Geo-Information Science and Earth Observation (ITC), Hengelosestraat 99, P.O. Box 6, 7500 AA, Enschede, The Netherlands – (santos, deby)@itc.nl
2. University of Twente, P. O. Box 217, 7500 AE, Enschede, The Netherlands - apears@cs.utwent.nl
3. The National Institute for Amazon Research (INPA), André Araújo Avenue 2936 – Petrópolis, 69.083-000 - Manaus – Amazonas – Brazil – (lcampos, celiomag)@inpa.gov.br

Commission IV, WG IV/2

KEY WORDS: Database schema, Metadata, Biodiversity Information Systems, Web Environment

ABSTRACT:

Biodiversity information has been collected and compiled during many unrelated and independent projects across Amazon region. Institutions on their own maybe unable to answer crucial questions, as their answers may depend on a multi-disciplinary context. Since their situation is still considerably isolated, some solutions adopted impose redundancy leading to high costs. The use of computer technology has been a fundamental resource for biological information management. However, such information is heterogeneous and to understand and automatically manage it accurate schema representation and formal metadata description are needed. This paper presents concepts of the CLOSi schema, which was conceived to ease and stimulate the development of biological collection databases. The schema represents functional groups that are specified in terms of object classes and their relationships. Another important issue presented, is the management of good quality bio-metadata that can be integrated with the data. For that, we implemented a solution to gather, manage and disseminate biological metadata, which is integrated with our CLOSi-based database. The FGDC metadata standard was adopted and represented as an XML schema. The schema was mapped to a well-formed XML template. A client/server infrastructure was tested to manage and disseminate data and metadata about Amazonian biodiversity.

1. INTRODUCTION

A vast amount of information has been compiled on the properties and functions of the Earth's living organisms, and an increasing proportion of this information is contained in large repositories in digital form. These include biodiversity data on the distribution of plants, animals and microbes; detailed genomic maps; compilations of the physiological functions of organisms, and information about the behaviour and function of species within ecosystems. Because these data have been collected and compiled during many unrelated, independent projects, their full research potential has not been reaped.

In the Amazon, even though the region is considered to have one of the highest concentrations of biodiversity, the data situation is critical. One of the most important assets are the biological collections, which are distributed across institutions. Biological collections are represented by samples, as well as by associated scientific information. The challenges that institutions are facing is to organize biological data in a Biodiversity Information System (BIS) with a consistent and efficient database implementation that allows the use of legacy data, and supports integration, intelligent and easy access for dissemination and sharing purposes.

To identify users and system requirements in this context, institutions were visited and problems identified in four domains: information production method, source of biodiversity data and information, degradation of information and inherent biodata modelling.

The "Instituto Nacional de Pesquisas da Amazônia" (INPA), is promoting computer-based solutions to help interdisciplinary collaborations and scientific advances. This aims to contribute to biodiversity preservation and sustainable development. The current interest of INPA is to share and disseminate data and information from its assets for a global scale audience. It started by implementing a strategy for information delivery, which includes: coordination, integration, analysis, presentation and delivery. To achieve that, understanding of the physical construction of its biological collections and scientific datasets is required. It also requires the knowledge about the data and objects information that will eventually be stored in the database. To identify users and system requirements, interviews and an evaluation were conducted.

The results of the study were grouped and clustered as schema objects, representing the items in collections. The schema provides a mechanism for database definitions that can be implemented in a (object) relational database management system.

Complementary to this, there was the need to describe and manage metadata and to integrate them with the data. We adopted the extended FGDC metadata standard, which incorporates the Biological Data Profile. Metadata has two purposes: (1) To help search and retrieval, or identify the location of data that meet a user's selection criteria, e.g., the "card catalog" type of descriptive information about a data set, and (2) To help a user to fully understand the data content and evaluate the usefulness of the data for his or her purposes, i.e., data set documentation.

This paper presents an implemented computer-based infrastructure for biodiversity data and metadata management. We focus on the support of CLOSi schemas for database implementations and the metadata standard for biological profiles. The standard has been implemented as XML schemas and transformed to an XML template. CLOSi databases can be integrated to Bio-Metadata through standard attributes and user-defined attributes. This method of integration allows users to access metadata information when querying collection objects, and to access data sets when browsing metadata descriptions.

2. SCHEMA REPRESENTATION FOR BIOLOGICAL COLLECTIONS

The design of database systems for managing biological collections requires the understanding of the physical structure of these biological collections. It also requires knowledge about the data and its characteristics. To acquire that, users were involved in the processes of data and systems requirements, since they play an important role, especially during data requirement analysis (Sonderegger *et al.* 1998; Campos dos Santos *et al.* 2000). At INPA, this phase was carried out in two ways: Interviews and descriptions evaluations.

Each participant in these processes was a specialist in some taxonomic group or a certain biological aspect of some taxonomic group. The interviews had an open format and researchers were asked the same general questions. This allowed us to elaborate a protocol to conduct the steps from field sampling to record information.

The data collected during a field mission consist of two parts: a general one, which holds the information that is normally collected in all studies (e.g., date, time, locality description), and a specific one, that corresponds to the scientific interest of a study (e.g., the altitude of a locality or the moon-phase may be of interest in one study but not in another). Interviewing scientists that worked on different studies and in different institutions helped to differentiate between information that is common to all and that which is used by just a few scientists. The result of the interviews was grouped by functionality and clustered as object types, representing the items in collections. The clusters were organized as: Collection Management, Collecting Event of Collection, Locality of Biodiversity Data, Taxonomy, Agent of Collection and Reference. This schema is called CLOSi, and clusters and their relationships are presented in Figure 1.

In the following, we present the main concepts of CLOSi; which include: clusters, object classes, relationships, specialization, attributes and its constraints, control value classes and notation.

CLUSTERS

A cluster groups a set of inter-related object classes. This forms the structure of a schema that can be used for the development of a database system. The schema combines a number of clusters and represents functional groups that are specified in terms of object classes and their relationships.

OBJECT CLASSES AND RELATIONSHIPS

An object class describes a population of objects in a collection and is identified by a unique name. One object class can have a relationship to other classes of any cluster. An object class is described by a class description and is associated with one or more attributes. A class can be a specialized class and for that, the class relationship is of type Is_a. An object class must belong to a unique cluster and can be associated with an optional list of control value classes (set of ordered atomic values).

IS_A OBJECT CLASSES

Specialization is an abstraction method that allows definition of object classes describing a subset of the population of another object class (super class). The inverse of specialization is generalization.

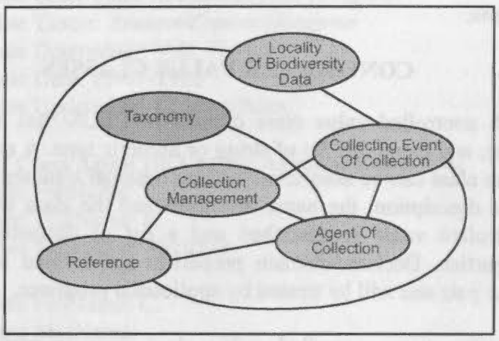


Figure 1 – CLOSi Clusters and their Relationships

ATTRIBUTES

An attribute is the basic unit of information about an object class occurrence. It may be local to the object class or inherited by relationship. CLOSi supports four types of attributes:

- **Standard** -- This attribute describes the common property of an object class, that is, attribute name, cardinality, the attribute data type and the attribute description. An object class can also be an attribute data type.
- **Composite** -- A composite attribute is formed by the concatenation of one or more attribute members and an attribute description.
- **Derivation** -- A derivation attribute represents a relationship between object classes and indicates that the attribute has an original object class (derivation_class_from), which derives to be an attribute of another object class (derivation_class_to).
- **Geo_attribute** -- It describes a geo-referential characteristic of the object-classes. It is described by an attribute name, cardinality and a Coordinates type or Distance. Coordinates of latitude and longitude are each defined by four values: degree, minutes, seconds and hemisphere. The Coordinate type can be of geographic, rectangular or nodes type. The

geographic types describe the latitude and longitude and are also described by four values: degree, minutes, seconds and the identification of the hemisphere (e.g., N, S, E or W). The rectangular type also describes the latitude and longitude, which are defined by two values: rectangular co-ordinates in meters and the hemisphere. The nodes type defines the nodes of a chain.

ATTRIBUTE CONSTRAINTS

An attribute associated with an object class has a cardinality constraint, which specifies the minimum and maximum number of values for attributes. In the absence of a specified cardinality constraint, it is by default `[0,1]` for single value attributes, and `[0,]` for set-of value or list-of valued attributes, where an unspecified maximum represents and unlimited number of values. Also, the CLOSi schema assumes referential integrity regarding attributes associated with controlled value classes: the value of such attributes must belong to their associated value classes.

CONTROLLED VALUE CLASSES

Each controlled value class defined in CLOSi has a unique name, and can either be of string or numeric type. A controlled value class can be associated with an optional standard value, a class description, the name of cluster and the class where the controlled value is described and a list of declared-domain properties. Declared-domain properties are defined as a tag-value pair and will be treated by application programs.

A string type controlled value class consists of a set of enumerated atomic values, which are all strings. If a standard value is declared, it must be one of the enumerated atomic values. Each string type of controlled value is associated with a `CODE_TYPE` associated with a data type for this code. Data type can be numeric type like: `INTEGER`, `SMALLINT`, `TINYINT`, `REAL`, `FLOAT`, `DECIMAL` and `NUMERIC`, or character string type like: `CHAR` and `VARCHAR`. A code can consist of alphanumeric characters and special symbols (. : + - * / \ ! =). The symbol "-" is allowed only when the code starts with A-Z or a-z.

Each controlled value also has a value description indicating for which cluster and class the controlled value class has been defined, as well as a standard value and a description of the controlled value class.

A numeric type control value class consists of a set of ranges, where each range is either a number or an interval defined by a lower and upper limit. For example[1], a numeric type controlled value class `INTERVALS` can be defined as a `1-10, 100, 200-300`. If a standard value is specified it must be within ranges defined for this controlled value class. Numeric type controlled values or range can have value descriptions but cannot be associated with a `CODE_TYPE`.

CLOSI NOTATION

A cluster is graphically represented by a double solid line rectangle with its name placed in the middle of the internal rectangle. There are two ways to represent classes: classes that belong to a current cluster (part of a current cluster) and those belonging to an external cluster. The first is graphically represented by double rectangle with dashed line in the external part. The latter classes are represented by a single rectangle with solid line.

Relations can exist between a cluster and its classes (cluster association), between classes to specify a derived relation, represented as solid arrow and between classes to define Is_a relation, represented as a dashed arrow.

Classes have attributes that are graphically represented as solid arrow. It is represented by attribute name and data type, including control value class and coordinates type.

The composite attribute is represented as a rectangle marked by two bullet point attachments to the attribute members. The attribute members are described as normal attributes.

2.1 CLUSTER TAXONOMY: AN ILLUSTRATION

To partially illustrate the CLOSi schema, we present in Figure 2, the cluster Taxonomy and its relationships. The cluster describes information about the taxonomic classification, identification and the ecological relations of the taxons. The object classes `Taxon_Name`, `Taxon_Relation`, `Classification` and `Determination` are the classes that describe the cluster. The object classes `Collection_Object`, `Agent` and `Reference_Work`, belong to external clusters and have relationships to classes within cluster Taxonomy.

The object class `Taxon_Name` describes the taxon at the lowest possible taxonomic level. The attributes of this class include: `Relation` (a list of ecological associations to other taxa), `TaxonAuthors` (the agents, persons, that made the original description of the taxon), `Repository` (the agent responsible for the specimens), `Synonym` (a list of names used for this taxon.), `Parent_Taxon` (the taxon rank that lies in the Linnean hierarchy immediately over this taxon), `TaxonOrigRef` (the reference that contains the original description of this taxon), `References` (the references in which the taxon is mentioned), `Classific` (the association of the taxon name with classification attributes). Additional attributes that describe this class include: `CommonNameandRegion` (a list of the common name of the taxon with its respective region), `Rank` (the Linnean rank of the taxon, e.g., Family, Order, Subfamily, etc), and `GeneralTaxonCoverage` (the geographic range of a taxon).

There are specific attribute relationships from classes within the same cluster to object class `Taxon_Name`. From `Determination`, there is the attribute `Taxon` (the taxon that represents the result of the determination); and from `Taxon_Relation`, there is the attribute `Relation_Host` (the taxon to which this relation exists).

Also, there are attribute relationships amongst object classes from the Taxonomy cluster to external object classes. The object class `Classification` defines a classification of the taxon in a defined structure and has four attribute relationships: `Author` (the agent - person, that made the original description of the classification); `References` (the references in which the classification is mentioned); `ClassificationOrigRef` (the reference that contains the original description of this classification); and `UpdatedBy` (a default object class attribute).

---

[1] This example was partially described in (Chen & Markowitz, 1996).

The object class `Determination` describes the information about the process of the determination of the taxon name. It has the relationship attribute `CollObject` (the identifier of exactly one collection object for which the determination has been made); `Determiner` (the list of agents that made de determination); and the `UpdatedBy` (a default object class attribute).

All CLOSi object classes have three default attributes: `Created` (date and time of the creation of the object), `Updated` (when the object was last updated) and `UpdatedBy` (identify the agent, in this case, the person that made the last update of the object).

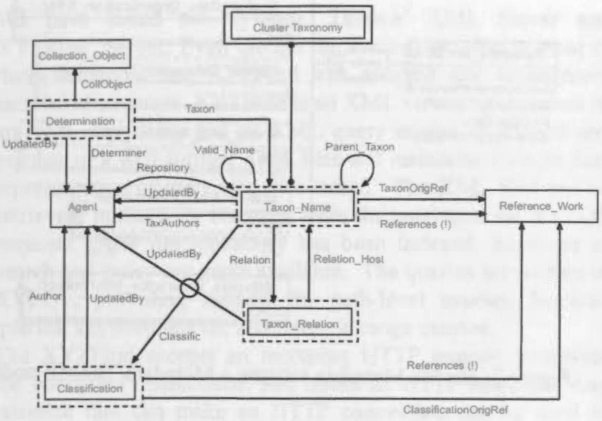


Figure 2 – Cluster Taxonomy and its Relationships

EXAMPLE OF CLOSI DESCRIPTIONS

This example uses real data from INPA's Ichthyology Collection. The information presented corresponds to the object classes and relationships of Figure 2. The Figure does not present the list of attributes of object classes.

Object Class Taxon_Name
Attribute CommonNameandRegion: *Amazonicopeus elongates*
Attribute Relation: Parasitism
Attribute Repository: INPA - Ichthyologic Collection
Attribute TaxAuthors: V. E. Thatcher
Attribute ParentTaxon: *Amazonicopeidae*
Attribute TaxonOrigRef: V. E. Thatcher, 1986
Attribute References (!): V.E. Thatcher,1991; M.A.P.M. Amado, 1995
Attribute Classific: Arthropoda, Crustacea, Maxilopoda, Copepoda, Poecilostomatoida, Amazonicopeidae, Amazonicopeus, Amazonicopeus elongatus

Object Class Taxon_Relation
Attribute RelType: Parasitism
Attribute Remarks: This taxon is a parasite of PESCADA fish *(Plagioscion quamosissimus)* and the place of fixation at the gill arches.
Attribute RelationHost: *Plagioscion squamosissimus*

Object Class Classification
Attribute Name: Arthropoda
TypeTaxonRankName: Phylum
Atribute Name: Crustacea
TypeTaxonRankName: Subphylum
Attribute Name: Maxilopoda
TypeTaxonRankName: Class
Attribute Name: Copepoda
TypeTaxonRankName: Subclass
Attribute Name: Poecilostomatoida
TypeTaxonRankName: Order
Attribute Name: Amazonicopeidae
TypeTaxonRankName: Family
Attribute Name: Amazonicopeus
TypeTaxonRankName: Genus
Attribute Name: Amazonicopeus elongatus
TypeTaxonRankName: Species
Attribute ClassificationOrigRef: K. Bowman
Attribute Author: K. Bowman
Attribute UpdatedBy: C. Magalhães

Object Class Determination
Attribute CollObject: *Amazonicopeus elongatus*
Attribute Taxon: *Amazonicopeus elongatus*
Attribute Determiner: V.E. Thatcher
Attribute Date: 15-07-1986
Attribute UpdatedBy: C. Magalhães

Object Class Agent [In Cluster Agent of Collection]
Attribute FistName: V.
Attribute MidName: E.
Attribute FamName: Thatcher

Attribute FistName: C.
Attribute MidName:
Attribute FamName: Magalhães

Object Class Refence_Work [In Cluster Reference]
Attribute Author: T. Bowman; L. G. Abele
Attribute Title: Classification of the recent Crustacea. P. 1-27. In Abele, L.G. (ed.), Systematics, the Fossil Record, and Biogeography. New Yourk, Academic Press. 319 pp. (The Biology of Crustacea, v.1).
Attribute Year: 1982

Attribute Author: V.E. Thatcher
Attribute Title: The parasitic crustaceans of fishes from the Brazilian Amazon, 16, *Amazonicopeus elongates* gen. Et sp. Nov. (Copepoda: Poecilostomatoida) with the proposal of Amazonicopeidae fam. Nov. and remarks on its patogenicity. Amazoniana, 10(10):49:5.
Attribute Year: 1986

Attribute Author: M.A.P.M. Amado; J-S. Ho; C.E.F. Rocha
Attribute Title: Phylogeny and biogeography of the Ergasilidae (Copepoda, Poecilostomatoida), with reconsideration of the taxonomic status of the Vaigamidae. Contributions to Zoology, 65(4): 233-243.
Attribute Year: 1995

3. METADATA STANDARD FOR BIO DATA

The Federal Geographic Data Committee (FGDC) coordinates the development of the National Spatial Data Infrastructure (NSDI). The NSDI encompasses policies, standards, and procedures for users to cooperatively produce and share geodata. Metadata or "data about data" describe the content, quality, condition, and other characteristics of data. The FGDC had approved the Content Standard for Digital Geospatial Metadata. The objectives of the standard are to provide a common set of terminology and definitions for the documentation of digital geospatial data. The standard establishes the names of data and associated elements to be used for these purposes, the definitions of these compound elements and data elements, and information about the values that are to be provided for the data elements (FGDC, 1998).

In 1999, the FGDC Steering Committee endorsed the Biological Data Profile of the Content Standard for Digital Geospatial Metadata (CSDGM). The purpose of this standard is to provide a user-defined or theme-specific profile of the CSDGM to increase its utility for documenting biological resources data and information. This standard supports increased access to and use of biological data among users on a national (and international) basis. It also helps to broaden the understanding and implementation of the FGDC metadata content standard within the biological resources community. This standard also serves as the metadata content standard for the National Biological Information Infrastructure (NBII). The extended standard can be used to specify metadata content for the full range of biological resource data and information. This includes biological data, which are explicitly geospatial in nature, as well as data, which are not explicitly geospatial (such as data resulting from laboratory-based research). This also includes "information" categories, such as research reports, field notes or specimen collections (FGDC, 2001).

3.1 INTEROPERABILITY OF BIO METADATA

XML, the Extensible Markup Language, provides an intuitive method for data structuring and organisation, based on tags written in text files. XML allows to set standards, defining the information that should appear in a document, and in what sequence. XML, in combination with other standards, makes it possible to define the content of a document separately from its formatting, making it easy to reuse that content in other applications or for other presentation environments. Most important, XML provides a basic syntax that can be used to share information between different kinds of computers, different applications, and different organizations without needing to pass through many layers of conversion.

XML provides a simple format that is flexible enough to accommodate diverse needs. Even developers performing tasks on different types of applications with different interfaces and different data structures can share XML formats and tools for parsing those formats into data structures those applications can use. It offers to users many advantages, including: simplicity, extensibility, interoperability, openness, and a core of experienced professionals. XML operates on two main levels: first, it provides syntax for document markup; and second, it provides syntax for declaring the structures of documents.
XML is derived from the Standard Generalized Markup Language (SGML). SGML has found its main customer base in organizations handling large quantities of documents. SGML's development provides the foundations for XML, but XML has a smaller and simpler syntax, targeted at Web developers and others who need a simple solution to document creation, management, and display.

The features of XML indicate it is suitable language for metadata representation. We use the XML schema for the definition of a BioMetadata XML file template implementing the biological standard profile, allows representing all the characteristics of the metadata elements like hierarchy, repeatability, optional or mandatory items, selection, attributes, data types and values based on integrity constraints. This processs allow schema validation to ensure well-formed XML file outputs. The Figure 3 presents the node BioMetadata schema, implemented from the FGDC standard specifications and mapped to a XML template.

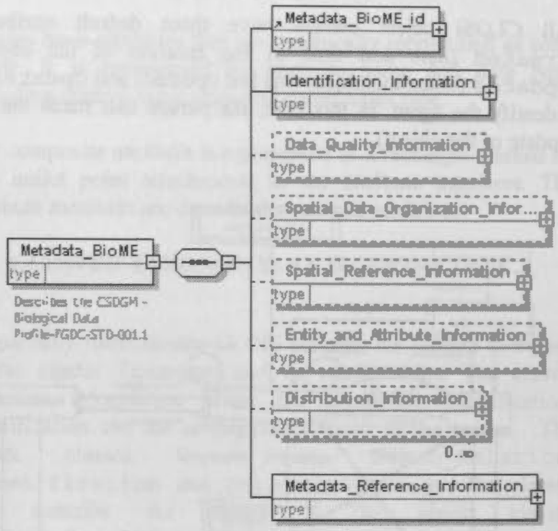

Metadata_BioME
type
Describes the CSDGM - Biological Data Profile-FGDC-STD-001.1
Metadata_BioME_id
type
Identification_Information
type
Data_Quality_Information
type
Spatial_Data_Organization_Infor...
type
Spatial_Reference_Information
type
Entity_and_Attribute_Information
type
Distribution_Information
type
0..∞
Metadata_Reference_Information
type

Figure 3: BioMe Metadata schema – Metadata_BioME node –

4. INFRASTRUCTURE FOR METADATA AND DATA MANAGEMENT

Aiming to provide users access to biological data and metadata information, we have implemented a computational infrastructure mainly based on Open Source Technology (OST), where data broker and providers can store, modify and disseminate data and metadata through Internet. OST allows the development of powerful and fast database-generated web applications. The systems selected represent a compromise between software functionalities and costs. The objective was to provide a technical solution with a low cost, matching the requirement imposed at some institutions in the Amazon, that is, lack of funds for such initiatives. The solution comprises of a Client/Server architecture, where the interactions are performed through the Apache[2] Server. At the server side, we implemented the BioME Components, XML Server and Database Server Services.

BIOME XML COMPONENTS

The XML metadata BioME schema implements the FGDC standard, which includes the biological profile. Since the XML schema can ensure a well-formed, XML corresponding document, a template was developed, that is the XML file for BioME. The Metadata XML Documents hold descriptive material about the metadata schema, the html version and FGDC full documentation. Additionally, we made available the

---

[2] Apache is a product of the Apache Software Foundation; which provides support for the Apache community of open-source software projects.

XMLmind XML Editor (XXE[3]), an editor featuring a word processor-like view. It has been developed to make users comfortable and productive at editing XML documents and XML data. Users at a client environment can download XXE together with the Metadata XML template. Allowing the deployment of the XXE tool to users was the best option since Internet connections in the Amazon region present limitations in performance, and also the task to produce a metadata is time consuming. The template can be loaded into XXE and users can insert information that describes their metadata. All documents available for download can also be accessed for browsing via the Web interface. Once the metadata has been concluded, it is required that the metadata be registered and posted to the BioME site.

XML SERVER

We have tested two systems, Tamino[4] XML Server and XYZFind[5] Server. Even though the systems overlap in most of their functionalities, XYZFind was adopted due to its more simplified structure. XYZFind is an XML server that consists of an XML repository and an XML query engine. It accepts any number of a well-formed XML files and maintains a single data representation of the entire files stored. The XML files can be retrieved, updated, or removed from the catalogue via XYZQL requests. Once the repository has been indexed, functions of search and query are made available. The queries are written in XYZQL, providing support for path-level queries, Boolean queries, keyword search, and numeric range queries.

The XYZFind accepts an incoming HTTP request, processes the contained instructions, and issues an HTTP response. Any resource that can make an HTTP connection can be used to write an application that takes advantage of the server functions.

DATABASE SERVER

CLOSi schema descriptions (biological collections object classes, relationships) can be mapped into a relational database management system. We selected MySQL[6], because it is a fast, multi-threaded, multi-user and robust SQL database server.

The implemented architecture is presented in Figure 4.

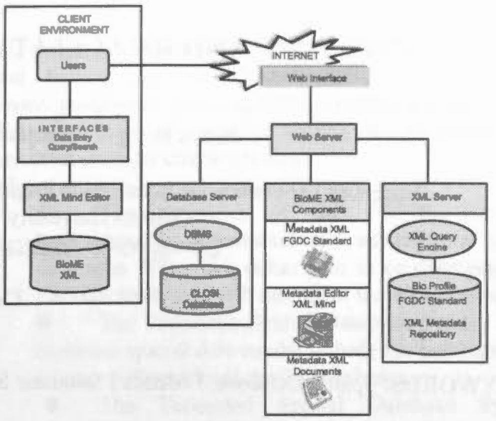


Figure 4: Client/Server architecture for data and metadata management

5. CONCLUSIONS

The present infrastructure has been developed and installed at ITC and is under test using data from INPA's Fish and Crustaceous collection. Researchers at INPA have started to produce metadata based on the XML template and editing it using the XMLmind Editor, both components deployed from the BioME XML Component site. A CLOSi schema of an entomological collection has been tested CLOSi partially. The results so far are encouraging and the Crustaceous data set has already started to be designed for its full extension. The current database implementation does not provide facilities to query via the web yet, as it is provided when querying the Metadata repository. The next step will be to provide XSL stylesheets for completed XML metadata files in order to render HTML pages over the web. In case of multiple output formats several XSL can be designed, that is, one stylesheet can render to HTML Navigator, Explore, Lynx, and to PDF for printing.

REFERENCES

Campos dos Santos, J.L., de By, R.A. & Magalhães, C. 2000. A case study of INPA's bio-DB and an approach to provide an open analytical database environment. International Archives of Photogrammetry and Remote Sensing, 33 (B4): 155-163.

Chen, I. A. & Markowitz, V. M. 1996. The Object-Protocol Model, Technical Report LBNL 32738, Lawrence Berkeley National Laboratory, Information and Computing Science Division, 1 Cycloton Road, Berkeley, CA 94720, June 1996.

DuBois, P. 2000. MySQL. New Riders Publishing, New York.

FGDC. 1998. Content Standard for Digital Geospatial Metadata. Federal Geographic Data Committee. Washington, D.C. pg. 85.

FGDC. 2001. Biological Data Profile Workbook. Federal Geographic Data Committee. Washington, D.C. pg. 148.

Graham, I.S. 1996. HTML Source Book. John Wiley & Son, London etc.

Sonderegger, J., Petry, P., Campos dos Santos, J.L. & Alves, N.F. 1998. An entomological collections database for INPA. In: Ling, T.W.; Ram, S. & Lee, M.L. (eds.), Proceedings of the 17th International Conference on Conceptual Modeling - ER '98. Singapore. p. 421-434.

---

[3] XXE Version 1.0 (October 2001) is a Pixware SARL product – Montigny Le Bretonneux, France.

[4] Tamino is a trademark of Software AG – Darmstadt, Germany.

[5] XYZFind is a trademark of XYZFind Corporation – Kirkland, WA, USA.

[6] MySQLis a trademark of MySQL AB. It can be used under the GNU General Public License (DuBois, 2000).

CONCEPT AND REALIZATION OF FEDERATED SPATIAL DATABASE

Jianya Gong Hanjiang Xiong Yandong Wang Lite Shi

State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS)
Wuhan University, 129 Luoyu Road, Wuhan, China, 430079
(jgong, wyd, xhj)@rcgis.wtusm.edu.cn, lilyshi@hp01.wtusm.edu.cn

Commission IV, WG VI/2

KEYWORDS: Spatial Database, Federated Database System, Geographic Information System

ABSTRACT:

Geographic information systems manage a large volume of spatial data. We often use it to solve some problems such as urban planning and land use management. Interoperability of GIS is the ability to access spatial data and associated services in a distributed and heterogeneous processing environment. There are many existing spatial databases and GIS applications that are built on those databases today. In order to solve problems, we often need the cooperation of many spatial databases, even if there are more complex spatial relationship and diversity between the existing spatial data models. In this paper, we will discuss the concept and methodology of a Federated Spatial Database System to implement the interoperability of GIS. At the end of paper, we introduced a project named VirtualWorld 2000. The goal of this project is to build a federated spatial database system to access the existing component spatial databases distributed in the LAN, which include graphic databases, DEM databases and image databases built with Oracle8i spatial cartridge, SQL Server 7.0 and GeoStar 3.0 file system.

1. INTRODUCTION

Geographic information systems manage a large volume of spatial data. We often use it to solve some problems such as urban planning and land use management. Interoperability of GIS is the ability to access spatial data and associated services in a distributed and heterogeneous processing environment [5]. There are many existing spatial databases and GIS applications that are built on those databases today. In order to solve problems, we often need the cooperation of many spatial databases, even if there are more complex spatial relationship and diversity between the existing spatial data models.

Interoperability among traditional geographical information systems requires solving two major problems:

- How to access spatial data distributed on the network.
- How to allow cooperation between existing heterogeneous spatial databases.

The first problem is dealt with by using networking techniques based on the client/server paradigm. For example a distributed object environment, such as CORBA can be used to provide interconnectivity among several geographical information systems.

The second problem has recently been the focus of extensive researches in the field of databases. Several solutions have been identified:

- Schema integration based solutions aim to combine all information into a single global schema.
- Canonical data-model based solutions hide heterogeneity and provide a good framework for multi-model translations.
- Multi-database language based solutions allow users to query several information sources at the same time.

In addition to the above problems, interoperability of GIS must address the following issues:

- Support for spatial data which exhibit more complex relationship than traditional applications;
- Support for spatial operators as spatial analysis and spatial query;
- Support for new types of heterogeneity specific to spatial information systems (generic, spatial contextual, models, and operators heterogeneity).

In section 2, we will discuss the distributed spatial data model of the federated spatial database system. We will build a distributed spatial data model that extends the Open GIS Simple Features Specification in order to manage the DEM and image data. In section 3, we will discuss the architecture of the federated spatial database system. We will build architecture with the middleware components and spatial agents. In section 4, we will discuss the spatial data model integration between the component spatial databases. In section 5, the spatial query and operators of the federated spatial database system will be discussed. We need to define the visual spatial query language for the federated spatial database system. In section 6, the spatial object migration in the federated spatial database system will be discussed as well.

At the end of this paper, we will introduce a project named VirtualWorld 2000, the goal of which is to build a federated spatial database system to access the existing component spatial databases distributed in the LAN, which includes graphic databases built with Oracle8i spatial cartridge, SQL Server 7.0 and GeoStar 3.0 file system, digital terrain databases and digital ground image databases obtained with photogrammetry or remote sensing technology. We can build the 2D or 3D applications on this system and keep those spatial databases autonomous, that is to say, all the existing applications built on them are not necessary to be rebuilt. Also, it is easy for us to add or delete a component spatial database.

2. THE CONCEPT OF THE FEDERATED SPATIAL DATABASE SYSTEM

In 1990, Kamel represented the concept of the Federated Database Management System [2]. This approach allows users and applications to access data across several heterogeneous databases (relational, object-relational and object –oriented DBMS) while maintaining their autonomy. Soutu. C discussed the methodology for developing a federated database system [3]. Many experts had discussed the query, schema transformation and other issues about the Federated Database System.

The integration of spatial databases has encountered the same problem. The cooperation of several spatial databases demands that the application should access those spatial databases transparently. With the theory, we can build a federated spatial database system to implement the interoperability of GIS. A Federated Spatial Database System (FSDBS) consists of autonomous component spatial database systems that participate in the federation to allow controlled sharing of their spatial data. There are several major characteristics:

- The members of the FSDBS may be heterogeneous, distributed and autonomous spatial databases. They can either join in or disengage the FSDBS freely and will not affect the whole system.
- The Federated Spatial Database System has a common spatial data model to hide the heterogeneity among different local spatial databases.
- The Federated Spatial Database System provides a set of global spatial query languages to query and access the local spatial databases.

3. THE COMMON SPATIAL DATA MODEL OF FSDBS

A Federated Spatial Database System has many local spatial databases. These local spatial databases will have different spatial data models. So we should define a common spatial data model to support a set of spatial concepts to hide heterogeneity between spatial models. The Open GIS Consortium (OGC) has presented us a good spatial model to integrate the different spatial models. Also, OGC has released an implementation specification in 1999 - the OpenGIS Simple Feature For OLE/COM Revision1.1. As more and more software packages have been adopting this specification, we can adopt it as the common data model of FSDBS as well. But the model has a deficiency, that is, the specification cannot define the image objects and DEM objects. We hope that the local spatial databases of the FSDBS can be DEM, image or 3D vector databases. The common model should be extended in the following way (Figure 1):

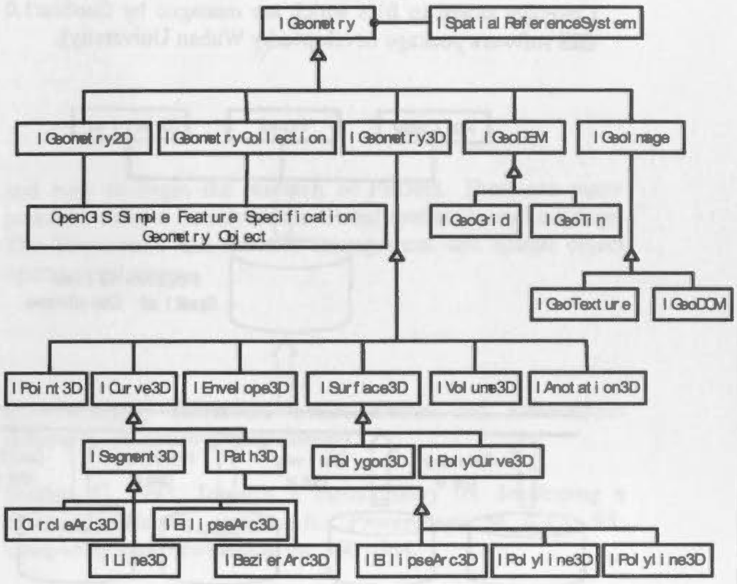

IGeometry
ISpatialReferenceSystem
IGeometry2D
IGeometryCollection
IGeometry3D
IGeoDEM
IGeoImage
OpenGIS Simple Feature Specification Geometry Object
IGeoGrid
IGeoTin
IGeoTexture
IGeoDOM
IPoint3D
ICurve3D
IEnvelope3D
ISurface3D
IVolume3D
IAnotation3D
ISegment3D
IPath3D
IPolygon3D
IPolyCurve3D
ICircleArc3D
IEllipseArc3D
ILine3D
IBezierArc3D
IEllipseArc3D
IPolyline3D
IPolyline3D

Figure 1 The Extended Spatial Object Architecture for FSDBS

4. THE SCHEMA DEFINITION AND INTEGRATION OF FSDBS

According to the methodology of developing a federated database system [3], the architecture of a federated database system can be described by a five-level schema. There is five-level schema: local schema, component schema, exported schema, the federated schema and the external schema.

Spatial Data Coverage is the core concept of the five-level schema. We can see the FSDBS as:

1) All the local spatial databases are consisted of several exported spatial data coverages.
2) FSDBS has several federated spatial data coverages. Each federated spatial data coverage is coming from one or more exported spatial data coverages.
3) From the users' point of view, the FSDBS is consisted of several users' spatial data coverages. Each user's spatial data coverage is also coming from the federated spatial data coverage.

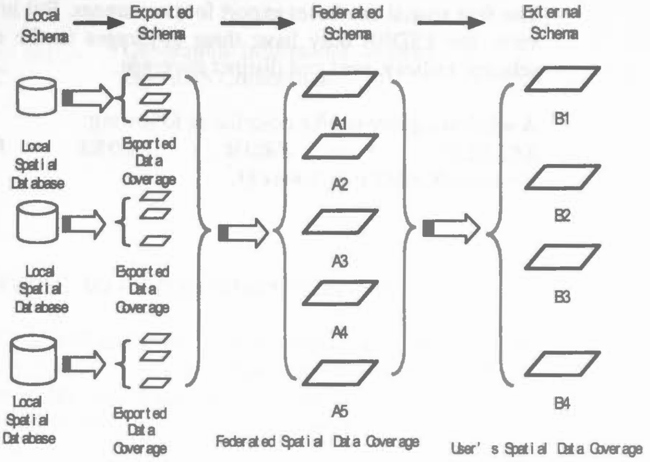

Local Schema
Exported Schema
Federated Schema
External Schema
Local Spatial Database
Exported Data Coverage
Local Spatial Database
Exported Data Coverage
Local Spatial Database
Exported Data Coverage
A1
A2
A3
A4
A5
Federated Spatial Data Coverage
B1
B2
B3
B4
User's Spatial Data Coverage

Figure 2 The Schema and Transformation of FSDBS

5. QUERYING THE FSDBS

In order to access the heterogeneous, distributed and autonomous spatial databases, we should have a spatial query language for the FSDBS. All the user's spatial data requests are translated into the global query language. The FSDBS discomposes the query into several sub-query commands and send them to each local spatial database. The spatial databases execute the command with the local mode, and retrieve the result to FSDBS. The FSDBS then integrates the result and retrieve the result to the user. In order to interpret the procedure of the execution of user's query, an example is given as follows:

Suppose that we have a Chinese digital map with a scale of 1:1000000. The digital map has three data coverages: the railway coverage stored in SQL Server 6.5, the road coverage stored in Oracle8.1.6, the Province coverage and County Coverage stored in files which are managed by GeoStar3.0 (a GIS software package developed by Wuhan University).

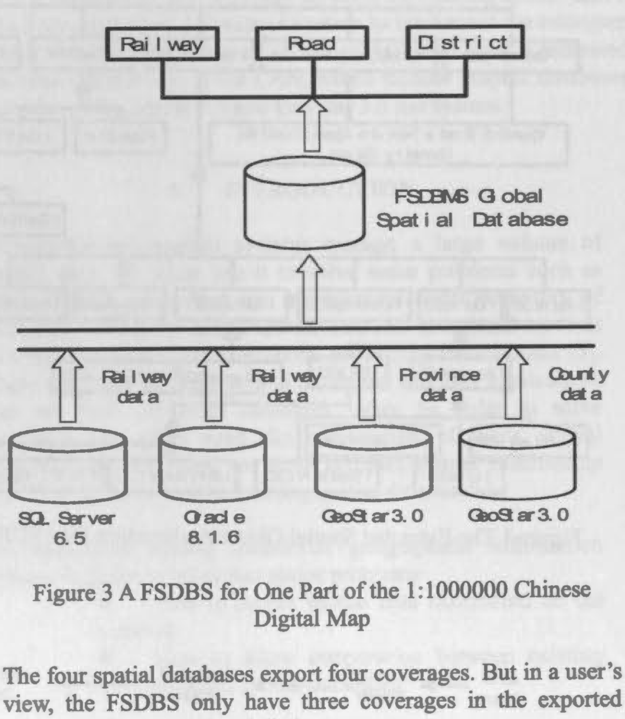


Figure 3 A FSDBS for One Part of the 1:1000000 Chinese Digital Map

The four spatial databases export four coverages. But in a user's view, the FSDBS only have three coverages in the exported schema: railway, road and distinct coverage.

A windows query maybe describe as following:
SELECT * FROM FSDBS WHERE Rect(minX,minY,maxX,maxY).

It notes that "FSDBS" means the query is execute in all the spatial databases that are members of the federal.

The procedure of the FSDBS spatial query will be as follows:

1) To query the global metadata database and decompose the query command into three absolute sub-query commands:
SELECT * FROM Coverage Railway WHERE Rect(minX,minY,maxX,maxY)
SELECT * FROM Coverage Road WHERE Rect(minX,minY,maxX,maxY)
SELECT * FROM Coverage Distinct WHERE Rect(minX,minY,maxX,maxY)

2) With the meta database, each of the three sub-query commands will be decomposed to several sub-queries. This decomposition should be insured that each sub-query command only relates to one local spatial database.

(1) The command, "***SELECT * FROM Coverage Railway WHERE Rect (minX,minY,maxX,maxY)***", will be sent to the wrapper of SQL Server 6.5.

(2) The command, "***SELECT * FROM Coverage Road WHERE Rect (minX,minY,maxX,maxY)***", will be sent to the wrapper of the Oracle 8.1.6.

(3) The command, "***SELECT * FROM Coverage Distinct WHERE Rect (minX,minY,maxX,maxY)***", will be decomposed into two sub-query commands: ***SELECT * FROM Coverage Province WHERE Rect(minX,minY,maxX,maxY)***, which will be sent to the wrapper of spatial database managed by GeoStar3.0, and ***SELECT * FROM Coverage County WHERE Rect(minX,minY,maxX,maxY)***, which will be sent to the wrapper of another spatial database managed by GeoStar3.0

3) Each wrapper converts the query from global spatial query language to local spatial query language or procedure. For example, the query: ***SELECT * FROM Coverage Road WHERE Rect(minX,minY,maxX,maxY)***, which has been sent to the wrapper of Oracle 8.1.6 should be converted as following: ***SELECT * FROM Road WHERE sdo_filter(Road.GeoObject, mdsys.sdo_geometry{2003,NULL,NULL,mdsys.sdo_elem_info_array(1, 1003,3),mdsys.sdo_ordinate_array(minX,minY,maxX,maxY)}, 'querytype=window')='TRUE'***, and sent to Oracle 8.1.6 DBMS.

4) Each local spatial database executes the sub-query and retrieves the result to FSDBS with the wrapper's schema transformation. The FSDBS also execute the schema transformation, integrate the result and retrieve it to the user.

6. THE IMPLEMENTATION OF FSDBS

We have built a project named as VirtualWorld 2000, the motivation of which is to build a FSDBS to manage the existed DEM, image and vector graphic databases and provide the services for urban planning, while keeping the old applications built on these databases not to be modified. The architecture of VirtualWorld 2000 is as follows (Figure 4):

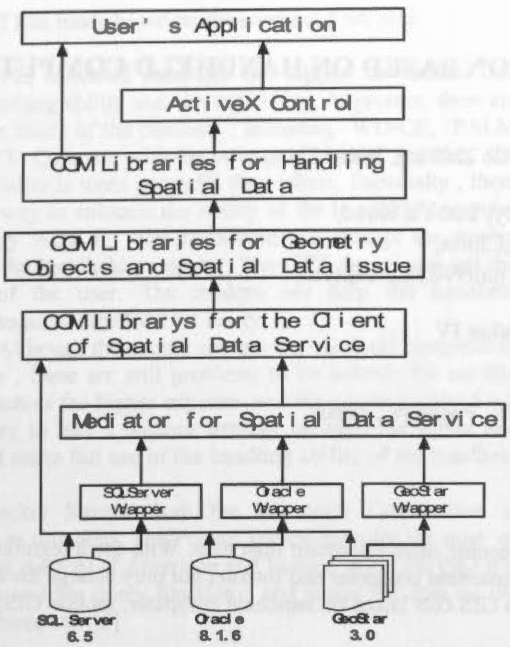

User's Application
ActiveX Control
COM Libraries for Handling Spatial Data
COM Libraries for Geometry Objects and Spatial Data Issue
COM Librarys for the Client of Spatial Data Service
Mediator for Spatial Data Service
SQLServer Wapper
Oracle Wapper
GeoStar Wapper
SQL Server 6.5
Oracle 8.1.6
GeoStar 3.0

Figure 4 The Architecture of VirtualWorld 2000

7. CONCLUSIONS

The interoperable GIS is becoming more and more important today. The FSDBS provides us a new idea to build an interoperable GIS platform. This paper shows what is a FSDBS and how to begin the research of FSDBS. There are many problems need to be solved: the Global spatial Query Language, The Transaction and Security management and spatial object operator and so on.

REFERENCES

Holger, G, 2001. Object-oriented modeling of data sources as a tool for the integration of heterogeneous geoscientific information. *Computers & Geosciences*, 27, pp. 975-985.

Kamel, M. N. and Kamel, N. N., 1990. The federated database management system: an architecture of distributed systems for the 90's. In: *Proceedings of the 2nd IEEE Workshop on Future Trends*, pp. 346 –352.

Sheth, A. P. and Larson, J. A., 1990. Federated database systems for managing distributed heterogeneous and autonomous databases. *ACM Computing Surveys*.

Soutou, C. 1993. Towards a methodology for developing a federated database system. In: *Proceedings of ICCI '93: Computing and Information*, pp. 560 –564.

Tryfona, N. and Sharma, J., 1996. On information modeling to support interoperable spatial databases. In: *Proceedings of the 8th International Conference on Advances Information Systems Engineering, CaiSE'96*, Vol. 1080, Springer, pp. 210-221.

ABBREVIATIONS

FSDBS	Federated Spatial Database System
GIS	Geographic Information System
OGC	OpenGIS Consortium

RESEARCH OF MOBILE GIS APPLICATION BASED ON HANDHELD COMPUTER

Liu Yong , Li Qing Quan, Xie ZhiYing, Wang Chong

WuHan University, LuoYu Road,
WuHan,China,
lycourage@263.net , qqli99@hotmail.com

Commission IV

KEY WORDS: Mobile GIS, handheld computer ,application mode , PocketPC, J2ME

ABSTRACT:

With the coming of the digital age, geography information will become more important than ever. With the integration of the new technology, GIS make new achievement. The development of the handheld computer and Internet not only enlarge the applying field of the GIS, but also bring the new opportunities and challenges to GIS.GIS based on handheld computer, mobile GIS, will become the new branch of GIS and bring the GIS into a brand new world.

In post-pc age, the mobile device including the handheld computer will become the important field to serve the public for information service. With handheld computer, people can breakthrough the limitation of the GIS based on PC and enjoy the geography information freely., Mobile GIS makes geographical information as carrier and integrate all kind of information and implement the dynamic updating of information. So mobile GIS will make GIS apply widely and become one of the popular tools for the public.

Compare with traditional GIS, Mobile GIS based on handheld computer has many new functions. With the mobility character and extension for hardware and software character, the handheld computer will become the important part of the mobile GIS. With the GPS and communication device, people can use handheld device anywhere, anytime to access any information they wanted. That will meet the increasing demand of geographical information in information age. More and More scholars realize the importance of the integration between geographical information and wireless INTERNET and start the research and application of these fields.

After a systematic analysis to the handheld and related technologies, the strategy of applying these technologies to the mobile GIS and three applying mode are presented. The whole frame of the system is put forward along with system development which based on it is finished. At last, the system is verified.

1. INTRODUCTION

Recently, with the new challenge in the work and life, personal computer can not meet the demand of the people in many situation. So not only the individuals but also enterprise customers hope to access the information under the mobile environment .The development of the handheld computer make us come into the post-pc age. Especially, as for the 80 percent of the whole information is related to the spatial information, people want to get the spatial information related to current location under the mobile environment. .[1]

Under this circumstance, the public and the professionals need spatial information service system .The spatial service application based on the handheld computer development very fast , because the handheld computer have mobility , practicability and extensibility character. These character make the handheld computer widely used by the people.

Mobile GIS make great achievement with the development of the wireless Internet and the mobile device . It is the new direction of the GIS and make the GIS easily accepted by the public.[2]. On one hand , the mobile device and the communication system develop very fast, the handling ability of the mobile device is equal that of the Pentium of the PC and the communication system is make a transition from second generation to third generation ; On the other hand ,the related software technology including GML(geographical extensible language), SVG(scalable vector graphics) and J2ME also accelerate the development of the Mobile GIS.In general ,Mobile GIS develop with the WEB-GIS technology, and there are many similar characters of two technology.

But mobile GIS show its own character because of the limitations of the mobile device and the wireless communication technology.

As we know , in the field of Mobile GIS there are still lots of problems to be solved. At present, many GIS companies research on Mobile GIS and bring up the product ,such as the MapInfo Mobile Location Suite(MMLS) , Intelliwherer Suit from the Intergraph, ArcPad from ArcInfo and so on. Because of the limit of the wireless communication and the mobile device, it is very hard to apply mobile GIS widely.

After the analysis of the current technology of handheld computer and wireless communication , the paper bring up three mode for mobile GIS based on handheld computer and the framework of the spatial information service system based on handheld computer. Considering the condition of the wireless communication and mobile device, the prototype of the system is finished.

2. SYSTEM APPLICATION MODE

At present, the hardware and software of the handheld computer develop very fast, the people can get the information with many method in mobile situation. As far as the method of the data service , there are three mode of mobile GIS based on the handheld computer.:

1) Off line mode based on the compact flash card

The handheld computer can support the mobile GIS for its handling ability and storing ability. At present, there are three main kinds of the product , including WINCE, PALM and LINUX. Compare with the other two kind of product, the handing ability is more powerful than others. Especially , there are many way to enhance the ability of the handheld computer powered by WINCE. The flash card can enlarge the storing ability of the handheld computer. The GPS device can get the location of the user. The modem can help the handheld computer to access the Internet easily.

Although the people can use the handheld computer to be on line , there are still problems to be solved for on line service such as the higher expense and the power supply. So it is necessary to take a balance strategy between the server and client and make full use of the handling ability of the handheld computer.

The Pocket Streets from the Microsoft Corporation is belonging to this kind. The Pocket Streets includes the map of the most of the city of American and Europe ,supports the GPS navigation and the query function and stores the data on the handheld computer.[5]

2) Wired Internet mode

This mode is similar to the web-GIS based on the PC. Figure1 show the structure of this mode. The client sends the request to the server ,the server send the content back to the client. The interactive of the server and the client is based on the HTTP. When client send a URL with the HTTP header, the application on the web could be the application based on static HTML page, CGI or SERVERLET and so on, both of them will send the content back with the HTML format, then the page will be showed on the web browser(Figure 1).

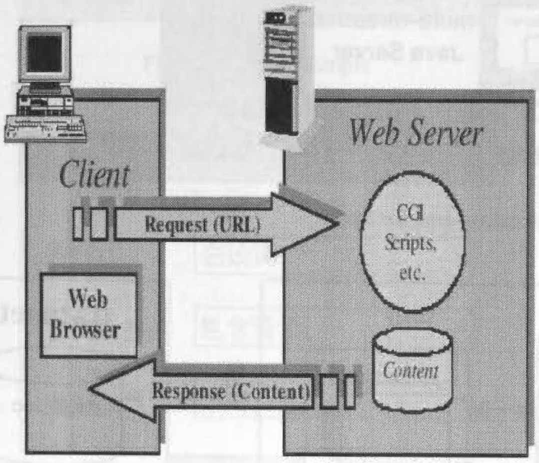


Figure 1. World-Wide Web Programming Model

In this mode, the handheld computer need the support of the browser or the client application. There are also two way of this mode , including the C/S mode and B/S mode.

As for the B/S mode , the handheld computer can exchange the data with server by modem .But for the limitation of the handheld computer , the complicate format HTML can not be showed on the browser of the handheld computer.

As for the C/S mode , it is necessary to finish the additional work to exchange the data with the server, but it is more flexible than B/S mode. The ArcPad from the ESRI is belonging to this kind. With the technology of the ESRI, ArcPad can support not only the raster data but also the vector data. It also can be the client of the ArcIMS [6]

3) Wireless Internet mode

Wireless Internet is new trend of the IT and it can take the advantage of the handheld computer. The handheld computer can access the wireless Internet by mobile phone or wireless modem. There are also two mode B/S mode based on WAP and C/S mode based on Java.

WAP , wireless application protocol is supported by most of the country in the world. The mobile user can access WAP server to query the information and the WAP server can transfer the image and text to the client .Because the handling ability of the mobile device is limited ,so the interactive function is limited and most of the function is fulfilled by the server. The following figure is the workflow of the WAP.[3] Figure2.

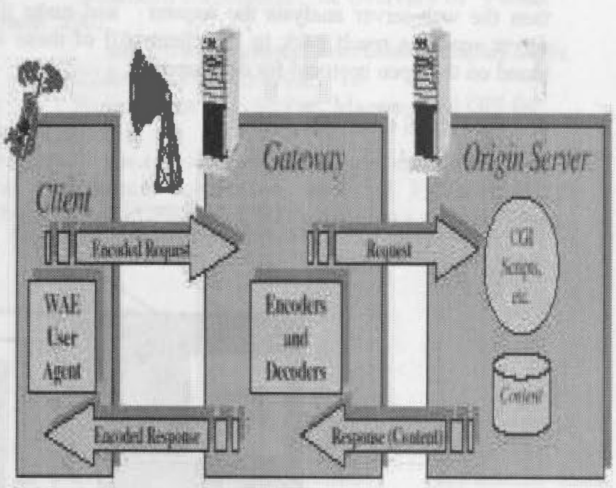


Figure 2. WAP Programming Model

J2ME is another way to support data service .It is technology from the SUN corporation. The most obvious advantage is that the interactive function can be fulfilled on the client-side and its platform independence[7] . With the development of the communication of the communication system , the mobile device will support more complicated interactive function . And the J2ME technology is the supplement of the WAP. The wireless part of the system is based on the WAP.

4) System mode

In conclusion, three mode have its own advantage and disadvantage. Off line mode can work independently ,it is fit for the navigation and information collection; And on line mode can get the dynamic information ,it is fit for the mobile target management and emergency help and so on. After the analysis on the mobile GIS, the solution is proposed. In the server side , MapXtreme for Java is serving as the map server, with the JSP/SERVERLET technology. In the client side , Pocket IE is serving as the browser to get the spatial and non-spatial information .Off line application , Mobile City 1.0 based on the PocketPc is finished with the support of the eMbeded Visual C++.

3. THE THEORY AND IMPLEMENTATION ON THE CLIENT SIDE

1. B/S mode (thin client)

The handheld computer powered by WINCE has its own browser for WEB and WAP. Especially the HTML and WML can be browsed on the PocketIE2002. So the three-tier structure can be adopt to the handheld computer, Not only the HTML but also WML can be showed on the WINCE

With the support of the javascript , the client can get the data including the image ,text and so on by modem. As for the wireless spatial service, The interactive function is fulfilled by wml and wmlscript. As for the wireless communication , WAP is an very important technology. WAP is wireless application protocol , it is an open communication protocol to share the data with the mobile phone , Internet ,PDA and so on.[3]. It is widely support by most of the country in the world . The workflow of the B/S mode is that the client send the request , then the web-server analysis the request and make the map server send the result back to the client .All of these work is based on the open protocol for development .

2. C/S mode (fat client)

Because the limitation of the handheld computer and bandwidth of the Internet , it is necessary to balance the burden between the server and client . The basic spatial information is stored on the handheld computer and the dynamic information will send to the handheld computer if necessary.

The key point of the design of the fat-client is to put the data used usually on the handheld computer . The following function will be fulfilled on the client, including the map function , query function ,navigation function and so on. The system is development by eMbedded Visual C++.

4. THE THEORY AND IMPLEMENTATION ON THE SERVER SIDE

The main function on the server-side is to handle the request from the client and query from the database ,then send the result back to the client. There are many technology can be used ,including the CGI ,APS ,PHP,JSP and so on. In general , the JSP/SERVERLET has it own advantage in many aspects. So the application on the server side is based on the mapxtreme for Java .(Figure 3). The workflow on the server side is to make different of the client , then send back the response result, including WML page, HTML page and self-defined data, the following is the structure of the system(Figure 4).

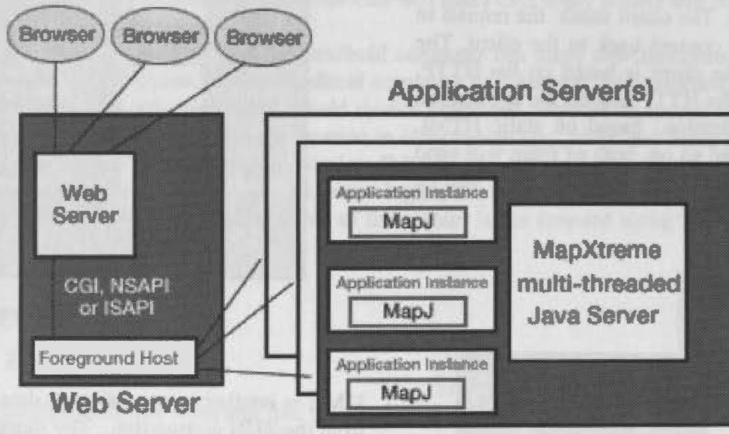

Browser
Browser
Browser
Application Server(s)
Web Server
CGI, NSAPI or ISAPI
Foreground Host
Web Server
Application Instance
MapJ
Application Instance
MapJ
Application Instance
MapJ
MapXtreme multi-threaded Java Server

Figure 3 The workflow of the mapxtreme for Java

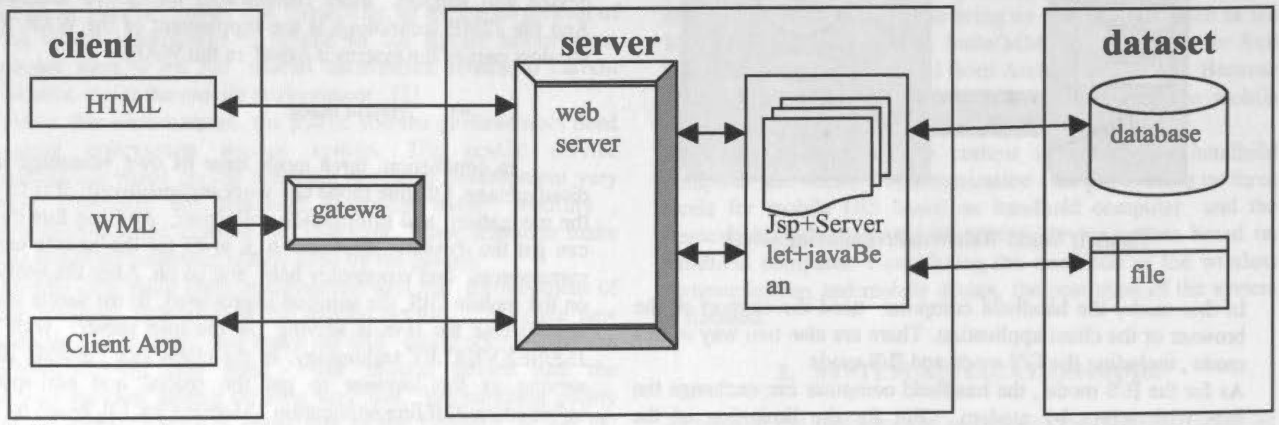

client
HTML
WML
Client App
server
web server
gatewa
Jsp+Serverlet+javaBean
dataset
database
file

Figure4 The structure of the system.

5. THE SYSTEM DEVELOPMENT AND CONCLUSION

The application is finished by the Jbuilder5 , the server side is based on the mapxtreme for Java and the client is based on the embedded Visual C++3, the client is based on the HP jornada548. The WuHan map is used for test and there are also plenty of the information including hotel, hospital ,market and so on. The system function including the map function , query function , road function and so on. The on line service and off line service can be accessed by the prototype of system. The following is the example of the system.

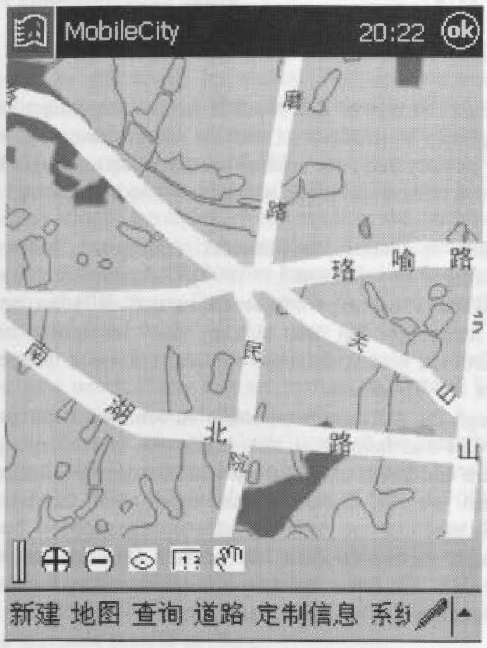
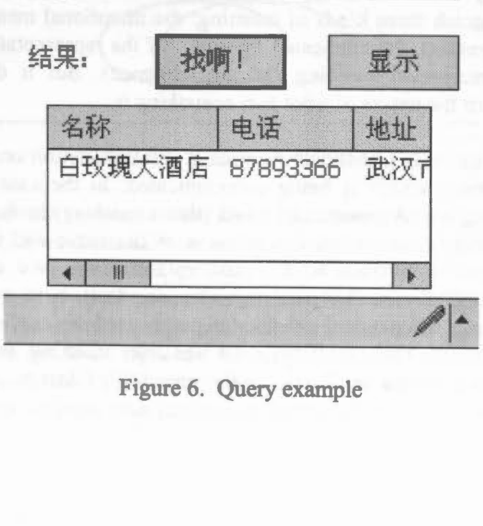


Figure 5. Map example



名称: 白玫瑰

查询类型: 包含查询

类别: 宾馆

结果:

名称	电话	地址
白玫瑰大酒店	87893366	武汉

Figure 6. Query example

The three modes are proposed in the paper based on the handheld computer and the prototype is finished based on the research. And we will modify the system to make it better in the future.

6. REFERENCES

1) Li DeRen , Li QingQuan □2001, The discussion of the integration between spatial information and communicatio , information science of WuHan university. 26□1□:1-7

2) Li QingQuan □Liu Yong , GIS Application based on wireless Internet, 2001 ,.ICC2001, BeiJing.

3) WAP Forum□WAP White Paper□ June 2000

4) MapInfo corporation , The workflow of the mapxtreme for java , 2002 ,http://dynamo.mapinfo.com/miproducts/Architecture.cfm?ProductID=1162

5) Microsoft corporation , the Overview of Pocket Streets ,2002, http://www.microsoft.com/catalog/display.asp?subid=22&site=11069&x=39&y=10

6) Esri corporation, ArcPad, Mapping and GIS for Mobile Systems, 2002http://www.esri.com/software/arcpad/index.html

Sun Microsystem corporation, Java^{TM} 2 Platform, Micro Edition (J2ME^{TM} Platform) ,2002, http://java.sun.com/j2me/

SEMANTIC HETEROGENEITY OF GEODATA

Zhu Xu[a, *], Y.C. Lee[b]

[a] Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Hong Kong, China,
zhu.xu@polyu.edu.hk
[b] Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B. E3B 6Y1, Canada,
yclee@unb.ca
[b] National Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, China (LIESMARS), Wuhan University, Wuhan, China

Commission IV, WG IV/2

KEY WORDS: Semantic heterogeneity, Spatial database generalization, Geodata sharing, Semantic resolution

ABSTRACT: Semantic heterogeneity is a major barrier to geodata sharing. The aim of this paper is to improve understanding of semantic heterogeneity. It provides discussions on what semantic heterogeneity of geodata is, what its sources are, what semantic heterogeneity is irresolvable and how to avoid. The conclusion is semantic poverty and poor conceptual modeling are major forms of semantic heterogeneity. We also argue that spatial database generalization is a critical technique to facilitate geodata sharing.

1. INTRODUCTION

Semantic heterogeneity is a general term referring to disagreement about the meaning, interpretation or intended use of the same or related data. This problem is poorly understood, and there is not even an agreement regarding a clear definition of the problem [Sheth and Larsen, 1990]. The importance of being aware of semantic heterogeneity and doing semantic reconciliation is to guarantee meaningful data sharing, i.e. data exchanged is correctly interpreted and used. It is now well accepted that semantic heterogeneity of geodata is the major impediment to geodata sharing.

The aim of this article is to improve our understanding of semantic heterogeneity of geodata and discuss possible solutions to semantic reconciliation if they are resolvable and how to avoid or compensate irresolvable semantic heterogeneity. The article is organized as follows. In next section, we discuss what semantic heterogeneity is. After briefing the database perspective, we present our linguistic view on semantic problem in geodata sharing and identify sources and forms of semantic heterogeneity. Followed is an analysis sources and forms of conceptual heterogeneity in section 3, which is the major source of irresolvable semantic heterogeneity. We emphasise the special characters of geographic conceptual modelling. Section 4 is devoted to resolvability of semantic heterogeneity. We point out major irresolvable heterogeneity. Finally, we summarize our study in section 6.

2. MEANING, COMMUNICATION AND SOURCES OF SEMANTIC HETEROGENEITY

The understanding of semantic heterogeneity has evolved much during the past two decades. Semantic heterogeneity is perhaps most studied in the domain of information sharing in general and interoperating database in particular. At early stage, semantic heterogeneity mainly refers to the difference in database models and the alternative ways of implementing a certain conceptual model using a database model, usually a relational model and more recently an OO model.

Such heterogeneity is now called structural semantic heterogeneity by Colomb [1997] and schematic heterogeneity by Bishr [1998] and it is found that the more troublesome problem is what is called fundamental conceptual heterogeneity. According to [Colomb 1997], fundamental semantic heterogeneity occurs when terms in two different ontologies have meanings that are similar, but not quite the same. Furthermore, neither database contains sufficient information to resolve the differences.

Although the database perspective on semantic heterogeneity is still valid and helpful, it also limits our understanding of the problem and limits us in a from-within-database solution to this problem. We will discuss in more detail about this later. In the next subsection, we will compare communication by natural language, human-database interaction and database-database interaction to have a deeper understanding of semantic heterogeneity and its sources and forms.

2.1 A working definition of meaning

As mentioned above, semantic heterogeneity occurs when there is a disagreement about the meaning, interpretation or intended use of data. Therefore, it seems necessary to inspect what meaning is, how meaning is conveyed, the process of communication, and how disagreements upon meaning may happen.

According to the American Heritage Dictionary, meaning is interpreted as "something that is conveyed or signified", "something that one wishes to convey, especially by language" or "an interpreted goal, intent, or end." The above statements are largely from the viewpoint of communication and seem to distinguish three kinds of meaning: the intentional meaning (of the speaker), the explicated meaning (of the representation) and the interpreted meaning (of the audience). But it does not concern the nature of what that *something* is.

It is extremely difficult to give a general definition of that *something* which is being communicated. In the case of data sharing, we take meaning of data (that *something*) as the sum of a conceptualization of the universe of discourse and the facts observed according to this conceptualization. We call this meaning the *sensible meaning*. Our emphasis is to make this meaning independent of the language used to explicate and convey it. The counterpart of sensible meaning is *formal meaning*, which we define as the meaning of data in a certain

---

[*] Corresponding author.

language that can be understood and processed by the speakers of that language. When the language is a natural language and the speaker is a human being, formal meaning is just sensible meaning. When the language is some artificial language and the speaker is computer, formal meaning is the part of sensible meaning that is explicated or formalized and can be processed by the computer. One benefit of making the distinction is that we can tell the different between a database system and an information system, i.e. the meaning level of database system is that of that the database language and the meaning level of information system is promoted to sensible meaning due to the participation of human operator.

2.2 A communication model

From language studies [Jeffries 1998], we know that to communicate effectively by natural language we need to *a)* grasp the grammar and a (symbolic) vocabulary of the language, *b)* be equipped with necessary commonsense or domain knowledge to make sense of the vocabulary and *c)* have a correct perception of the context. To illustrate these notions, we give some simple examples. When reading the sentence *"Tom likes coffee"*, we are likely to interpret it as "Someone, named Tom, likes drinking coffer", if we know English to some extent. However, the sentence *"An algebra is a set together with operations defined in the set that obey specified laws"* doesn't make much sense to those who are totally new to algebra. It is not the new words if any that are hard to understand but rather that one lacks the knowledge to make sense of it. Yet, the sentence *"Tom is chasing Jerry"* cannot be unambiguously interpreted without referencing a certain context. Finally, while the statement *"This spatial database is of scale 1:5000"* might be used to mean "the spatial database is created by digitizing 1:5000 paper maps", and it is not a rigorous statement since the concept of scale of spatial database has no well-accepted definition. This shows domain knowledge needs to be built and shared to enable precise exchange of meaning.

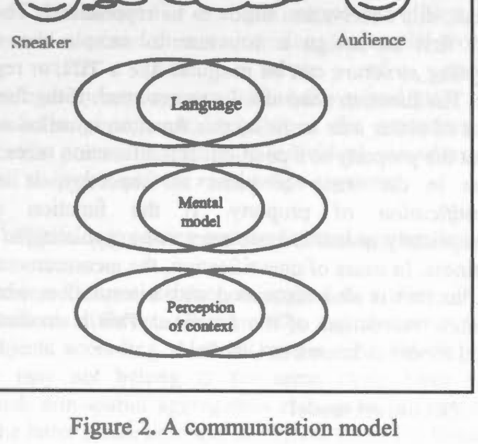


Figure 2. A communication model

To identify the sources and forms of semantic heterogeneity, we base further discussion on a communication model as shown in figure 3. In this model, the communication environment consists of a pair of speaker and audience and a conversation context. Both the speaker and the audience have their own mental models about the topic and may speak different languages. They may also have different perceptions of the context. A speaker has an opinion (the intentional meaning) and explicates it by first representing it using his language and then issuing a piece of utterance. The audience receives the utterance and tries to understand its meaning if he knows the language. First, the utterance is parsed according to the language grammar to obtain a structure of the language items. Then, the audience applies his/her vocabulary and perception of the context to interpret the utterance. During the process of interpretation, the audience may need to do some disambiguation on words, the structure of the utterance and the context. Finally, the audience applies his/her conceptualization to make sense of the interpretation to obtain the meaning. It should be noted that the audience should understanding the meaning of the speaker but does not necessarily agree with it. The disambiguation process may repeat sever times until the audience believes the "right" meaning is obtained.

Interactive information processing employing a database system is comparable to communication using a natural language. When we issue a query against the database, we get an answer, which is assumed to be semantically correct and complete. We now examine how meaning is exchanged between the user and the computer system and how semantic heterogeneity is avoided in the human-computer interaction, in particular human-database interaction. In this interaction, the common language used is the database language, which is free of grammatical ambiguity. Users need to understand the database language. Documents have to be provided to users for them to operate the views/tables in a meaningful way. All terms are defined including table names and filed names, i.e. the vocabulary is fixed. The database system can understand the programs and process data according to its own language and its own "mental model", i.e. the data schema. To the computer, data has only formal meaning in the sense of the database language and no intentional meaning in the sense of the specific application. In summary, in this interaction the user adapts the database language in the communication. The user must stick to the vocabulary and data schema, of which any change or deviation will produce errors unless the program is informed and/or updated. In fact, the computer does not care if the audience/speaker is a human user or a remote computer as in the case of homogeneous distributed database system. It should be stressed that the human operator is in charge of promoting the formal meaning of data to sensible meaning that makes sense to the application or decision-making.

	human-database interaction	Database-database communication
Speaker/ Audience	*Human operator*	*Database system*
Audience/ Speaker	*Database system*	*Database system*
Common Language	*Database language*	*A common database language*
Common Context	*That of the database system*	*Compromised context*
Common mental model	*Data schema*	*Integrated schema*
Common vocabulary	*Name of tables, fields, views etc.*	*Reconciled vocabulary*
Common Meaning level	*Formal meaning of database language*	*Formal meaning of database language*

Table 1. Communication setting of human-database interaction and a possible communication setting (in federated database system) for successful database-database communication.

Communication would fail if a common setting of the language, the context, the mental model can't be reached. Misunderstanding may occur if there is any disagreement upon the communication setting but not identified. In general, it's easy for human being to reach a common communication setting and intentionally change to another setting, given our intelligent capability while it is difficult for computer systems to reach a common communication setting. A possible communication setting for successful database interoperation is shown in table 1.

2.3 Sources of semantic heterogeneity

Based on the above discussions, we can identify sources of semantic heterogeneity of geodata as shown in figure 2, which include formalization difference, conceptualisation difference and difference in assumed/perceived context. By formalization heterogeneity, we mean the difference in formalizing a conceptualisation using data languages. It results from two factors. One factor is that we can use different data languages to formally represent a conceptualisation and these different languages may be of different semantic richness. For representing and processing geographic information, many GIS software packages with varying sophistication have been developed, of which all have their own data languages. There are also many data languages designed for exchanging geodata. The languages employ different spatial data model and language syntax. The spatial data models of some geodata languages are quite complex, providing plenty of geometric and topological data types while some are relatively simple. The other factor is that there are alternative ways to formalize a conceptualisation using one language or different languages. [Podczasy 1994] gives a comparison of some spatial data models and [Lee 1990] provides a framework for comparing syntax of different spatial data formats. For relational data model, [Kim and Seo 1991] gives a systematic analysis of representation difference.

Semantic heterogeneity
|→ *Formalization heterogeneity*
| |→ *Language heterogeneity*
| |→ *Representation difference*
|→ *Conceptualisation heterogeneity*
|→ *Context heterogeneity*

Figure 2. A classification of semantic heterogeneity

Conceptualisation heterogeneity is distinguished from formalization heterogeneity in the sense that it is the result of fundamental differences in modelling the real world. It usually means that the data is not readily suitable for a certain application. [Garcia-Solaco *etc.*1996] gave a systematic analysis of conceptualisation difference, in which O-O model is employed as the conceptual model. Regarding difference in conceptual spatial modelling, there have also been some studies. [OGC 1998] gave six cases of semantic relationship between terms defined in feature catalogues of different geodata standards. [Bishr 1998] discussed semantic heterogeneity basing on a language analogy. [Xu etc. 2000] applied the method of [Garcia-Solaco *etc.*1996] to analyse semantic heterogeneity of geodata. These studies helped us understand semantic heterogeneity of geodata. However, a common limitation of these studies is that geographic conceptual modelling is considered as an ordinary database modelling and the special problems of spatial modelling were ignored.

The third source of semantic heterogeneity is related to the context of geodata. In linguistic studies, context refers to the words around a word, phrase, statement, etc. often used to help explain (fix) the meaning. The more general meaning of context refers to the general conditions (circumstances) in which an event, action etc. occurs [Akman and Surav 1996]. For databases, [Sciore etc. 1994] used context to refer to the (implicit) assumptions made when an interoperating agent routinely represents or interprets data. When different agents try to cooperate, there often exists context heterogeneity that need to be resolved. For geodata, unit, reference system, and projection used when creating geodata are among the most significant contextual factors.

All the three kinds of semantic heterogeneity need to be tackled to achieve meaningful geodata sharing. In the rest of this article, we are particularly concerned with conceptualisation heterogeneity because it is relatively less studied and fundamental differences in conceptualisation may render geodata not sharable. We will examine conceptualisation difference in the course of geographic conceptual modelling in detail in the next section.

3. CONCEPTUALISATION HETEROGENEITY

In order to inspect conceptualization difference, we need to base further discussions on a meta model that is used for geographic modeling. In particular, we need to identify the overall idea used in geographic modeling and those modeling constructs. We base our discussion of geographic conceptual modelling on the well-accepted field-based model and feature-based model and for the latter, the Object-Oriented Model is taken as the conceptual formalism.

3.1 The field model

Space can be observed (i.e. applying a *function* to) from the viewpoint of a certain *Spatial Property*, which can be an *attribute* or *operation*. Spatial property can be natural, such as height, or artificial, such as "are there roads?" This observation happens in a certain spatial and property *resolution*. To be formal, this observation needs to be represented. Therefore, we need first to design a structure to sample the space. The *sampling structure* can be irregular like a TIN, or regular like a grid. The function need also be represented. If the function takes value of either *true* or *false*, this function is called an *assertion* about the property at a position. If the function takes continuous value in the sense of point set topology, it is called a *quantification* of property. If the function cannot be appropriately quantified, we may make it *quantified* with some fuzziness. In cases of quantification, the measurement scheme of the function is also associated with a resolution, which we call *thematic resolution* of the function. This is in fact the field-based model as known in GIS field.

3.2 The object model

Besides field-based modeling, feature-based modeling is the other essential way of spatial modeling, in which we see individual objects, i.e. geographic features. To form an identifiable object, we need not only a spatial property but also an identifying criterion. The process of identification is basically cognitive and the formal result will create an object and associate it with an identifier, which is often of sensible meaning that enables us to reference a geographic object without explicitly referencing its geometry. The criterion of identification is often complicated and involves a set of asserting rules. Individual objects are grouped into feature classes according to similarity in the criterion of identification.

As mentioned above, the Object Oriented Model is a well accepted and widely used meta model for object-based modeling, in which classification, specialization/generalization and aggregation are basic modeling methods. Here we are particularly interested in the special character of specialization and generalization when they are applied to geographic objects.

3.2.1 Spatial specialization and non-spatial specialization

After we have identified geographic features, we can attach other properties to them. Every time we introduce a new property to a class of objects, we are in fact making a further classification, i.e. specialization. For example, when adding *gender* property to *person* class, *person* is categorized into *male* and *female*. Here we make a distinction between spatial specialization and non-spatial specialization. If the criterion of specialization is spatial, it's called a spatial specialization. Otherwise, it is called an non-spatial specialization. The point is that a spatial specialization creates new spatial objects by dividing one object of superclass into component objects of subclasses while a non-spatial specialization creates new classes without creating new objects. For example, the specialization of *school* (a geographic feature class) into *primary school/secondary school/college etc.* is a non-spatial specialization. The specialization of *road* into *pedestrian path/carriageway/carriageway with walkway etc.* is a spatial specialization because this specialization creates new spatial objects, i.e. the starting and ending positions of which are generally not coincident with those of original *roads* identified using other criterion.

Therefore, the big difference between spatial and non-spatial specialization is that spatial specialization is accompanied by a process of feature identification. Further, only if the object identification of the superclass is respected, i.e. the spatial specialization is done by fragmenting objects of superclass into component objects, can we derive objects of superclass from objects of subclasses. For example, suppose we have a class called *named road* and then we introduce the spatial property of *usage* which takes value of *pedestrian path/carriageway/carriageway with walkway* and get a subclass called *road-usage-segment*, only when the secondary object identification process is applied to individual *named roads*, can we derive *named road* from *road-usage-segment*.

3.2.2 Spatial aggregation and non-spatial aggregation

The criterion for specialization is applied to a class of objects. Aggregation is different. Conceptually, aggregation is to group some objects according to some criteria. The objects grouped may or may not belong to the same class. Here we first distinguish non-spatial aggregation and spatial aggregation, of which the latter create new spatial objects while the former does not. Then, we distinguish spatial-dominant aggregation and thematic-dominant aggregation according to criterion of aggregation.

Non-spatial aggregation:

Although aggregation of geographic objects will result in new geographic objects, however, the location of the aggregates may not be of concern for a certain application. For example, a university may have several campuses in different locations. Generally, campus is a geographic object, so does a university. However, the location of the university might be not significant in some applications and in database the spatial object represent university is not explicated.

Spatial aggregation

A spatial aggregation is one that creates new spatial objects. For example, the buildings together with other accessory facilities might be aggregated and considered as a new feature, a campus, and the spatial object representing it is of significance. In this example, the criterion of aggregation is non-spatial. In fact, the aggregation is to group a set of features that functions as a whole to serve as a campus. Such an aggregation is called *thematic (spatial) aggregation* here. Some aggregations are based on spatial distribution pattern. For example, the buildings enclosed by major streets can be considered as a street block and an area with a high density of building and roads can be considered as a so-called built-up area. Such aggregations are called *pattern-based (spatial) aggregation.*

The importance of distinguishing pattern-based aggregation and thematic aggregation is that in the case of pattern-based aggregation, the processing can in theory be automated by developing spatial pattern recognition techniques while a thematic aggregation can not be automated because the criteria for aggregation is not derivable from spatial pattern. However, there is no distinct separation of thematic and pattern-based aggregation. Some thematic aggregation can be considered as pattern-based aggregation to some extent because features are related to nearby features more than faraway features. [Devogele *etc.* 1998] gave an example of spatial aggregation. They discussed the problem from the perspective of data integration and suggested to use location as an identifier to match one node to a set of *waysections*. Their basic idea is that the correspondence between the traffic circle node and a set of *waysection* arcs can be found by first forming a proper buffer of the node and then find out those *waysection* arcs falling into the buffer (see figure 3). However, it should be noted that such matching is not applicable if the case is spatial data generalization rather than data integration, in which we don't have a start point to do matching. In theory, we can apply pattern recognition technique to detect a traffic circle and do the generalization. But in practice, it is rather difficult and requires much cognitive intelligence.

Both spatial specialization / generalization and aggregation/segregation may be done recursively and thus form classification hierarchy and aggregation hierarchy. What we emphasize here is that the modelling constructs of specialization and aggregation are special to geographic features. Spatial specialization requires an object identification procedure and it is preferable that sub-identification process is applied to features of

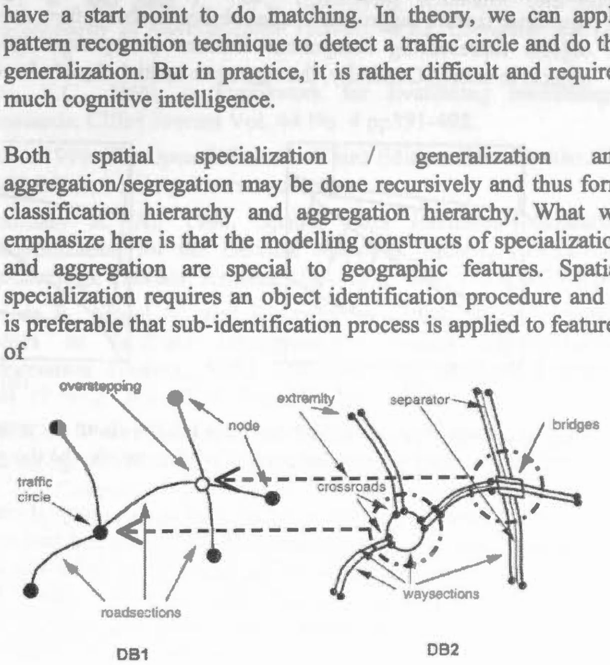


Figure 3 Thematic Spatial Aggregation (adapted from [Devogele *etc.* 1998])

superclass in order to allow the formation of a spatial classification hierarchy and allow spatial generalization be computable. In order that pattern-based spatial aggregation can be automated it is preferable that the ISPARTOF relation between component features and aggregate features be explicated. For thematic spatial aggregation, this explicitness is a necessity for automating the aggregation process.

c) Spatial relationship and non-spatial relationship
In object-based modelling, relationship between objects is an important part of conceptualisation. GIS tool is well known for its capability of computing spatial relationship between features. However, attention should be paid to some non-spatial relationships that look like spatial relationships. For example, the traffic connectivity of road segments often deviates from physical connectivity of road segments. Also the *building-along-road* relationship is not fully computable only from position of buildings and roads. In most cases the *Road* to which a *Building* logically "belongs", will be the closest. However, this will not always be the case and one *Building* may belong to two or more different *Road Elements* if the *Building* has several entrances located at different *Road* [CEN TC 278, 1995]. Therefore, whether or not a relationship is a computable spatial relationship largely depends on whether additional information is available.

3.2.3 Representation dimension and semantic constraints

Although geographic features always occupy some portion of a 3-dimensional space, conceptually they are considered as point, line or area features in most cases to achieve computational efficiency. Different applications may represent the same class of features differently. Although it is possible to geometrically derive a lower dimensional representation from a high dimensional representation, result of such geometric generalization may not make sense for a certain application.

We have discussed the criterion of spatial aggregation above and we now discuss the derivation of spatial objects representing the aggregate feature (aggregate spatial object, in short) from spatial objects representing component features (component spatial objects in short). Generally, the aggregate spatial object can only be derived if the component objects cover the whole aggregate object. The above specialization of *road and road-usage-element* example is such a case. The derivation from lower order administrative area into higher order of administrative area is another example. However, in many cases, the component objects do not cover whole area of the aggregate object. For example, the represented road elements constituting a traffic circle do not cover the area of the traffic circle and the represented buildings and facilities do not cover the area of a building complex, such as a school. In these non-covering cases, there exists the problem of how to derive a suitable aggregate spatial object from its components.

What's more, features of the same class or features of different classes are often assumed to conform to some semantic constraints, especially topological constraints. Many spatial analyses rely on such assumptions. Semantic constraints are application dependent. For example, in order to analyse the accessibility of road-building, the application may require that buildings accessible from a certain road be represented as points lie on the road line or polygon have part of road lines as its boundary to facilitate evaluating accessibility by examining topological relation between geometric objects representing them. Some semantic constraint is somewhat betraying the real world facts such as the one mentioned above. While some semantic constraint respects real world facts. For example, the constraint that road lines should intersect or meet if the road they stand for intersect or meet. In such cases, often a spatial data model support topology is preferred. It should be stressed that even though the spatial semantic constraint respects real world fact and such constraint might be maintained by applying some spatial data model supporting such constraint, it can be generated and maintained only when the geometric objects representing them have enough high accuracy. Otherwise, the computer can't generate false spatial relationship due to poor accuracy. Such an observation although is widely known but worth stressing in cases in which the geometric objects are obtained by such kind of map generalization process, especially when it is an automated processing. It is often that we are able to get a generalized representation but it is hard for us evaluate its accuracy. Figure 4 illustrates the problem.

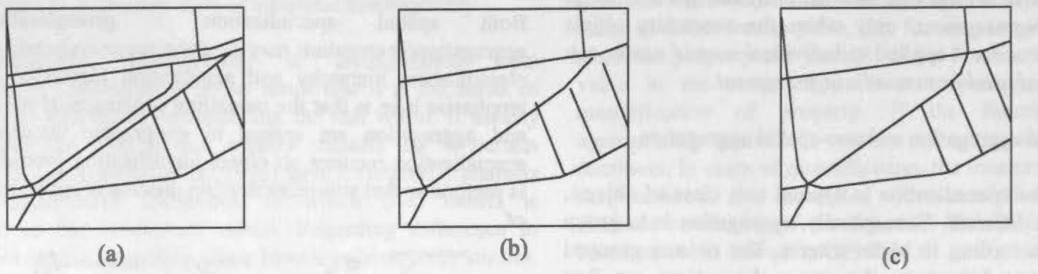


Figure 4. Topology can not be correctly built without accurate estimation of accuracy of representational generalization.
(a) the source representation of road network, (b) the generalized represented of (a), (c) the desired representation

when we have no accurate estimation of the accuracy of map generalization, we can't use topological cleaning and building process to obtain and maintain the connectivity constraint on road lines. If it is case, connectivity information should be incorporated from the original data if available, manually or automatically. In the example given above, the semantic constraint can be derived. However, it is not always the case because the representation dimension may change or the semantic constraint has changed.

4. RESOLVABILITY

The problem of geodata sharing can be considered as that of deriving an intended representation of real world from an available representation. For data created using field-based modelling, the problem of conversion among various sampling structures and deriving lower resolution representations from that of higher resolution has been widely studied. Generally, the conversion and derivation are possible but sometimes it is quite difficult and different methods are needed to deal with fields of different properties to produce useful information. We would not elaborate on this problem further. In fact, the less studied problems are those in feature-based modeling.

When sharing data created using feature-based modelling, the derivation of intended data from source data is in fact a problem of database transformation, which can generally be resolved by a series of (possibly complex and subtle) view definitions [Colomb 1996] or schema transformation [Xu and Lee 2000].[Xu etc. 2000] and [Devogele etc.1998] proposed candidate schema mapping languages for geodata translation and geodata integration respectively. For detailed discussion of this approach, please refer to foregoing literatures. However, the prerequisite of schema transformation approach is that the information needed in the target database is contained in the source database. Based on the discussion of featured-based modelling, it can be observed that semantically poor geodata set render the data not sharable. The sources of poor geodata set include:

- Non-general definition of feature classes
- Few thematic properties available
- Low thematic resolution
- Poor strategy of spatial object identification in spatial specialization
- Implicit object identification criterion in thematic aggregation
- Few relationships between features available

In fact, we have some examples of inappropriately identified road boundaries, semantically poor database of buildings and poorly modelled road network database, which can not be shown here due to limited space.

When some information is absent, such as the identifying property for thematic aggregation, thematic information need to be computed from spatial pattern of the features and human interpretation is often a necessity. Further, the computed or interpreted information need be confirmed and checked by field survey. When all necessary information for doing spatial database transformation are present, specialized techniques are still needed to deal with spatial objects derivation and semantic constraint maintenance when there are reduction in abstraction level. These issues are addressed traditionally under the title of map generalization and more recently spatial database generalization.

5. CONCLUSION

Geodata sharing is highly desirable because geographic information is public domain information which is used extremely wide and very costly to collect. In this paper, we inspect the greatest barrier to geodata sharing, the semantic heterogeneity. Our emphasis is the special characters of spatial conceptual modelling. We have shown that spatial specialization and aggregation create new spatial objects, difference in representation dimension needs representational generalization of spatial objects while at the same time relationship between generalized features need to be derived from component features. It is found that semantic poorness is the major source of semantic heterogeneity, which includes narrowly defined feature classes when creating data, few feature properties and poor thematic resolution. Another significant problem is geodata is poorly modelled. This includes using unmeaning identifiers for spatial objects, ignoring the specialization and aggregation hierarchy of feature classes. This can bring great trouble for spatial database generalization, which heavily rely on information that identify ISA and ISPARTOF relationships. We believe the cause of semantic poorness and poor conceptual modelling is that when collecting geodata we have borne in mind only traditional application of geographic information, i.e. map production, and largely ignored to take into account of GIS-based applications. It should also be stressed that spatial database generalization is a critical technique in geodata sharing as abstraction level is a very common semantic heterogeneity when geodata is used for unintended purposes.

ACKNOWLEDGEMENTS Zhu Xu thanks The Hong Kong Polytechnic University for funding his Ph.D. study.

REFERENCES

Akman V. and Surav M., 1996, Steps Toward Formalizing Context, AI Magazine 17(3):55-72.

Bishr, Y.A., 1998, *Overcoming the semantic and other barriers to GIS interoperability*, Int. J. Geographic Information Science, Vol. 12, No.4, pp299-314.

Burrough P.A. and McDonnell R.A., 1998, Principles of Geographic Information Systems, Oxford University Press, 333p.

CEN TC 278, 1995, Geographic Data Files, Version: 3.0 FIRST DRAFT.

Colomb, R. M., 1997, Impact of Semantic Heterogeneity on Federating Databases, THE COMPUTER JOURNAL, Vol. 40, No. 5, p235-244

Devogele T., Parent C. and Spaccapietra S., 1998, On spatial database integration, , IJGIS vol.12, no.4, p335-352.

Garcia-Solaco, M, Saltor, F. and Castellanos, M., 1996, *Semantic heterogeneity in multidatabase systems*, In Object-oriented Multidatabase System---A Solution for Advanced Applications (Bukhres, O.A. and Elmagarmid, A.K., Eds.), Prentice-Hall Inc., pp129-202

Jeffries L., 1998, Meaning in English: an introduction to language study, MACMILLAN PRESS LTD, 267p.

Kim, W. and Seo, J., 1991, *Classifying schematic and data heterogeneity in multidatabase systems*, IEEE Computer, 24:12, pp. 12-18.

Lee, Y.C., 1990, A Framework for Evaluating Interchange Standards, CISM Journal Vol. 44 No. 4 pp391-402.

OGC, 1998, *The OpenGIS Guide* (Third Edition, Eds. Buehler, K. and McKee, L.), http://www.ogis.org

Podczasy K. A., 1994, Spatial Data Exchange Standards Harmonization for the Defense Mapping Agency, GIS/LIS'94 Proceedings, Phoenix, Arizona, USA, 25-27 Oct. 1994, p629-638.

Sciore E., Siegel M. and Rosenthal A., 1994, Using semantic values to facilitate interoperability among heterogeneous information systems, ACM Transaction on Database Systems, Vol. 19 No.2 , June 1994, Page 254-290

Sheth, A. P. and Larsen, J., 1990, Federated database systems for managing distributed, heterogeneous and autonomous databases. *ACM Comput. Surveys*, **22**, 183.236 (special issue on heterogeneous databases).

Xu Z., Lee Y.C. and Chen Y.Q., 2000, Schema transformation for semantic geodata translation, International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B6. Amsterdam 2000, p195-202

WG IV / 3:

DATA GENERALIZATION AND DATA MINING

GENERALIZATION OF 3D BUILDING DATA BASED ON SCALE-SPACES

Andrea Forberg and Helmut Mayer

Institute for Photogrammetry and Cartography, Bundeswehr University Munich, 85577 Neubiberg, Germany
andrea.forberg@unibw-muenchen.de

Commission IV, WG IV/3

KEY WORDS: Three-dimensional Generalization, Simplification, Building Model, Automation, Visualization, Feature Analysis

ABSTRACT:

For an interactive visualization of three dimensional (3D) city models a Level of Detail representation is extremely useful to avoid unnecessary computations. Differently detailed models of the same object are displayed depending on the distance of the object. This implies that less detailed models have to be generated from complex models (generalization). In this paper an approach for the automation of 3D generalization especially for building data is presented. It uses the formally well-defined scale-space theory which includes morphology and curvature space. Practical investigations for the scale-space operations making use of the geometry kernel ACIS 3D Geometric Modeler show good results for morphology. For curvature space important preparations concerning the distinction between convex and concave object parts have been carried out. Tasks still to be performed include the coherent modeling and the implementation of continuous curvature space, the combination of morphology and curvature space, and the rectification of angles.

1 INTRODUCTION

The performance of the interactive visualization of three dimensional (3D) polyhedral data depends on the number of polygons that have to be rendered. To improve the speed, often the Level of Detail (LOD) concept is used where objects that are far away are represented with less detail than close ones (cf. Fig.1).

Many approaches for automatic polygon-reduction exist in computer graphics and computational geometry. A good survey on approaches for surface simplification is given in (Heckbert and Garland 1997). Further examples for automatic LOD generation can be found, e.g., in (Varshney et al. 1995) and (Schmalstieg 1996). An approach close to our approach dealing with computer aided design (CAD) objects is presented in (Ribelles et al. 2001).

Unfortunately, most of these approaches do not consider the specific properties of an object type such as right angles of urban objects. Opposed to this, model generalization of cartography takes into account these properties, but there is hardly any work on 3D building data. Basics about automatic generalization can be found in (Mackaness et al. 1997; Meng 1997; Staufenbiel 1973, and Weibel and Jones 1998).

In this paper we base the automatic creation of less detailed models, i.e., model-generalization, on the formally well-defined scale-space theory. The basic idea is a multiscale representation in which for a coarser scale, fine scale information with high frequency is eliminated.

On one hand, we modify the scale-spaces "morphology" and "curvature space" to deal with polyhedral, mostly rectangular building data. On the other hand, we extend the generalization of building data to 3D.

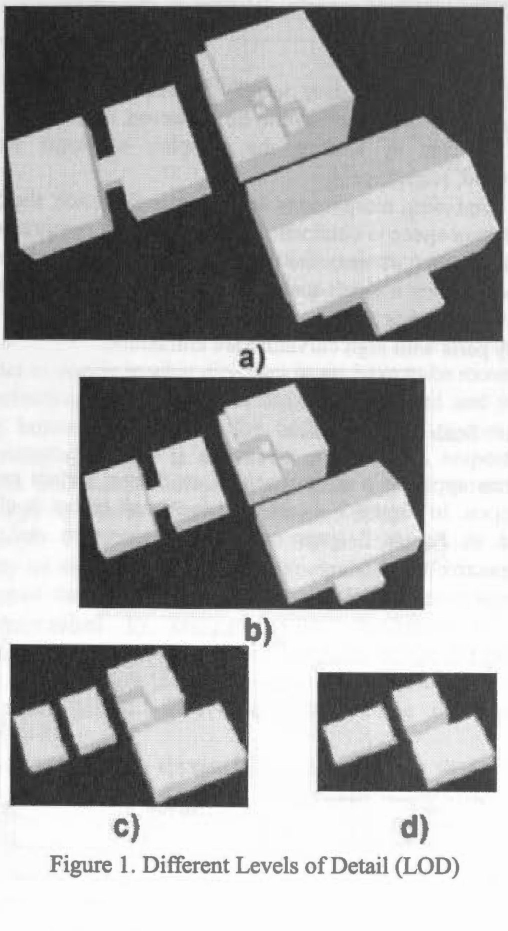


Figure 1. Different Levels of Detail (LOD)

2 SCALE-SPACES AND SCALE-SPACE EVENTS

2.1 Scale-Spaces

One basic scale-space is the linear scale-space, where causality, isotropy, and homogeneity are combined. This scale-space family continuously smoothes the image function and satisfies the so called "diffusion equation" for which the convolution with the Gaussian Kernel $\mathbf{g:R^2\times R_+\backslash\{0\}\rightarrow R}$ is the solution for an infinite domain (Koenderink 1984):

$$g_\sigma(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

Another scale-space is mathematical morphology (Serra 1982). It is based on the two basic operations erosion and dilation and the two combined operations opening and closing, here defined for grey-scale images:

$$\begin{aligned}
&\mathit{Erosion} \quad (f\ominus g)(\vec{x}) = \inf{}_{\vec{x}\in G}(f(\vec{x}+\vec{x}')-g(\vec{x}))\\
&\mathit{Dilation} \quad (f\oplus g)(\vec{x}) = \sup{}_{\vec{x}\in G}(f(\vec{x}-\vec{x}')+g(\vec{x}))\\
&\mathit{Opening} \quad (f\circ g)(\vec{x}) = ((f\ominus g)\oplus g)(\vec{x})\\
&\mathit{Closing} \quad (f\bullet g)(\vec{x}) = ((f\oplus g)\ominus g)(\vec{x})
\end{aligned} \qquad (2)$$

Because building data usually consists mostly of straight segments, which also have to be preserved, erosion and dilation are realized by shifting the complete segments inward or outward, respectively.

By combining morphology and linear scale-space, the reaction-diffusion-space is obtained. The reaction part comprises erosion and dilation, whereas the diffusion part, also termed curvature space, is for a small-scale parameter equivalent to the linear scale space. For a large-scale parameter it diverges in a way that only parts with high curvature are eliminated.

2.2 Scale-Space Events

When applying a scale-space to an image, certain events can happen. In Figure 2 a part with too small extent is eliminated and in Figure 3 gaps are filled by erosion and dilation, respectively.

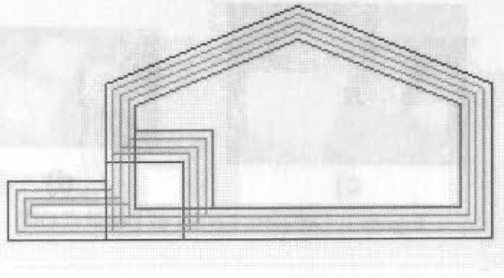


Figure 2. Erosion: blue – original object; green – incremental steps; red – object after erosion; annex is eliminated.

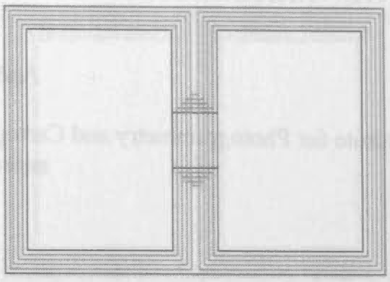


Figure 3. Dilation: blue – original object; green – incremental steps; red – object after dilation; gap between two building parts is filled.

There are two things happening in parallel while transforming the objects from fine scale to coarse scale. Firstly, the information is reduced by means of scale-space events. Secondly, the elimination of object parts such as an annex (cf. Fig.2) or the closing of gaps (cf. Fig.3) often results in a simplification or abstraction of an object. This abstraction capability makes the scale-spaces well suited for generalization.

2.3 Opening and Closing in 3D - Internal as well as External Events

Dilation and erosion are in 3D synonymous with moving all facets of the polyhedral building data in the direction of the normals. Instead of erosion and dilation opening and closing are used to reset the object to its original range of size. For small objects such as an annex inward moving facets collide while opening the object and hereby the annex is eliminated (cf. Fig.4). This is called an "internal event", because it only affects topologically neighboring facets. By employing the scale-space in small steps, i.e., incrementally, events can be handled by simple basic operations.

So-called "external events" emerge when topologically non-neighboring facets of one or more buildings touch or overlap while opening and closing. For instance, a building consists of two big blocks connected by a narrow block in fine scale (cf. Fig.5). Two buildings are created after eliminating the narrow block while opening. On yet a coarser scale, the two big blocks are merged while closing, resulting again in one building (cf. Fig.6).

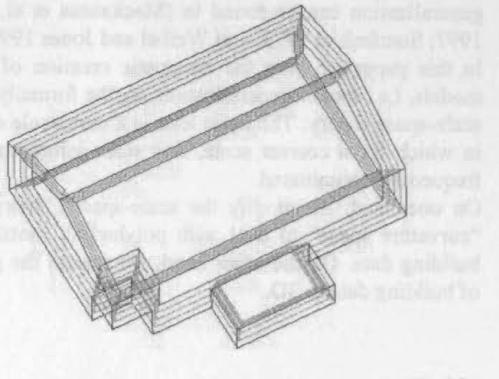


Figure 4. Elimination of the annex while erosion in 3D: blue – original object; green – incremental steps; red – object after erosion.

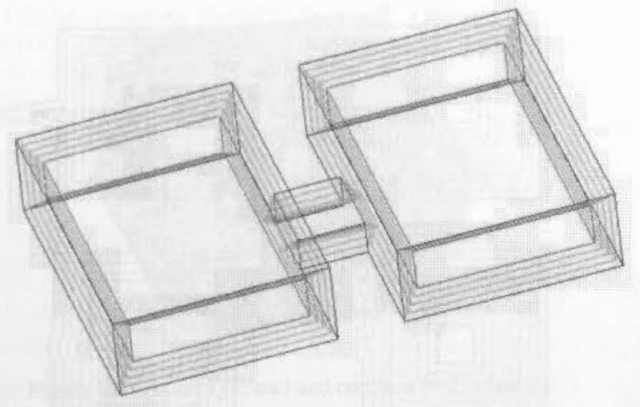


Figure 5. Splitting of two object parts while erosion in 3D: blue – original object; green – incremental steps; red – object after erosion

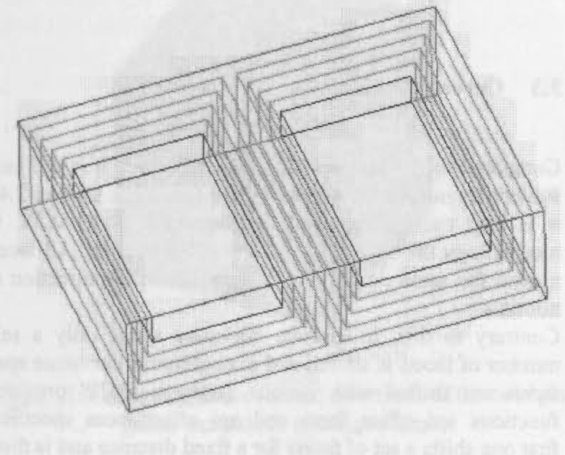


Figure 6. Melting of two objects while dilation in 3D: blue – original object; green – incremental steps; red – object after dilation

2.4 Discrete and Continuous Curvature Space

While morphology provides the only means to aggregate or split objects, it cannot be used to eliminate step-structures and inward or outward pointing boxes (cf. Fig.1a) upper centre). This is the strong point of curvature space.

In curvature space the facets are moved in a way that steps and boxes are eliminated. This is shown in two dimensions (2D) in Figure 7 for Z- and L-structures. Curvature space in 3D can be treated in a discrete or a continuous way.

In discrete curvature space only specific facets under a certain size are shifted. E.g., for Z-shapes the two long segments are moved in opposite directions in a way that the short segment becomes even shorter. For more than one short segment all segments are shifted inward or outward at the same time (cf. Fig.7). In 3D the procedure is similar. Only the facets belonging to the step or the box which has to be eliminated are shifted inward or outward.

In contrast to this, in continuous curvature space all facets are moved, but the speed of the movement is weighted by the area of the facets and by the length of the corresponding edges. The different properties of discrete and continuous curvature space are shown in Table 1 (Mayer 1998).

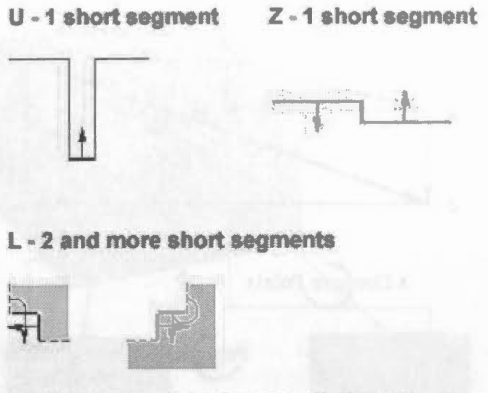


Figure 7. Movement of the facets to eliminate U-, Z-, and L-structures

	Discrete Curvature Space	Continuous Curvature Space
Generality	Depends on threshold	General
Movement	Convex facets are fixed	All facets move with different speeds
Weighting	Area	Length of segments (concave only) + area; (e.g.: length^2 * area)

Table 1. Basic properties of Discrete and Continuous Curvature Space

In order to decide in what direction facets have to be moved, the determination of elements like steps and outward and inward going boxes is necessary for both curvature spaces. The differentiating feature is convexity or concavity, respectively. For the detection of concave segments (step-structures and inward going boxes) different approaches exist:

In (Mayer 1998) the relation of the normals which depend directly on the 2D polygons of the neighboring facets are used to control the movement of the facets. The concave segments are determined by comparing the orientation of three subsequent points of the ordered point-list of the 2D polygon with the orientation of the whole polygon. When the orientations are different, the points define two concave segments (cf. Fig.8).

The problem of this approach in 3D lies in the decision if a structure goes inward or outward. Especially with nested structures as seen in Figure 10 this becomes an intricate problem.

Therefore in this paper a new approach to determine concave segments which can deal with this problem is proposed. It investigates the interior angles between neighboring facets and determines, if points created by extending segments are in- or outside of the polyhedron. The approach is described in detail in section 3.4.

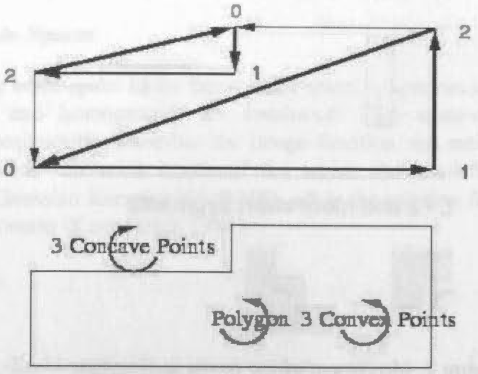


Figure 8. Concave segments have a different orientation from that of the polygon.

3 PRACTICAL INVESTIGATIONS

3.1 ACIS 3D Geometric Modeller and VRML

The prove of concept was done in (Mayer 1998) based on CGAL (Computational Geometry Algorithm Library, www.cgal.org) with a fixed sequence of opening, closing, and curvature space. This led to acceptable results, but great effort was needed to change the topology of the polyhedra after the scale-space events.

Therefore, it was decided to use a more suitable tool, particularly the ACIS 3D Geometric Modeler (Spatial Corp., www.spatial.com) that offers a more general solution for 3D topology. Functions that can directly be used for the erosion and dilation as well as for the change of the angles between single facets of the object already exist. The latter function can be used to handle non-orthogonal structures. Short descriptions of the functions mainly used are given in the following sections.

As input format VRML (Virtual Reality Modeling Language) as a common file format for interactive 3D models is used. It offers the possibility to build test objects in a fast and simple way and many programs provide the possibility to export their own formats to VRML. To construct the ACIS objects out of the VRML models a converter was implemented.

Until now in ACIS only separate tests for morphology and discrete curvature space have been made. The investigations concerning the ACIS functionalities are not yet finished, but the main tasks are already functional.

3.2 Morphology

For morphology all facets need to be moved simultaneously inward or outward. In ACIS this is termed offsetting. The function api_offset_body works for all kinds of objects, but for the task of generalization it is only well suitable for orthogonal structures. Inclined structures disappear far too slow, especially when the inclination is low (cf. Fig.9 lower right).

For perpendicular structures, small segments are eliminated by incrementally offsetting the whole object until ACIS reports a change in topology.

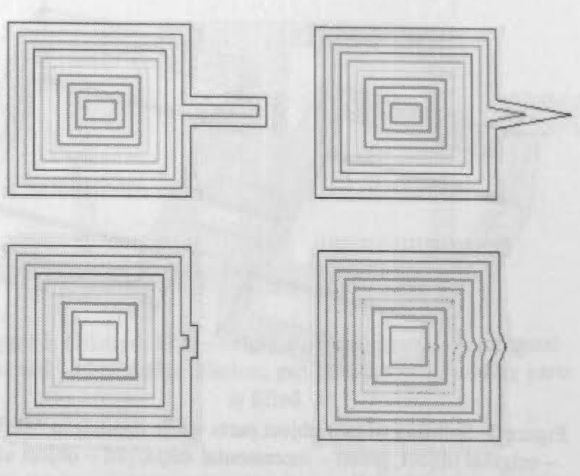


Figure 9. Erosion, seen from above: inclined structures especially with a low inclination (cf. lower right), disappear slow compared to orthogonal structures.

3.3 Curvature Space

Compared to morphology for curvature space it is not only yet sometimes unclear how to move the facets (cf. section 2.4), but it is also more complex to implement it with ACIS. Using morphology the whole object is handled at once. All facets are moved the same distance in or opposite to the direction of the normals.

Contrary to this, in discrete curvature space only a selected number of facets is shifted and in continuous curvature space all facets are shifted with various distances. ACIS provides the functions api_offset_faces and api_offset_faces_specific. The first one shifts a set of facets for a fixed distance and is therefore suited best for discrete curvature space. The latter moves each facet individually and can therefore be used for continuous curvature space.

Until now mainly the discrete curvature space operations were tested in ACIS. The first task is to decide which facets have to be moved. For this, and in order to obtain the sign for the movement of a particular facet, one has to know, if a vertex is concave or convex.

Yet, the determination of the convexity and concavity of vertices is only the basic problem. The concave and convex parts can construct rather complicated nested structures. The remainder of this section deals with the determination of the hierarchy of the concave and the convex structures.

3.4 Concave and Convex Parts

First attempts to determine the convexity of a vertex were made by computing the interior angles between the facets which are neighbors to the vertex. If all interior angles at a vertex are 90°, the vertex describes a perpendicular corner which is convex, if they are 270°, it is concave (cf. Fig.10).

A more general solution is made by extending the coedges belonging to, i.e., ending at a vertex (cf. Fig.11 and 12). An edge according to ACIS is adjacent to two facets which results in two coedges with different directions, each associated with a loop in one of the two facets

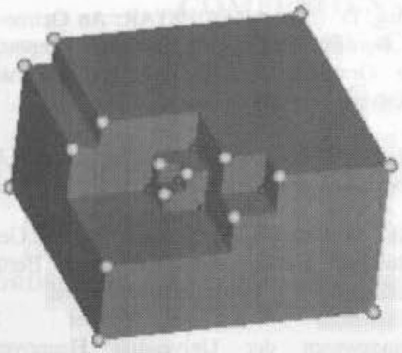


Figure 10. Convex (yellow) and concave (red) vertices

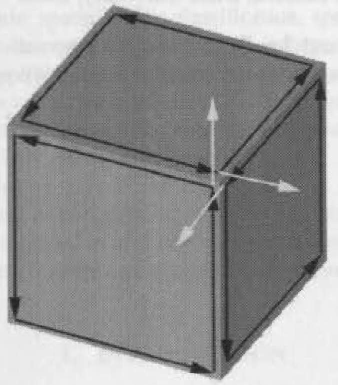


Figure 11. Coedge structure (black: directions of the coedges; green: extensions of the coedges)

The coedges are only very slightly extended (linear equation) and the new endpoint is tested for its relation to the object itself by the ACIS function api_point_in_body. Possible relations are inside, outside, or on the boundary of the object. If all extended points belonging to a vertex lie inside the object, the vertex describes a concave corner, if they are all outside of the object, it is convex (cf. Fig.12).

Unfortunately, the determination of concave and convex vertices is not sufficient for the decision which facets have to be moved outward or inward, respectively. If an object consists of nested structures as seen in Figure 10, additionally different kinds of concave structures have to be distinguished.

The five concave vertices in Figure 10 (represented as red spheres) belong to different types of box structures. To determine if a box points inward or outward, the facets which are neighbors of the obtained concave vertices are investigated. If all related facets do not contain any other concave vertex, this is a fully concave, inward pointing box without any other object part contained (cf. Fig.13, red facets). Else, it is part of a nested structure (magenta and yellow facets).

Next the concave structures which are higher in the hierarchy (magenta) have to be distinguished from the convex structures they include (yellow). The number of concave vertices for each of the facets is known. If a facet contains only one concave vertex, it is convex (yellow). If there are two of them, the facet is concave (magenta).

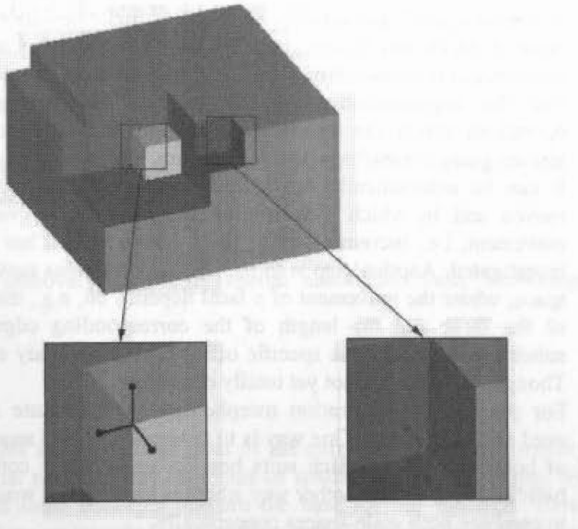


Figure 12. Coedge-extension to distinguish convex and concave vertices: If all ends of the extended edges at a vertex lie outside, it is a convex vertex (left), if they lie inside, it is a concave vertex (right)

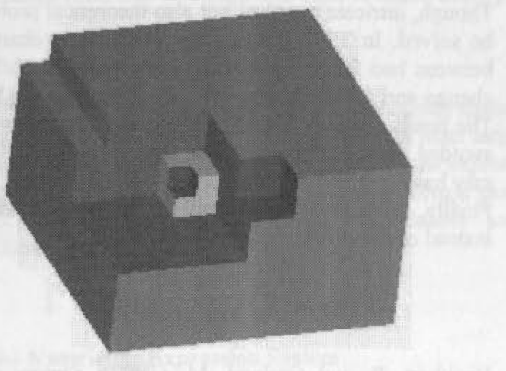


Figure 13. Determination of different types of outward or inward going boxes

After the determination of the three types of facets, the sign of the movement is clear. While the concave facets (red and magenta) have to be moved outward, the convex facets (yellow) have to be moved inward. In Figure 12, e.g., we have eliminated the smallest of the inward pointing boxes in the centre of Figure 13.

What is left to determine is the distance for the movement. At the moment it is, e.g., investigated if an incremental offsetting of the specific facets is suited best, or if the offset distance has to be fixed by the length of the shortest edge contained in a set of facets. The latter is due to the fact that the ACIS functions only produce a cleaned topology if facets are coplanar.

4 CONCLUSIONS AND OUTLOOK

An approach for the generalization of orthogonal building structures using the scale spaces "morphology" and "discrete curvature space" was introduced. The test implementation was done in ACIS and shows good results for morphology where annexes can be removed or objects can be split or aggregated.

For the implementation of the discrete curvature space operations which can handle step-structures and outward and inward going boxes important preparatory work has been done. It can be automatically determined which facets have to be moved and in which direction of the normal. The way of movement, i.e., incrementally or fixed distances, still has to be investigated. Another step is to realize the continuous curvature space, where the movement of a facet depends on, e.g., the area of the facet and the length of the corresponding edges. A suitable function for the specific offset in ACIS already exists. Though, the theory is not yet totally consistent

For practical generalization morphology and curvature space need to be combined. One way is to determine a fixed sequence of both operations, which suits best for most of the common building structures. Another way which is investigated would be to combine both scale-spaces conceptually.

As the work at the moment is restricted to orthogonal models, additionally rectification is necessary. Contrary to the test models, real data in many instances will not be orthogonal. In ACIS exists the function api_edge_taper_faces for changing the angles between the facets. An algorithm for automatically detecting and correcting slightly distorted angles has to be developed. The code to derive the interior angles between facets mentioned in section 3.4 could be used as starting point. Though, intricate practical but also theoretical problems have to be solved. In 3D it is not sufficient to only change the angle between two facets because all other angles are affected by the change and are also changed even if they should be preserved. The result could be a much more skewed model which must be avoided. In the future also the main directions of the building may have to be taken into account.

Finally, the scale-space operations are to be tested on real data instead of the simulated test objects.

REFERENCES

Heckbert, P. and Garland, M. (1997): Survey of Polygonal Surface Simplification Algorithms. Multiresolution Surface Modeling Course SIGGRAPH '97

Koenderink, J. (1984): The Structure of Images. Biological Cybernetics **50**, pp. 363-370.

Mackaness, W., Weibel, R., and Buttenfield, B., (1997): Report of the 1997 ICA Workshop on Map Generalization. Publication of the ICA Working Group on Map Generalization, http://www.geo.unizh.ch/ICA/, Gävle, Sweden.

Mayer, H. (1998): Three Dimensional Generalization of Buildings Based on Scale-Spaces. Technical Report, Chair for Photogrammetry and Remote Sensing, Technische Universität München, Germany.

Meng, L., (1997): Automatic Generalization of Geographic Data. Technical Report, VBB Viak, Stockholm, Sweden.

Ribelles, J., Heckbert, P., Garland, M., Stahovich, T., and Srivastava, V. (2001): Finding and Removing Features from Polyhedra. Proceedings of DETC'01, ASME Design Engineering Technical Conferences, Pittsburgh, Pennsylvania, USA.

Schmalstieg, D. (1996): LODESTAR: An Octree-Based Level of Detail Generator for VRML, Technical Report, Institute of Computer Graphics, Visualization and Animation Group, Vienna University of Technology.

Serra, J. (1982): *Image Analysis and Mathematical Morphology*. Academic Press, London, Great Britain.

Staufenbiel, W. (1973): Zur Automation der Generalisierung topographischer Karten mit besonderer Berücksichtigung großmaßstäblicher Gebäudedarstellungen. Vol. 51, Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Hanover, Germany.

Varshney, A., Agarval, P. Brooks, F., Wright, W., and Weber, H. (1995): Generating Levels of Detail for Large-Scale Polygonal Models, Technical Report, CS-1995-20, Department of Computer Science, Duke University, USA.

Weibel, R. and Jones, C. (1998): Computational Perspectives on Map Generalization. GeoInformatica **2**(4): pp. 307-314.

COMPREHENSIVE KNOWLEDGE DISCOVERY: THEORY, CONCEPT AND APPLICATION

Sha Zongyao [a], *, Bian Fuling [a]

[a] School of Remote Sensing and Information Engineering, Wuhan University, 430079, WuHan, China - (zongyaosha, Fuling Bian)@263.net

KEY WORDS: comprehensive knowledge discovery, knowledge discovery algorithm, spatial association rule, knowledge expression system, data mining

ABSTRACT:

The principle of comprehensive knowledge discovery is proposed in this article. Unlike most of the current knowledge discovery methods, comprehensive knowledge discovery considers both the spatial relations and attributes of spatial entities or objects. We first introduce the theory if spatial knowledge expression system and some concepts that are the base for our research. Those concepts include: spatial object classification, spatial relations, comprehensive knowledge discovery, comprehensive knowledge and SUIT, etc.. In theory, SUIT records all information contained in the study objects, but in reality because of the complexity and varieties of spatial relations, only those factors that have interest to us are selected. The selected factors constitute the to-be processed data which are a subset of SUIT. In this study, we select spatial association relation as the research emphasis. In order to find out the comprehensive knowledge from spatial databases, an efficient comprehensive knowledge discovery algorithm called recycled algorithm (RAR) is suggested.

As an example, we give a case study based on our proposed algorithm to get comprehensive knowledge. The research areas are agricultural land in two adjacent counties in northern China. The climate of northern China is very in lack of rain and only those crops that accustomed to arid environment can survive. By RAR, spatial association rules can be mined and used to make decision on crop planting distribution in agricultural planning.

1. INTRODUCTION

Knowledge discovery in databases (KDD) and data mining (DM) have been an area of increasing interests during recent years. Because data mining can extract desirable knowledge or interesting patterns from existing databases and ease the development bottleneck in building expert systems so they have become common interest to researchers in machine learning, pattern recognition, statistics, artificial intelligence, and high performance computing (Leung Yee, etc., 2001; W. Lu, etc., 1993; KERRY TAYLOR, 1999). According to knowledge types, it can be grouped into spatial information generalization, spatial association, spatial classification and spatial clustering. In data mining research area, there exist two trends: mining attribute data mining omitting spatial character and mining spatial relation omitting attribute associated with spatial objects. Both of those trends are not complete in expressing the real world. Purely attribute data miner ignores the fact that 80% of the earth data surrounding us has spatial character while purely spatial data miner usually emphases less on the importance of spatial object attribute which are deep and full description of an object. Both experience and theory study prove that knowledge discovery from spatial databases (KDSD) cannot be independent of spatial objects attributes and should be associated together to give a complete expression of spatial objects. The association rule discovery problem in particular has been widely studied and has been the focus of many studies in the last few years and spatial association rule mining has become one important aspect of KDSD (S Rahayana and A Siberschatz, 1998). General transaction association rule mining cannot explore the implicit spatial rule in database, so method that can integrate mining both spatial and attribute features is urgently needed. Since the spatial relations are complex and are not easy to express, it is very important to construct a suitable spatial data mining model which can ease the process of KDSD.

We present in this paper an approach to integrate mining spatial relation and the attribute character of spatial objects from large spatial data repositories. We also use this approach to explore the rules that lies behind large database collected from a case study area. The rules mined proved to be valuable and understandable.

2. THEORY AND CONCEPTS

2.1 Spatial Knowledge Expression System

Let $S=(U, C, D, V, f)$, and $U=\{u_1, u_2, \cdots, u_n\}$. U is a finite set of objects, $A=C \cup U$ is attribute set , $C=\{a_1, a_2, \cdots, a_m\}$ is the condition attribute set (note should be taken that C contains spatial constraint conditions), $D=\{d_1, d_2, \cdots, d_x\}$ is the decision attribute set, V is the field set composed of $C \cup U$, viz. $V=\cup_{p \in A} V_p$, V_p is the field of attribute p, f is an information function, viz. f: $U \times A \rightarrow V$. S is defined as formalization definition of spatial knowledge expression system (or SKES).

In the view of the form of SKES, there is no difference between SKES and the general knowledge system often outlined in artificial intelligence. However, the condition attribute set of SKES includes both spatial and attribute constraints. As for spatial constraints, different spatial relation type may have different forms. For example, if we consider spatial clustering, a spatial object may be given a constraint that it must be a certain clustered group. At the same time if we research on spatial association, we should first classify the spatial objects (into n categories) and then construct an attribute set with n-dimensions. The value of each object in the n-dimensions set will be given spatial index value (fuzzy index type) or 0-1 (Boolean index type) according to the spatial association we intend to include in a research project.

Spatial knowledge data mining aims at mining some interesting patterns that are unknown before analysis. Because of the complexity of spatial database, which not only contains attribute data but also spatial relations (topology relation, metric relation, orientation relation, etc.), mining spatial knowledge imposes more challenges. Currently, the main research area of spatial knowledge data mining is basic theories, optimised algorithms and applications (Bian Fuling, etc., 2001; Ziarko W, 1995). SKES is intend to abstract irregular data from the real world that contains valuable information and simplify the disposing processes.

2.2 Some Concepts

Spatial object classification: Classification assumes that homogeneity and heterogeneity exist between objects. The standard of classification is the essential factors of objects that can be used to identify a given object. According to such standard, spatial objects then can be grouped into a several divisions. The research of spatial relations is undergone based on the frame of spatial classification. Without classification, all spatial objects are referred to be one thing and the relation does not exist. If we say object A and B have an association relation, we know A and B do not belong to the same group.

Spatial relations: Researching spatial relations is a key area in GIS theory and application, and an important function of GIS is embodied by spatial analysis (Sauchyn, DJ, Yong Xongchao, 1991; T.Q. Zeng and Q. Zhou, 2001; Zhang T, etc., 1997). The footstone of spatial analysis is to understand spatial relations. The methods for describing spatial relations include intersection-based model, interaction-based model and hybrid method based on *voronoi* graph. According to semantically relation, spatial relations can be divided into topological relation, ordinal relation and metric relation, etc.. But in reality, only one or two is selected to research on.

Comprehensive knowledge discovery: Comprehensive knowledge discovery is to analyse comprehensively on spatial character as well as attributes of spatial entities and to find out deep regulations that are stored implicitly in attributes information and spatial information of research objects. For example, in the process of analysing on influential factors upon crop yield, we usually only consider possible attribute factors such as climate, soil fertility, soil texture, etc. but ignore spatial information (climate and soil fertility distribution) that may contain spatial association patterns. Those unknown patterns can then be used to support decision for crop planting area planning and yield evaluation.

Comprehensive knowledge: This is referred to be the rules that are found out by the method of comprehensive knowledge discovery. The patterns, containing both generalized spatial and attribute information, are understandable and having potential applications.

Spatial union information table (or SUIT): The above-mentioned spatial information includes graphical information, topological information and attribute information of spatial entities. SUIT is defined as information table containing graphical information, topological information and attribute information of spatial entities. This table can be separated into two parts with spatial relations (SR) that record classification and relations of spatial entities and attribute information (AR) which records attribute fields of spatial entities. In a formalized form, SUIT is expressed as SUIT（T, SR, SRV, AI, AIV）, while T stands for the whole set of spatial entities, and SRV and AIV are index value or representative mode of spatial relations and attribute value of spatial entities. It is possible to obtain all possible valuable information (such as spatial information generalization, spatial association rules, spatial classification and clustering, etc.) by processing on SUIT. In practice, it is impossible to consider all factors. For specific purposes, some simplifications have to be made. Let SUIT$'$ as the actual study goal, and T$'$, SR , and AI' are subsets of spatial objects, spatial relation objects and attribute of spatial objects respectively, viz. $T' \subseteq T$, $SR \subseteq SR$, $AI' \subseteq AI$. As shown in Figure 1, the sub-sets of spatial objects: T$'$={A, B, C, D, A}, the classification of spatial objects is: SR ={A, B, C, D}, suppose attribute field sets AI'={area, perimeter}, the content of SUIT$'$ is shown in Table 1. The value of the elements in Table 1 shows the relative neighbourhood index between entities shown in Figure 1. In the following sections, we will use SUIT to stand for SUIT'.

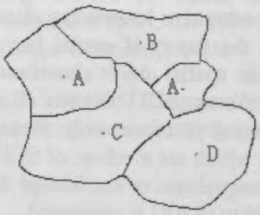


Figure 1. Diagram showing spatial object relation

T	A	B	C	D	area	peri
A	0	0.5	0.7	0	1.95	2.74
B	0.9	0	0.3	0	1.90	2.88
C	1.1	0.3	0	0.6	2.13	3.11
D	0.5	0	0.6	0	1.89	2.69
A	0	0.4	0.4	0.5	0.84	1.39

Table 1. SUIT$'$ for spatial objects

Figure 1 shows the fact of spatial adjacency relation of spatially neighbouring objects. Possible conditions other than adjacency relation can also be integrated in SUIT. The difference between spatial adjacency relation and other relations lies in the different application meaning of elements in Table 1. We are now researching on a method called influential field to qualitatively represent spatial relations by extending adjacency relation. In this method, spatial objects do not need share edge (in Figure 1 spatial objects share edge). They can be crossing, separating or adjacency. Figure 2 shows two separating spatial objects O1 and O2, and their influential fields. Each object in a given area is influenced by numerous fields caused by other objects. Those spatial objects can also be classified as A, B, etc. so for any object, its relation index value with other objects can be calculated according to the influential field model. By this way the possible relations between spatial objects can be extended.

3. OVERVIEW OF ASSOCIATION RULE AND SPATIAL ASSOCIATION RULE MINING

3.1 Association Rule and Spatial Association Rule

Association rule describes item relations in a database. In a mathematic language, let I={i_1, i_2, …, i_m}and it is an itemset called dataset, let D is a collection of all possible itemset.

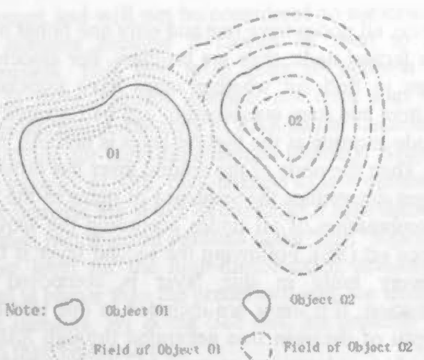


Figure 2. Influential field for two separating spatial objects

Transaction T is a sub set of I, vis. $T \sqsubset I$, and every transaction is identified by TID. For dataset X, we call T includes X if and only if $X \subseteq T$. Association rule is usually presented in the shape like: X=>Y, here $X \subseteq I$, $Y \subseteq I$, and $X \cap Y = \emptyset$. X is called the condition of the rule while Y is result of the condition X. The confidence level for the rule X=>Y on the base of set D is defined as c% which means that out of all transactions in D there are c% records include both X and Y, while the support level of rule X=>Y is defined as s% that means there are s% out of all transaction records include X□Y. Confidence reflects intensity of rule and support reflects frequency of a rule. Rules whose support level are higher than the predefined support level are called frequent rules and both confidence level and support level are higher than predefined ones are called intensive rules. So we can see association rule represents relations between objects in macro view. This is different from association relation between two items or two spatial objects that represents a certain relation between items or objects in micro view.

The focus of spatial association rule mining is on spatial information. In formalized expression, it is described as: for spatial objects A and B (A, B do not belong to the same type) and complete objects set U, let R be the spatial relation term, ARB if A and B have an association relation R. For example, if along the sides of river A in 5 kilometres arrange area 80% of all area is distributed by agricultural fields then we say agricultural fields and river have a certain association relation. There exists a large difference in association mining between spatial database and transaction database. Firstly, spatial association rule in spatial database is more difficult to be mined because spatial data structure is usually unstructured. So data miners have to be familiar with a particular data structure and take primary data preparation to translate data structure into another one that is convenient to be handled. Secondly, spatial association relation is fuzzy. Items in transaction database have either associate relation or not, often referenced as 0-1 algebras. There is no middle state. For spatial database, however, association relation is co-determined by objects that belong to one or several types. Suppose object X, grouped as type O, has relation objects X1, X2, X3 and X4, and X1, X2 belong to type A, X2 and X3 belong to type B and type C respectively. As for object X, its contribution to spatial relation between type O and type A is the sum of relation index value of X1 and of X2. A predefined standard has to be set to determine whether object X is spatially associated with X1 (and X2).

3.2 Review of Algorithms in Mining Association Rules

The general method for mining association rule include the following procedures: (1)to find out all frequent itemsets; (2)to form association rules from frequent itemsets. For the given full item set U, if $A \subseteq U$ is a sub set of U and sup(A)/sup(U)>*Confidence*, where sup(X), *Confidence* stand for support level and confidence level for spatial type X, then the rule A=>U-A can be induced. In the above two steps, the first step is central. Once frequent itemsets have been got, it will be easy to form rules.

Classic association rule mining algorithms such as Apriori and DHP (Direct Hashing and Pruning), etc. are usually used to draw rules from transaction databases. The principle of Apriori is to generate large candidate itemsets by scanning database and calculate the happening times of each candidate itemset. One-dimensioned large itemset L_1 extracted from the large candidate itemsets. Next is to generate two-dimensioned large itemset L_2 based on L_1 and the database. Following the same method, n-dimensioned large itemset L_N can be formed and the n+1 dimensioned large itemset no longer exist. Sequential large itemsets $\{L_1, L_2, \ldots, L_N\}$ can be got. Because Apriori is very time consuming in generating large itemsets Park proposed hashing pruning called DHP algorithm.

Besides the above mentioned algorithms, literature (Agrawal R, etc., 1993; Lavingto N, etc., 1999) also proposed generalized association rule, multi-level association rules, quantitative association rules mining respectively. But almost all of those algorithms need scanning database many times, which greatly reduce their efficiency.

All of those methods are directed to transaction database, however. As for spatial association rule mining, it is also possible to apply after a little modification. But the spatial database must be based on suitable data model. Even so, the efficiency is low. So in section 5, we will give an efficient algorithm (RAR) to find spatial association rules. As preparation, neighbourhood index value and the generation of SUIT are in the field of data model construction.

4. INDEX VALUE CALCULATION FOR SPATIAL NEIGHBOURHOOD RELATION

Spatial relations include many categories. To simplify our study, we take spatial neighbourhood relation as the specific subject to research its association index value.

Here we take polygon objects as example. It is usually regarded as spatially neighboured if two spatial objects share *voronoi* edge, but this definition does not give the way how to calculate the qualitative value for spatial neighbourhood relation, viz. it cannot explain Object A is more neighboured to Object O than Object B. In Section 2.2 we simply illustrate the method of influential field model to express spatial association, in the following study we use another method which can calculate spatial relation of adjacency instead of this model due to the computational complexity is heavy. Figure 3 shows Polygon 1 and Polygon 2 share common edge AB. In order to give qualitative spatial neighbourhood value, it is necessary to set a standard that can express neighbourhood index value and this value is an index for spatial association rule. Define neighbourhood index N_q for spatial objects, which do not have containing, or contained relation. N_q is positively ratio to the length of sharing edge and negatively ratio to the distance between objects canter. The central points of Polygon 1 and Polygon 2 are O1, O2, as shown in Figure2. The length of AB is l_{AB}, we will get $N_q = l_{AB} / l_{O1O2}$, where l_{O1O2} is the distance from O1 to O2. When a study object has more than one neighbourhood objects that belong to the same type, the neighbourhood index is the sum of neighbourhood objects that sharing an edge. As show in Figure 3, in the three spatial objects Polygon 1, Polygon 2, Polygon 3, if the neighbourhood objects of Polygon 2 (they are Polygon 1 and Polygon 3) are grouped as a same type A, then they can be merged into one. But the neighbourhood index value between Polygon 2 and type A is $\Sigma\ 1/l_{O2\text{-}P}$, where O2-P is the distance from canter of Polygon 2 to the canter of its neighbourhood objects. When the index value is higher than pre-set value, then the two types of spatial objects are spatially neighboured. If two objects have containing or contained relation, they are *absolutely* neighboured. From association (here specifically neighbourhood index value), it is possible to analysed spatially associated rules both qualitatively and quantitatively. We can extend this concept to spatial buffer analysis, which is useful when we want to get the neighbourhood index for a buffer area of a given object. Although this method has its limitations, it still can be used to evaluate regular objects and some irregular spatial objects as objects having enclosure relations.

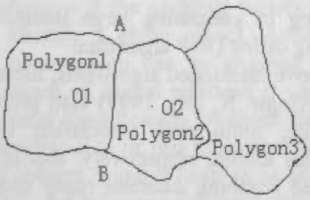


Figure 3. Association relation diagram showing adjacent neighbouring polygon objects

It is possible to get association relations other than adjacency as mentioned in Section 2.2, using the method of influential field method to calculate the association index value for separating and crossing spatial objects. Since the form of SUIT for all spatial relations are the same we will use spatial relation of adjacency (neighbourhood index value) in the following study as example.

5. SPATIAL ASSOCIATION MINING ALGORITHM

In this paper, we present a new algorithm for efficient association rule mining, which we apply in order to discover association rules in spatial databases. Our algorithm, which is called Recycled Association Rule mining (RAR), is based on the designed data structure SUIT. Other spatial relations finding are the same except a little difference of the construction of SUIT.

5.1 Description of RAR

Three steps are included in spatial-attribute comprehensive discovery: (1) to find out all large itemsets, (2) to generate rules that have confidence value higher than the predefined confidence point, and (3) to minimum rules generation. The first two steps in fact are spatial association rules mining and I/O operates on T, SR and SRV that are elements of SUIT, while step (3) is to find out comprehensive knowledge by integrating the results made from the first two steps and attribute of spatial entities and form a logically correct knowledge base or find out any logically incorrect rules from mined rules attained. In the above three steps, (1) and (3) are the key points while step (2) is relatively easy to do. Because step 3 has been introduced in our research (Bian Fuling, etc., 2001), we only give algorithm RAR for step (1).

The principle of RAR is based on a decision-tree generation method called Systematic Set Enumeration (SSE) proposed by Rymon, R (Rymon R, 1992). A decision tree is composed of tree nodes and connection line between nodes. SSE supposes all nodes have a relation of hierarchy layer structure. Except the root node, all nodes have one and only one father node. If nodes share a farther node, they are brothers. For association mining, we want to find out whether items have association relation. Every item and item combination can be regarded as node. The root node locating at the highest layer is nothing, represented as set {}. Then the node of the second layer has only one item and this layer determines the sequence of items in the decision tree. The combination of all nodes according the sequence we call sequence set (SS). Following the second layer is the third layer and every node in this layer is composed of 2 items combination. If n items are included in one transaction record, the depth of decision tree generated through SSE will be n+1 and the lowest layer has at most one node, which is composed of all items. The content of node in middle layers are made of two parts: head and tail. The head part is a item set that is completely derived from its father and the tail part is the item set of SS subtracting head item set.

The pruning strategy of decision tree can be described as following: if father node $\{N_1\}$ is not a frequent itemset then there is so use to generate its children nodes because they must not be frequent itemset. This can greatly reduce computing time.

5.2 Procedures

We use one bit segment (8 bit segment constitute one byte) to present an association flag (yes or no) and RAR to tract all possible spatial association frequent itemsets. Suppose the largest possible dimensions is *m*, the total count of record is *n*. In order to complete step 1, one time database scanning is needed with the intent to find out large itemsets and twice scanning are needed if quantitative association rules are to be mined, with the second to find out index value of association relations. The base of the implementation of RAR is on SUIT. Scanning database is meant to scan SUIT. The whole procedures is detailed in the following:

Step1: define *primary table* with two dimensions N(p1)(p2) and one-dimensioned *sum table* A(k × p3), while p1=n (total record count); $p2=MOD[(\Sigma C_m^i +7)/8]$ (i=from 1 to m); $p3=\Sigma C_m^i$(i=from 1 to m). The implication of formula p2 stands for all possible nodes count of resulted decision tree with n items processed by SSE (In the equitation, /8 is to get the possible largest bytes) and p3 stands for the largest possible nodes count of resulted decision tree with n items. k is a constant value(usually 4). Every element of *sum table* records support level of corresponding elements for items in *primary table.*

Suppose the largest possible itemsets containing 10 items, then from the above step we will get the total count of bit segments is 1023 and thus p2=128, p3=1023 (note: {Ø} is not included). In the 1023 bit segments, the first 10 is the initial storage region for data import (the data stored in this region is called input attribute) and all the other segments are for temporal data storing region (the data stored in this region is called valuation attribute). For convenient purpose, the initial value of elements in both *sum table* and *primary table* are set to 0.

Step 2: Initialise the initial storage region. Fill the elements of the initial storage region by scanning SUIT. The value of elements in the initial storage region will be filled by 1 if the content of corresponding element in SUIT is not null and by 0 on the other side.

Step 3: make summary by column after the initial storage region has been initialised. The result is to fill the corresponding elements of *sum table.* If the value of the corresponding element of *sum table* is smaller than predefined support level, this column (item) is deserted because it is not frequent itemset and will not be considered to construct higher dimension itemsets because it cannot be used to generate frequent itemsets. This step is to keep all itemsets that are impossible to form frequent itemsets out from step 4 so the whole computational complexity of RAR can be decreased.

Step 4: searching for high-dimensioned itemsets. According to the pruning strategy of SSE, two itemsets of low dimension that are frequent itemsets are select to construct a high-dimensioned itemsets. Those two selected itemsets make "and" algorithm and form support level for the high-dimensioned itemsets. If the support level is higher than the predefined one, the itemsets will be frequent itemsets. By doing like this, all possible frequent itemsets can then be found out.

Step 5: mining quantitative spatial relations. From step 1 to step 4, frequent itemsets in transaction database can be easily minded, but it is not complete for spatial rule mining because SUIT not only represent spatial association but also contains the information of index value for spatial association. In order to find out quantitatively spatially associated frequent itemsets, a second database scanning is necessary. It is regarded as quantitatively spatially associated frequent itemsets if statistical spatial index value is higher than the predefined support level.

Generally speaking, spatial associated frequent itemsets can be got by RAR. Further work will generate spatial association rules from spatially associated frequent itemsets and integrates spatial and attribute value so we can then get comprehensive knowledge.

Our emphasis here is to find out spatial association rules between spatial entities, so the whole procedures can be simplified. Spatial association relation of adjacency is entity-to-entity relation; the basic data structure that RAR is based on is a two-dimensioned table (as seen in table1). In order to illustrate RAR more clearly, we take mining spatial association relation as an example. The detailed procedures are presented in the following.

Data preparation: We translate outer data into coverage (Arc/Info data structure) because it stores the topological information, and then extract every spatial entity and its neighbouring entities and calculate the neighbourhood index value to form Spatial entities and their neighbourhood index value or SENIV (see table 2) in appliance to the neighbourhood-expressing model. Note that B and 0.5 represent the neighbouring entity name and neighbourhood index value with B respectively.

Entity	Neighbouring entity and index value
A	B:0.5; C:0.7
B	A:0.5; C:0.3; 0.4
C	A:0.7; B:0.3; A:0.4; D:0.6
D	A:0.5; C:0.6
A	B:0.4; C:0.4; D:0.5

Table 2. Spatial entities and their neighbourhood index value

Construction of SUIT: Summing neighbouring entities for each class in the same record in SENIV and their index value to form SUIT. The result of Table 1 actually comes from Table 2 after this process.

Construction of neighbourhood matrix of spatial entities: Summing each entity in SUIT according to their classification in Column T (see Table 1, entities are classified 4 types) and entity neighbourhood index value to form neighbourhood matrix of spatial entities. The result is shown in Table 3.

T	A	B	C	D
A	0	0.9	1.1	0.5
B	0.9	0	0.3	0
C	1.1	0.3	0	0.6
D	0.5	0	0.6	0

Table 3. Neighbourhood matrix of spatial entities

From table 3 we can see it is a symmetric matrix that shows the neighbourhood relations between spatial entities. The result shows that spatial class A and B, A and C have high neighbourhood index value. The value is 0.9 and 1.1 respectively.

5.3 Computational Complexity Analysis of RAR

The work is done sequentially by 5 steps. From the procedures as described in Section 5.2, we can see that the computational complexity of RAR depends mainly on Step4 and Step 5. Step1 is a constant time consuming complexity and because Step 2 is simply to scan SUIT and initialise the storage region while Step3 is simply to make summery according to the predefined classification of spatial entities so both the computational complexity of Step2 and Step 3 are O(*n*) where *n* is the total number of the research entities. Step4 is to find out all frequent itemsets and its computational complexity is O(*n*log*n*). If we only consider 2-itemsets as relation of adjacency between spatial entities, the computational complexity will be O(*n*). Step 5 scans databases again and also has to find out the frequent itemsets using the pruning strategy of SSE. So the computational complexity will be O(*n*) + O(*n*log*n*). The result is O(*n*log*n*). Under the worst condition, the computational complexity of RAR will be O(*n*log*n*).

6. A CASE STUDY

6.1 Background

The research areas are agricultural land in two adjacent counties in northern China. The climate of northern China is very in lack of rain and only those crops that accustomed to arid environment can survive. Those crops include peanut, cotton, maize, sorghum, etc. In reality, we make random field investigation on some crop fields by inquiring the farmers. Then we analyse the crops yields and find that some kinds of crop yields have significant difference between the two study areas. In order to find out the deep reasons that may account for the difference, we use the above proposed methods to make sure if there exists spatial association relations between planted crops. Each area of the two has an aviation image for analysis

6.2 Data Preparation

The two images are processed to extract spatial entities by the image analysis software of IDRISI. According to our study goal, we divide the planted crops into 5 categories (viz. peanut, cotton, maize, sorghum and the others). Here <the others> stands for all the other spatial entities except the mentioned 4 crops. We use supervised classification to extract the 5 kinds of spatial entities from remote sensing images. At last we convert the image data structure into vector to create crops covering polygons. We use coverage, data model used for Arc/Info, to present spatial data and use AML, the scripting language for

Arc/Info, to construct SENTV and then get SUIT. From SUIT, neighbourhood matrix of spatial entities can be built.

6.3 Generation of Spatial Association

The spatial object types have been divided into 5 and by joining the neighbourhood index value, SUIT is then built. According to the step 1 of RAR, the column of primary two-dimensioned table have 4 bytes in length and to initialise the initial storage region (5 bits). By RAR, spatial association rules can in the end be mined. The neighbourhood association rule is two-dimensioned like "A is neighbouring to B (s=70%, c=50%)". Take attribute set of spatial objects, comprehensive knowledge as if "if A is neighbouring to B, then A has higher yield (s=70%, c=50%)". For example in our study we get a rule "if cotton is planted surrounded by sorghum, it has high yield".

After extracting spatial object set that has neighbourhood association from all the objects sets, we compare the attribute (average yield) between Region A and Region B. The process is described as follows:

Suppose the spatial object sets can be divided into $C=\{C_1, C_2, \cdots, C_n\}$, and the average value of attribute A of C_i ($i=1, 2, \cdots, n$) in region A is X_A and X_B in region B. Let u_A, u_B and δ_A, δ_B are real value and mean square deviation of attribute A. Let $U=(X_A-X_B)/(\delta_A/m+\delta_B/n)^{1/2}$, then U is fitting to normal distribution N(0, 1). In order to verify hypothesis H: $u_A=u_B$, set confidence level *a* first. If $|U|\geqslant u_{1-0.5a}$, then abandon H, which means X_A and X_B have significant difference. In our case study, by analysing the yield of C_i (cotton) between region A and B, we found they have significant difference. And in our study, we found other factors have no significant influence on yield after the analysis of rule generation, so we can draw that it is spatial association that causes the difference. Further exploration reveals the true reason: cotton that plant surrounded by sorghum has stronger resistance ability to disease. This rule can be used to make decision on crop planting distribution in agricultural planning.

Although the spatial association rules like those mined in the case study is incidental, we can get consequential association rules from incidental production operation through comprehensive knowledge discovery. That means the valuation factor of spatial objects that are fitting to a certain rule has higher (or lower) value than those that do not show this rule. And this rule can be utilized in practice experience.

7. CONCLUSIONS

Knowledge discovered from spatial databases has been recognized as valuable knowledge acquisition in environment management, resource utilization and planning of industry and agriculture. Based on the general discussion of spatial knowledge discovery and spatial rule mining, this article proposed the principle of comprehensive knowledge discovery, concept and mining algorithm, which has a wide application in comprehensive knowledge discovering and utilization. We propose the importance to integrate mining both spatial information and their attribute description and give the way how to attain this goal through theoretical analysis and case study. Although the comprehensive knowledge discovery proposed here focuses on spatial association rule mining and attribute data, it can also be applied in other comprehensive knowledge discovery areas such as spatial classification, spatial clustering, etc., which will be included in our next research.

REFERENCES

Agrawal R, Imielinski T, Swami A, 1993. Mining association rules between sets of items in large databases. In: *Proc. of the 1993 ACM-SIGMOD: International Conference on management of Data*, Washington, DC., pp. 207~216.

Bian Fuling, Sha Zongyao, Chen Jiangping, 2001. Generation of Minimum Rules from Rough Rule Sets Based on Object-Spatial Information Table. *Journal of Wuhan Uniniversity (information version in Chinese)*, 26(5), pp. 399~403.

KERRY TAYLOR, GAVIN WALKER and DAVE ABEL, 1999. A framework for model integration in spatial decision support systems. *Int. J. Geographical information science*, 13(4), pp. 533~555.

Leung, Yee, Ma Jiang-Hong, Zhang Wen-Xiu, 2001. New method for mining regression classes in large data sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(1), pp. 5-21.

Rymon R, 1992. Search through systematic set enumeration. In: *Proceeding of the 3rd International Conference on Principle of Knowledge Representation and Reasoning*. Massachusetts, USA. pp. 539~550

Sauchyn, DJ, Yong Xongchao, 1991. Structure and application of URIFT: An interface between image analysis and vector GIS. *Canadian Journal of Remote*, 17(4), pp. 332-338.

S. Lavingto N, Dewhurst E, Wilkins A Freitas, 1999. Interfacing knowledge discovery algorithms to large database management systems. *Information and software technology*, 41(9), pp. 605~617

S Rahayana, A Siberschatz, 1998. On the discovery of interesting patterns in association rules. In: *Proc. of the 24th VLDB Conference*, New York, USA, pp. 368~379

T.Q. Zeng and Q. Zhou, 2001. Optional spatial decision making using GIS: a prototype of a real estate geographical information science. *Int. J. Geographical information science*, 2001, 15(4), pp. 287~306.

W. Lu, J. Han and B. C. Oci, 1993. Discovery of General Knowledge in Large Spatial Databases. In: *Proc. Far East Workshop on Geographic Information Systems*, Singapore, pp. 275~289.

Ziarko W, 1995. Introduction to the special issue on Rough sets and knowledge discovery. *International Journal of Computational Intelligence*, 11(2), pp. 223~226.

Zhang T, Ramakrishnan R, Livny, 1997. BIRCH: A New Data Clustering Algorithm and Its Applications. *Data Mining and Knowledge Discovery*, 3(1), pp. 141~182.

ACKNOWLEDGEMENTS

This study was partially supported by the China's National Surveying Technical Fund (No. 20007). The authors also deeply appreciate the constructive suggestions made by the paper reviewers to polish this article.

A VARIABLE-SCALE MAP FOR SMALL-DISPLAY CARTOGRAPHY

Lars Harrie [a], L. Tiina Sarjakoski [b], Lassi Lehto [b]

[a] National Land Survey of Sweden, SE-801 82 Gävle, Sweden - lars.harrie@lantm.lth.se
[b] Finnish Geodetic Institute, P.O.Box 15, FIN-02431 Masala, Finland, - tiina.sarjakoski@fgi.fi, - lassi.lehto@fgi.fi

Commission IV, WG IV/3

KEY WORDS: Visualisation, Algorithm, Data structure, Multi-scale database

ABSTRACT:

The aim of this paper is to develop methods for presenting geodata for personal navigation in real-time to a mobile user. Ideally, the user should have an overview map in the vicinity of his position; required for e.g. choosing the right road in a crossing. At the same time the user requires a small-scale map where he can see the final target of his route. A solution to obtain these user requirements is to create a variable-scale map that is constantly changing to make the user position always be in the large-scale part of the map. In the paper a formula for a variable-scale map is derived and its properties are demonstrated. Furthermore, real-time generalisation methods are described, these methods are used to adapt the original cartographic data to the small-scale areas of the map. A prototype system of a variable-scale approach was created using the emerging XML-based vector-data standards (GML and SVG), where the generalisation and scale-variations were performed in an XSLT transformation. The paper describes a minor case study which shows that variable-scale maps have a potential for personal navigation.

1. INTRODUCTION

1.1 Background

In the near future an increasing number of maps will be displayed in portable computers and mobile telephones. These maps are often denoted mobile maps. One major problem with mobile maps is the small displays, which put high demands on the selection of the cartographic data to be shown. This is problematic when the user requires much cartographic information. In personal navigation, for example, the user often requires both a detailed map around the current user posistion as well as an overview map. That is, in cartographic terms, the user requires both large-scale and small-scale cartographic data. This paper concentrates on a method for giving the user both large-scale and small-scale map information simultaneously on a small portable display.

There are, principally, three different approaches to provide the user both with large-scale and small-scale maps on a single display:

1) The user switches between a small-scale map and a large-scale map.
Advantage: The user gets full view of both the small-scale and the large-scale map.
Disadvantage: The user does not see the two maps simultaneously; therefore, the user might have difficulties in connecting the information of the two maps. Furthermore, switching between maps is necessary.

2) The large-scale map is shown in the ordinary map window and the small-scale map in a key-map.
Advantage: The user sees the two maps simultaneously.
Disadvantage: Even though that the user sees the two maps at the same time it is often problematic to identify common objects in the maps. In addition, the key map hides data in the ordinary map window or it is often difficult to get sufficient space for both the maps at the same time in a small display.

3) Large-scale and small-scale cartographic data are presented in the same map. In the vicinity of the user large-scale data is shown and far from the users position small-scale data is shown.
Advantage: The large-scale data and small-scale data are integrated.
Disadvantage: The map has a variable scale as well as discrete jumps between different representations.

This study is devoted to the third approach; that is, the paper is concentrated on variable-scale maps. Even though many studies have been performed using variable-scale maps (see Section 1.2) we think that the subject requires more attention. The technical development has given to the variable-scale approach new possibilities in at least three aspects. Firstly, today it is possible to have a mobile device showing cartographic data where the user always will be in the centre of the map (possible since the user position is known e.g. by cellular network or Global Positioning System (GPS)). Secondly, new methods for real-time integration and real-time generalisation make it possible for tailoring the map for specific user requirements. Thirdly, the emerging vector-data standards will enhance the possibility to transfer cartographic data to a mobile user. That is, in principle, it will be possible to create a variable-scale user-tailored map, that is updated while the user is moving.

This paper starts with a survey of some previous studies of variable-scale maps. In the second section theory of our variable-scale maps is given. In our approach a map in the map projection plane is required as input data, and the mapping introduced here is a distortion to this original map. The third section focuses on the technical environment as well as on vector-data standards used in the study. The type of variable-scale map presented here requires that some data be generalised in real-time. This is further described in Section 4. The fifth section contains case studies. The paper concludes with discussion and conclusions.

1.2 Previous studies of variable-scale maps

There are basically two aims of variable-scale maps. Firstly, a variable-scale map is used when a constant-scale map is not sufficient for visualising all the required data. Secondly, a variable-scale map is used to make the density of a variable uniform; such maps are normally called cartograms (see e.g. Gusein-Zade and Tikunov, 1993). In this paper we are only interested in the first aim of a variable-scale map.

Much research has been devoted to develop map projections that give as little distortions as possible. To create a variable-scale map the aim is quite the opposite; the scale is intentionally varying so that the map can visualise data with uneven distribution or with the aim of focusing on an area of interest. A few map projections have been designed for this specific purpose. Snyder (1987) derived some azimuthal projections with a magnifying glass effect. These projections give a map with circular cap where the radial scale is constant out from the centre point (of the azimuthal projection); outside the circular cap the radial scale is smaller. As mentioned above, however, the mapping used in this paper uses coordinates in a map projection plane as input coordinates and not geographic coordinates (as in map projections).

There are a few examples of applied use of variable-scale maps. Hägerstrand (1957) used a logarithmic azimuthal map projection to visualise migration patterns. He found that most people only moved within a region, but a few moved between regions; to visualise this migration patterns he realised that a constant-scale map was insufficient. Commercial examples of variable-scale maps are the city maps by Falk-Verlag, Germany.

Fairbairn and Taylor (1994) proposed variable-scale maps for urban areas. They argued that the density of objects is larger in the city centre and therefore the scale should be larger there. In their study they implemented a variable-scale map where the scale decreases linearly from the centre of the map towards the edges (formulas for similar scale variations were also derived by Kadmon, 1975). Furthermore, Lichtner (1979) used a transformation that magnified several parts of urban thematic maps. Similar ideas were presented by Guerra and Boutoura (2001). They proposed that a magnifying glass could be moved around in a city map and that the area of the glass should be presented at another window where the data are magnified non-linearly. It should be noted that all these studies used coordinates in the map projection plane as input coordinates to the scale variation transformation.

Special types of variable-scale maps are the schematic maps (some authors use the term cartograms for this kind of maps, see e.g. Kadmon, 1982). These maps have great geometrical distortions, but are topologically correct; they are mainly used for presenting subway lines and other transportation networks. Recently, research has been devoted to establish automatic methods to generate and query schematic maps (Avelar and Müller, 2000; Avelar and Huber, 2001).

Agrawala and Stolte (2001) presented a study for maps in personal navigation. They aimed at producing a navigation map that resembles a hand drawn map sketch. Agrawala and Stolte argued that humans are mainly interested in information about turning points of the routes and not the links between the turning points. Therefore, it is better to shrink some road elements (i.e., use a smaller scale locally) to be able to give all the required data about the turning points. The study presented in this paper is based on the same basic philosophy as the study in Agrawala and Stolte (2001); the main difference lies in that while the user in their study was given a single map the user in our study will receive an updated map while moving along the route (when a complete prototype exist).

2. VARIABLE-SCALE MAPS

2.1 Basic properties

The aim in this study is to create a map for a mobile user. This user communicates with a cartographic database. This implies that the map can be tailored for the current location of the user. A main idea of the variable-scale map, used in this study, is to show large-scale data for a circular cap around the user and small-scale data outside this circular cap (see Figure 1).

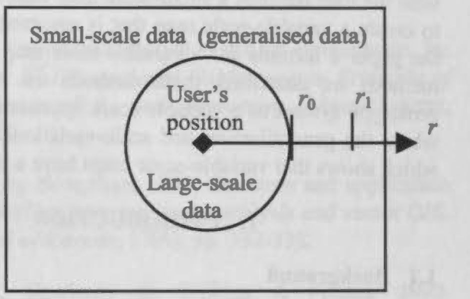


Figure 1. The figure illustrates the circular cap (with radius r_0) where large-scale data is shown and the area outside this cap where small-scale data is shown. Inside the circular cap the scale is uniform (and equal to s_l) and outside radius r_1 the radial scale is constant (and equal to s_s).

2.2 Mapping function

A characteristic of a variable-scale map is that is not using a uniform scale in the whole map; in our application only the data within the circular cap is shown in a uniform scale s_l (Figure 1). This scale is equal to the preferred presentation scale of the original data set. Outside the cap the scale is decreasing according to the following rules:

- The scale should be continuous over the whole map.
- The scale (in radial direction) should be equal to s_l at the radial distance r_0 from the user position. The scale (in radial direction) is then decreasing by a constant factor to be equal to s_s at a radial distance of r_1. For distances longer than r_1 the scale (in radial direction) should also be equal to s_s.

These rules can be satisfied by a mapping from the original data (stored at a scale of s_l) to the new data to be displayed. This mapping can be constructed by a radial displacement function where r_{new} (the new radial distance) is a function of r (the original radial distance). It should be noted that the angle is not changed in the mapping. By applying the rules above we get the following radial displacement functions:

$$r_{new}(r)=r\,,\ r<=r_0$$

$$\begin{aligned} r_{new}(r)&=r_0+\frac{1}{s_l}\int_{u=r_0}^{r}\left(s_l+(s_s-s_l)\cdot\frac{u-r_0}{r_1-r_0}\right)du=\\ &=r+\frac{(r-r_0)^2\cdot(s_s-s_l)}{2\cdot s_l\cdot(r_1-r_0)},\ r_0<r<=r_1 \end{aligned} \tag{1}$$

$$\begin{aligned} r_{new}(r)&=r_0+\frac{1}{s_l}\int_{u=r_0}^{r_1}\left(s_l+(s_s-s_l)\cdot\frac{u-r_0}{r_1-r_0}\right)du+(r-r_1)\cdot\frac{s_s}{s_l}=\\ &=r_1+\frac{(r_1-r_0)\cdot(s_s-s_l)}{2\cdot s_l}+(r-r_1)\cdot\frac{s_s}{s_l},\ r>r_1 \end{aligned}$$

where r_0 and r_1 are defined according to Figure 1. In Figure 2 a radial displacement function is shown.

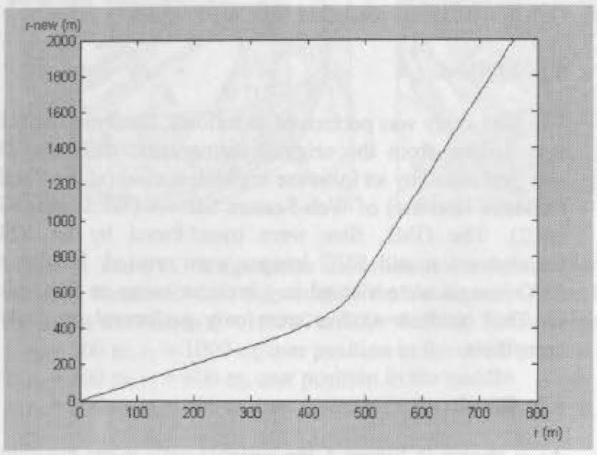


Figure 2. The radial displacement function for the input parameters: r_0=300 m, r_1=600 m, s_l=1:10 000 and s_s=1:50 000

2.3 Scale variations

It is important to note that the mapping function above is not conformal. In fact, there exists no conformal mapping that satisfy our requirements of the scale variations in radial direction (unless the uninteresting case when $s_s=s_l$; this follows from that a mapping between two annulus is conform only if the ratio between the radius are constant, see e.g. Nehari, 1975).

From the requirements used in deriving the mapping function we can directly determine that the scale in radial direction ($scale_r$) is equal to (this formulas are also given by the partial derivatives $\partial r_{new}/\partial r$):

$$scale_r(r)=1,\ r>=r_0$$

$$scale_r(r)=1+\frac{(r-r_0)}{(r_1-r_0)}\cdot\frac{(s_s-s_l)}{s_l},\ r_0<r<=r_1 \tag{2}$$

$$scale_r(r)=\frac{s_s}{s_l},\ r>r_1$$

The scale in tangential direction ($scale_t$) is simply given by:

$$scale_t(r)=\frac{r_{new}(r)}{r},\ \forall r \tag{3}$$

As is shown in Figure 3 the scale in tangential direction is larger than the scale in radial direction outside the cap. This is also seen in Figure 4, where the original grid has a hyberbolic form (cf. figures in Lichtner, 1979). This property is unavoidable if we want the central part of the map in larger scale. Figure 5 gives an example on how this transformation change a topographic map.

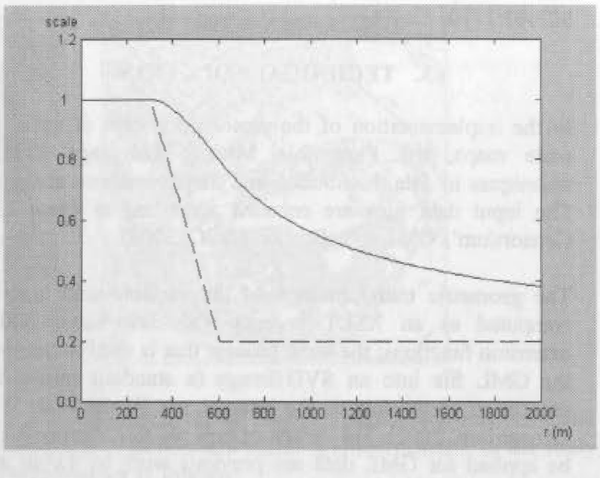


Figure 3. Plot showing the scale in radial (dashed) and tangential (solid) direction of the mapping function (as a function of the original radial distance (r)). The following input parameters were used: r_0=300 m, r_1=600 m, s_l=1:10 000 and s_s=1:50 000.

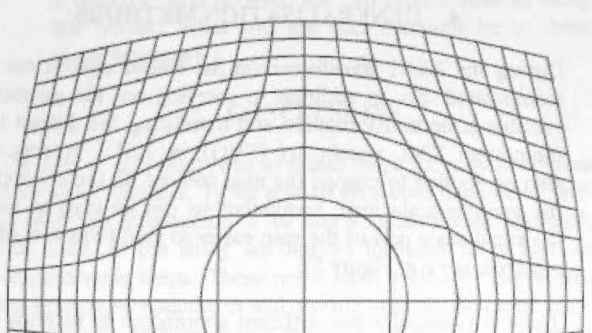


Figure 4. Distortions to an initial grid by variable-scale mapping. The following input parameters were used: r_0=300 m, r_1=600 m, s_l=1:10 000 and s_s=1:50 000. The circle represents the circular cap with radius r_0 (cf. Figure 1).

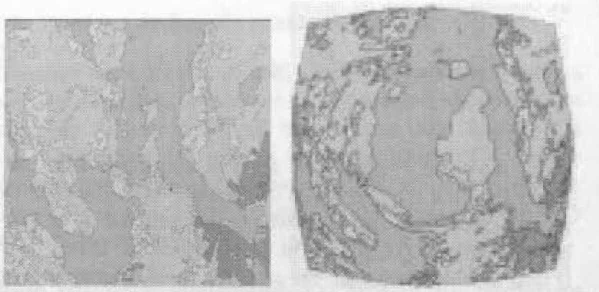


Figure 5. Distortions to a topographic map (same input parameters as in Figure 4). From Terrain map, © National Land Survey of Sweden, Gävle 2002. Printed by permission M2002/2770.

3. TECHNICAL SOLUTION

In the implementation of the prototype system of variable-scale maps, the Extensible Markup Language (XML) techniques in data distribution and transformations are used. The input data files are encoded according to Open GIS Consortium's GML specification (OGC, 2002).

The geometric transformation of the variable-scale map is computed as an XSLT process with Java-based XSLT extension functions; the same process that is used to translate the GML file into an SVG image (a standard format for presenting vector data on the Internet, see World Wide Web Consortium, 2002). For details of how an XSLT process can be applied for GML data see previous work by Lehto and Kilpeläinen (2000, 2001a,b).

The free-of-charge XML software components applied in the prototype system include an XML parser (Xerces), an XSLT processor (Xalan) from Apache community (Apache, 2002) and the SVG Viewer plugin from Adobe (Adobe, 2002).

4. GENERALISATION METHODS

During the XSLT transformation the source dataset can be manipulated. So, in addition to carrying out the geometric variable-scale transformation and translating the dataset into the correct XML vocabulary (SVG), an XSLT process can also be applied to change the map content. In the context of the variable-scale map, manipulations can be used to make the small-scale part of the map easier to read (which is often required, see e.g. Figure 5).

There are basically two different approaches to increase the readability of the small-scale part of the map. The first approach is to use two different data sets: a large-scale data set in the cap and a small-scale data set outside the cap. A difficulty with this approach is to integrate the two data sets along the border of the cap. To enable such integration a multiple representation database (MRDB) is preferably used (an MRDB consist of different data sets that are connected by links between objects that represent the same physical entities; see e.g. Buttenfield, 1993; Kilpeläinen, 1997; Timpf, 1997; Harrie and Hellström, 1999).

The second approach is to use only a large-scale data set. This data set is then generalised outside the cap. This implies that there is no need to integrate two different data sets; instead this approach requires suitable methods for real-time generalisation. Even though there has been research in the field of real-time generalisation (see e.g. van Oosterom and Schenkelaars, 1995; Jones et al., 2000; Lehto and Kilpeläinen, 2000, Lehto and Kilpeläinen, 2001a, b) the task of real-time generalisation is still far from being solved.

In this study we chose the second approach (i.e., only using one single data set). The implemented generalisation operators in this study include building simplification and selection. In the implemented selection function individual spatial objects can be selected or rejected for inclusion in the result dataset based on their feature type. Decisions are also based on properties derived during the transformation. Building simplification is based on an algorithm that constructs the minimum enclosing rectangle of the building and orients a scaled version of it along the longest side of the building.

5. CASE STUDY

A case study was performed on data from City of Helsinki. Only building and road data were used in these studies.

5.1 Method

The case study was performed as follows. Firstly, a GML file was derived from the original cartographic database. This was performed by an in-house implementation (at the Finnish Geodetic institute) of Web Feature Service (WFS, see OGC, 2002). The GML files were transformed by an XSLT transformation and SVG images were created. Finally, the SVG images were viewed in a browser using an SVG plug-in. That is, these studies were only performed on desktop computers.

5.2 Result

As is shown in Figure 6 the variable-scale maps are able to show data of different level of details and of different scales within a single map. As is also shown, this is done at the expense of other properties of the map: the rectangular street network is no longer rectangular and the feeling of the distribution of the buildings is somewhat deteriorated.

6. DISCUSSION

This paper describes an ongoing project; there are several issues that require more studies. Some technical issues are:

- So far we have only tested variable-scale maps in a desktop environment. To really evaluate the method, it should be tested using mobile devices. This study is a part of a larger project GiMoDig (GiMoDig, 2002); our aim is to evaluate the variable-scale maps displayed in mobile devices later within the frame of this project.
- The issue of time for transformation and distribution of data are crucial for mobile maps, but the time respond problem has not been treated in this study.
- More and better methods for real-time generalisation are required. The methods for real-time generalisation in this case study consist of the operators simplification and selection. Although the mobile maps use to show a very simplified picture of the reality, in a real usage situation these simple routines are not sufficient. Besides the aggregation operator, earlier demonstrated as an XSLT transformation by Lehto and Kilpeläinen (2001a,b), also other operators are needed.

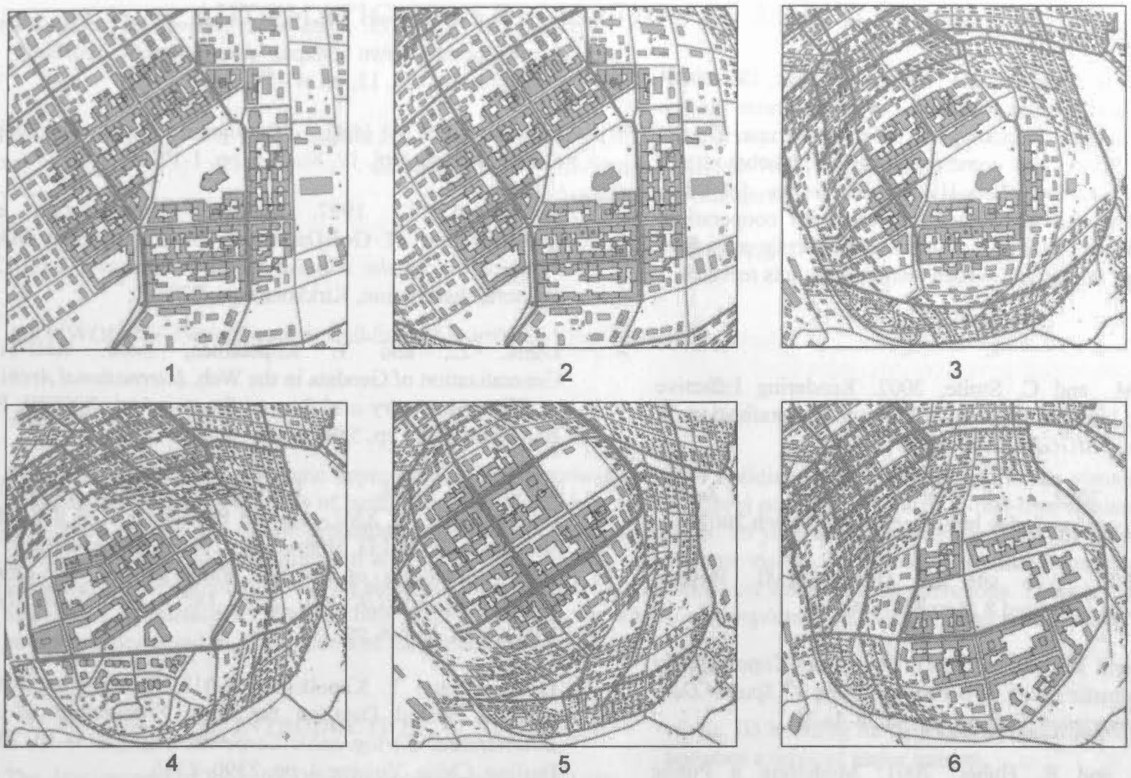


Figure 6. The SVG image series above illustrates the transformation variation depending on the changed parameters. Besides scale-variable, the small-scale map outside the circular cap r_0 has been generalised by selection and building simplification. In the study the scales were set to s_l = 1:1000 and s_s = 1:4000. The remaining parameters were set as follows:

1: $r_0 = 300$ m, $r_1 = 1000$ m, user position in the middle
2: $r_0 = 300$ m, $r_1 = 600$ m, user position in the middle
3: $r_0 = 200$ m, $r_1 = 400$ m, user position in the middle
4: $r_0 = 200$ m, $r_1 = 400$ m, user position in the lower left corner
5: $r_0 = 200$ m, $r_1 = 400$ m, user position in the upper left corner
6: $r_0 = 200$ m, $r_1 = 400$ m, user position in the lower right corner

- So far all our studies have been based on only one dataset. Whether one or several datasets should be used for variable-scale maps is a manifold question. Using only one dataset would imply that we avoid the problems of inconsistency between the datasets. Using two datasets, in turn, suffers from a data integration problem.

To conclude the last two technical issues above, we face a complicated problem whether we use one or several data sets. It is questionable if an XSLT transformation is sufficient for solving these problems (without extensive own coding). A possible solution might be to use a geometrical and topological Java library. Parallel to this project, studies are performed of using such a Java library (Java Topology Suite; see Vivid Solutions, 2000) in an XML environment. We intend to test this library for real-time generalisation and integration of geodata in future studies.

Besides the technical aspects, the user aspects are of course important. So far our studies have not incorporated any user studies. To compare the advantages and disadvantages of variable-scale maps the user studies should concentrate on questions such as:

- Will the user be able to make use of the integrated large-scale and small-scale data?
- Since the mapping is not conformal the mapping will destroy known geometries. What are the user reactions to distortions of e.g. rectangular street networks?
- A problem of the variable-scale map is the discrete jumps of representation along the border of the cap. For example, if the large-scale data represent buildings at the border area as building objects and the small-scale data the same area as built-up area objects there will be a discrete jump between the two representations along the border. What will the user reactions be to these discrete jumps?

7. CONCLUSIONS

The aim of this study was to create maps for mobile devices with small displays. To be able to give both detailed as well as overview information to the user a variable-scale approach was used. In the study we derived formulas for a type of variable-scale maps. These maps have a circular cap where the scale is homogeneous and outside this cap the radial scale is constantly decreasing to a threshold value. The map is not conformal; the distortions are visualised in figures in the paper.

The variable-scale approach was implemented in an XML environment; the transformations, as well as real-time generalisation of the geodata outside the cap, were performed as XSLT transformations.

A case study shows that variable-scale maps are able to present integrated large-scale and small-scale data. We think the result is promising and will continue to work with technical improvements and user studies.

Acknowledgements

This research is a part of the GiMoDig project, IST-2000-30090, which is funded from the European Union via the Information Society Technologies (IST) programme. Thanks to Monika Sester for comments, Stefan Jakobsson for mathematical advice, Lars-Håkan Bengtsson for data transformations and to Mikael Johansson for cooperation. Test data from the City of Helsinki and the terrain map from National Land Survey of Sweden were used in this research.

References:

Agrawala, M., and C. Stolte, 2002. Rendering Effective Route Maps: Improving Usability Through Generalization. In *Proceedings of SIGGRAPH01*, pp. 241-249.

Adobe, 2002. "The SVG Zone", www.adobe.com/svg/index.html (accessed 8 March 2002).

Apache, 2002. Web site of Apache XML Project, xml.apache.org (accessed 8 March 2002).

Avelar, S., and M. Müller, 2000. Generating Topologically Correct Schematic Maps. *Proceedings of the 9th Spatial Data Handling Symposium*, Beijing, pp. 4a.28-4a.35.

Avelar, S., and R. Huber, 2001. Modelling a Public Transport Network for Generation of Schematic Maps and Location Queries. In: *Proceedings of the 20th International Cartographic Conference*, Beijing, China, p. 1472-1480.

Buttenfield, B. P., 1993. *Research Initiative 3: Multiple Representations*, Closing report, National Center for Geographic Information and Analysis, Buffalo.

Fairbairn D., and G. Taylor, 1995. Developing a variable-scale map projection for urban areas. *Computers & Geosciences*, Vol. 21, No. 9, pp. 1053-1064.

GiMoDig, 2002. Geospatial info-mobility service by real-time data-integration and generalisation, http://gimodig.fgi.fi/ (accessed 11 March 2002).

Guerra, F., and C. Boutoura, 2001. An Electronic Lens on Digital Tourist City-Maps. In: *Proceedings of the 20th International Cartographic Conference*, Beijing, China, p. 1151.

Gusein-Zade, S. M., and V. S. Tikunov, 1993. A New Technique for Constructing Continuous Cartograms. *Cartography and Geographic Information Systems*, Vol.20, No. 3, pp. 167-173.

Harrie, L., and A.-K. Hellström, 1999. A Prototype System for Propagating Updates between Cartographic Data Sets. *The Cartographic Journal*, Vol. 36, No. 2, pp. 133-140.

Hägerstrand, T., 1957. Migration and area: migration in Sweden. *Lund Studies in Geography*, Series B, Human Geography, No. 13, pp. 27-158.

Jones, C. B., A. I. Abdelmoty, M. E. Lonergan, P. van der Poorten, and S. Zhou, 2000. Multi-Scale Spatial Database Design for Online Generalisation. *Proceedings of the 9th Spatial Data Handling Symposium*, Beijing, pp. 7b.34-7b.44.

Kadmon, N., 1975. Data-bank Derived Hyperbolic-scale Equitemporal Town Maps. *International Yearbook of Cartography*, Vol. 15, pp. 47-54.

Kadmon, N., 1982. Cartograms and topology. *Cartographica*, Vol. 19, No. 3-4, pp. 1-17.

Kilpeläinen, T., 1997. Multiple Representation and Generalization of Geo-Databases for Topographic Maps. *Publications of the Finnish Geodetic Institute*, No. 124, Doctoral dissertation, Kirkkonummi, 229 p.

Lehto, L., and T. Kilpeläinen, 2000. Real-Time Generalization of Geodata in the Web. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII, Part B4, Amsterdam, pp. 559-566.

Lehto, L. and T. Kilpeläinen, 2001a. Real-time Generalisation of XML-encoded Spatial Data on the WEB. Kidner, D. B. and G. Higgs, eds., 2001, GIS Research in the UK, *Proceedings of the GIS Research UK, 9th Annual Conference GISRUK 2001,* April 18th – 20th, University of Glamorgan, Wales, pp. 182–184.

Lehto, L. and T. Kilpeläinen, 2001b. Generalizing XML-encoded Spatial Data on the Web. *Proceedings of 20th International Cartographic Conference,* August 6–10, 2001, Beijing, China, Volume 4, pp. 2390–2396.

Lichtner, W., 1979. Kartennetztransformationen bei der Herstellung thematischer Karten. *Nachrichten aus dem Karten- und Vermessungswesen,* Reihe I, Heft 79, Frankfurt am Main, pp. 109-119.

Nehari, Z., 1975. *Conformal mapping*. Dover Publications, Inc., New York.

OGC, 2002. OpenGIS® Implementation Specifications, http://www.opengis.org/techno/implementation.htm (accessed 13 March 2002).

Snyder, J. P., 1987. "Magnifying-Glass" Azimuthal Map Projections. *The American Cartographer*, Vol. 14, No. 1, pp. 61-68.

Timpf, S., 1997. Cartographic Objects in a Multiscale Data Structure. In Craglia, M., and H. Couclelis (eds), *Geographic Information Research – Bridging the Atlantic*, Taylor & Francis, pp. 224-234.

van Oosterom, P., and V. Schenkelaars, 1995. The Development of an Interactive Multi-Scale GIS. *International Journal of Geographical Information Systems*, Vol. 9, No. 5, pp. 489-507.

Vivid Solutions, 2000. Java Topology Suite, http://www.vividsolutions.com/jts/jtshome.htm (accessed 13 March 2002).

World Wide Web Consortium, 2002. Scalable Vector Graphics (SVG), http://www.w3.org/Graphics/SVG/ (accessed 8 March 2002).

AUTOMATIC GENERALISATION OF 3D BUILDING MODELS

Martin Kada

Institute for Photogrammetry (ifp), University of Stuttgart, Germany
Geschwister-Scholl-Strasse 24D, D-70174 Stuttgart
martin.kada@ifp.uni-stuttgart.de

Commission IV, WG IV/3

KEY WORDS: Generalisation, Buildings, Algorithms, Visualisation

ABSTRACT:

The paper presents an automatic approach for the generalisation of 3D building models with regard to the visualisation of urban landscapes. Simplified versions of such models are not only needed for level of detail structures in real-time rendering, but also for web-based 3D GIS and for the presentation on mobile computing devices. To yield more sophisticated building models compared to already known surface simplification algorithms from the field of computer vision, the presented solution is based on least squares adjustment theory combined with an elaborate set of surface classification and simplification operations. This concept allows for the integration of surface regularities into the building models which are important for visual impression. These regularities are stringently preserved over the course of the generalisation process.

1. INTRODUCTION

The development of tools for the efficient collection of 3D city models has been a topic of intense research for the past years. In addition to Digital Height Models and 3D data representing streets and urban vegetation, building models are the most important part thereof. Meanwhile, a number of algorithms based on 3D measurement from aerial stereo imagery or airborne laser scanner data are available for automatic and semi-automatic collection of 3D building models. As an example, Figure 1 shows a 3D model of Stuttgart collected by the approach of (Haala and Brenner, 1999). A good overview on the current state-of-the-art of experimental systems and commercial software packages is for example given in (Baltsavias, Grün and van Gool, 2001). Almost all of these systems describe the reconstructed buildings by general polyhedrons, since a building representation by planar faces and straight edges is feasible for most cases. The resulting 3D boundary representation is either provided directly or constructed from a number of building primitives which are used during the measurement process.

Originally, simulations for the propagation of electromagnetic waves used for the planning of antenna locations were the major application areas for 3D building models. Meanwhile visualisation in the context of three-dimensional car navigation systems, virtual tourism information systems or city and building planning has become the key market for that type of data. In our opinion one of the most important development-driving forces for the application of 3D city models is the widespread use of mobile devices for the provision of location based services. Features like personal navigation or telepointing, i.e. the provision of spatial information by pointing to regions of interest directly on the display, presume a realistic visualization of the 3D urban environment on these mobile devices. Due to the limited amount of computational power and small size of the displays on the one hand and the huge amount of data contained within a 3D city model on the other hand, the amount of information to be handled, stored and presented has to be reduced efficiently. Thus, the generalisation of the 3D building models as it is described within this article becomes a topic of major interest.

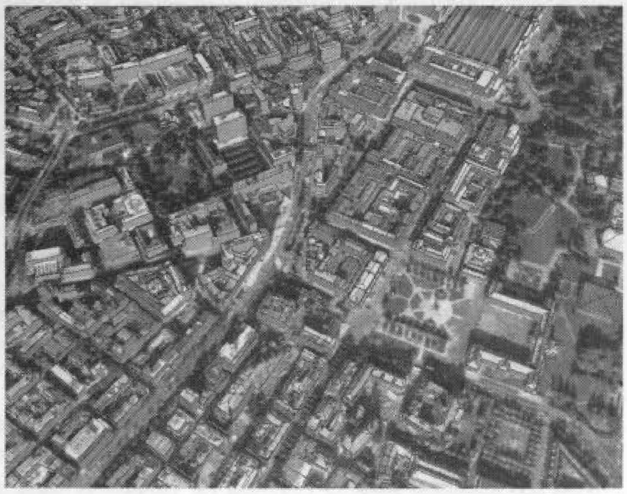


Figure 1. A 3D city model of Stuttgart.

In general, this process presumes the elimination of unnecessary details, whereas features, which are important for the visual impression, have to be kept. Especially for man-made-objects like buildings, symmetries are of major importance. For this reason, during the process of generalisation the preservation of regular structures and symmetries like parallel edges, perpendicular intersections or planar roof faces has to be guaranteed. In principle this geometrical regularisation is also required during data collection, since otherwise geometric errors introduced during measurement could result in erroneous structures of the building. Compared to the geometric regularisation during measurement, as e.g. described in (Grün and Wang, 2001), this problem is even aggravated during the simplification process due to the increasing deviations from the true building geometry.

2. RELATED WORK

The focus of this paper is on automatic simplification of polyhedral building models. A vast amount of research efforts have already been put into the generalisation of building ground plans and on general surface simplification algorithms. Whereas the former is a typical domain of cartography, surface simplification is a widely used technique in the field of computer vision to speed up the visualisation of highly complex models. As of today, very little research has been done to extend existing model generalisation techniques to work with three-dimensional building models or to adapt surface simplification algorithms to the specific needs of buildings.

Both techniques – model generalisation and surface simplification – which are combined in our work will be discussed briefly.

2.1 Model Generalisation

Model generalisation is the transformation of objects into representations of simplified geometry, topology and semantics. An early approach for the simplification of building data is described in (Staufenbiel, 1973), where the outline of the building is based on the intersection of straight lines and a set of rules is proposed for which are too small for presentation. Recent approaches incorporate object-oriented structures and rules to strive for a more holistic solution. (Barrault et al., 2001) e.g. present a hierarchical multi-agent system where a set of agents are delegated to a building, each aiming to improve the overall situation with respect to some attached constraint. Such rules often require a minimal building size, rectification of angles and enlargement of narrow objects inside the building.

A general concept using least squares adjustment theory for the simplification of building ground plans was first introduced by (Sester, 2000). It allows for the introduction of observations in terms of constraints in order to determine unknown parameters in an optimisation process

Methods for the generalisation of 3d building data have also been proposed in recent years. (Meyer, 2000) suggests using a sequence of opening, closing and rectification operations to gain simplified data which can be used to build up a level of detail structure. A different approach is presented in (Coors, 2000): a well known surface simplification algorithm (as described in (Garland and Heckbert, 1997)) is extended to enhance significant features of the model and to aggregate only the less important ones.

2.2 Surface Simplification

In the field of computer graphics, computer vision and computational geometry, a wide range of surface simplification algorithms have been developed. A good overview is given by (Heckbert and Garland, 1997). Those methods are usually applied for the simplification of general objects which are either given as polygonal or as triangular surface meshes. The most important algorithms are either based on vertex clustering or edge collapse operations.

The algorithm introduced by (Rossignac and Borrel, 1993) divides the object's bounding volume into a regular grid of boxes and all vertices inside a cell are clustered together into a single vertex. A simplified model is then synthesized from the remaining vertices according to the original topology. The simplification algorithms presented by (Hoppe, 1996) and (Garland and Heckbert, 1997) both iteratively contract edges to simplify models, but differ in the underlying error metric measuring the geometric error introduced into the model by an edge collapse operation.

Another interesting simplification approach for general polygonal models is described in (Ribelles et al., 2001). Small features including bumps, holes, tabs, notches and decorations are isolated, ranked and removed using a splitting and hole filling operation.

Building regularity as requested for our problems have not been used so far.

3. ALGORITHM OUTLINE

The presented generalisation algorithm is designed for polyhedral three-dimensional building models. Without loss of generality, it is assumed that the 3D building model is given as a 2-manifold M, composed of a set of vertices V and a set of polygonal faces F. Each face may additionally contain a number of interior points, determining the parameters of the associated planar surfaces. These interior points are provided during data collection, e.g. from stereo measurements, or result from vertices that are removed during the geometric simplification of the building model. The algorithm uses these interior points in the least squares adjustment to resolve the new coordinates of vertices after each generalisation step. This approach ensures a minimum deviation of the generalised building model to every vertex of the original model and thus to its original shape.

Our algorithm is based on the fact that most walls are oriented in parallel to the principal axes of the building, which are again often rectangular. It can therefore be assumed that the faces of a building model are usually coplanar, parallel or rectangular to other faces in the same model. A generalisation must preserve these properties as correct as possible. For this reason, the presented algorithm considers the aforementioned properties between faces as constraints during the simplification process. As this information is usually not explicitly available for a building model, the first step in the generalisation algorithm is to create the so-called constraint building model, which is basically the polygonal building model enriched by a set of constraints.

These constraints are not stored for pairs of faces, however, as this would lead to a large number of constraints, but as a hierarchy of constraints. The lowest element in this hierarchy is the coplanarity constraint, which simply groups a set of faces together, each being coplanar to any other face in the same set within some given tolerance. Sets of coplanar faces are then again grouped together by a parallelism constraint if their faces are parallel to faces of another face set. Finally, two or three sets of coplanar or parallel faces are grouped by a rectangularity constraint if the faces of each set are rectangular to faces in the other two or three sets.

Following the generation of the constraint building model, the geometry of the model is iteratively simplified as depicted in Figure 2. First, a feature detection algorithm searches for features like extrusions, intrusions, notches, tips etc. and evaluates their significance to the overall appearance of the model. A feature removal step next eliminates the features of least importance, i.e. which only slightly influence the silhouette of the building. The feature removal step not only alters the geometry of the constraint building model, but also the constraints that are affected by it. This is important, as constraints become obsolete through the process of feature removal. For example, sets of coplanar faces may often be merged together after simplification. Vertices that are removed from the geometry, however, are not just discarded from the model, but the algorithm stores their coordinates as additional interior points as mentioned in the beginning of this section. The last step of the algorithm uses least squares adjustment to find a new position for every vertex in the constraint building model in order to fulfil all its constraints.

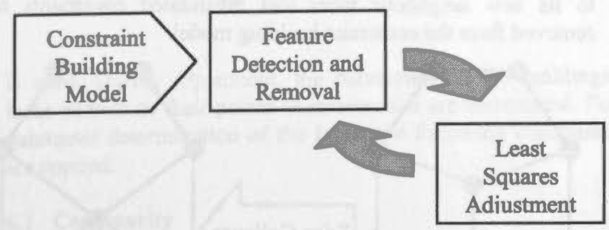


Figure 2. The generalisation algorithm iteratively simplifies the constraint building model using least squares adjustment to preserve the building regularities.

4. CONSTRAINT BUILDING MODEL

The initial step of the generalisation algorithm is to build the constrained building model, which stores constraints between two or more faces of the polyhedral building model. The following adjustment step of the algorithm is based on this information to optimise the coordinates of all vertices after each feature removal step. The simplification step on the other hand is designed to avoid violating constraints until they become obsolete. Simplification operations that are carried out on the building model also aim on preserving the coordinates of affected vertices and rely on the adjustment stage to determine their final position. Thus, the quality of the final, generalised building model directly depends on the quality of the constraints that are stored within the constraint building model.

It is the author's belief that not every constraint can be found by an automatic approach. Dependent on the quality of the input model, a number of constraints will almost always be missed due to errors introduced in the generation of the model. Those absent constraints might reduce the quality of the final model if missed in high quantities. An application should therefore offer the possibility to identify and insert more constraints into the constraint building model in a semi-automatic fashion to work around those errors and to improve the overall quality of the final building model. A semi-automatic tool also helps surveying the effects of certain constraints on the generalisation process by manually adding or removing those constraints.

4.1 Prerequisites

It is assumed that for each face F_i, a non-ambiguous plane can be computed using the coordinates of all its vertices and interior points. Each planar Face F_i is then given by the following plane equation:

$$F_i : A_i x + B_i y + C_i z + D_i = 0 \qquad (1)$$

where A_i, B_i, C_i is the normal vector of its plane and D_i the closest distance of the plane to the origin of the coordinate system. The angle ϕ between two planar faces F_1 and F_2 can then be computed using:

$$\phi = \arccos(A_1 A_2 + B_1 B_2 + C_1 C_2) \qquad (2)$$

The parameters and the angle between two planar faces are not only used to find properties between faces, but also to classify features of the building in later stages.

4.2 Properties between Faces

The algorithm identifies faces being coplanar, parallel and rectangular to other faces as described in the following subsections.

Coplanarity: Two faces are assumed to be coplanar if the angle between the normal vectors is close to 0° or 180° and the difference of the absolute value of the distances lies under a given threshold. It is worth mentioning, that coplanar faces may have normal vectors pointing in opposite directions. This is totally legal as the other constraints are expressed independently of the true direction of the faces.

Parallelism: Two faces are assumed to be parallel if the angle between their normal vectors is close to 0° or 180°.

Rectangularity: Two faces are assumed to be rectangular if the angle between their normal vectors is close to 90°.

4.3 Organisation of Constraints

As mentioned earlier in this paper, the algorithm does not store one constraint for each pair of faces as this would lead to a large number of constraints. Rather, the algorithm generates groups of faces, so that there exists a coplanarity constraint between any two faces placed in the same group. As each one of those face groups defines a unique plane inside the buildings own models space, the real goal of detecting coplanar faces is to find a minimal set of planes and to associate every face with exactly one of those planes.

The algorithm further groups two or more planes into sets of parallel planes if the faces in those planes are parallel to faces in the other planes. Finally, two or more (parallel) planes are grouped by a rectangularity constraint to yield the aforementioned hierarchy of constraints.

5. FEATURE DETECTION AND REMOVAL

In order to simplify the geometry of a building model, it is not sufficient to just remove arbitrary vertices or edges. Even if the introduced geometric error is small, the symmetry of the building model will irretrievably get disturbed. It is thus necessary to take notice to the regularity of the model during its simplification. Our feature detection and removal algorithm for generalisation allows the use of a manifold set of surface simplification operators, each designed to remove one specific class of feature types. In contrast to the rather simple operators used in traditional surface simplification algorithms, our operators remove entire features in one continuous process, while preserving the integrity of the remaining parts of the building model.

Prior to removing some feature, it must first be detected and identified, because each feature type requires its particular feature removal operator. Three classes of features types can be distinguished, each based on one of the three primitive types: the extrusion, the notch and the tip (Figure 3).

The presented algorithm detects features by looping through all primitives of the building model and by testing them against the feature types of the particular feature class. Once the algorithm detects a feature, its impact to the appearance of the silhouette of the building model is evaluated. This can be a very complicated process as small features may be important due to their semantic meaning. At this point, we use a simple metric to measure a value which corresponds to the maximum distance moved by a vertex during the removal of the feature.

After feature detection is completed, the algorithm removes the features of lowest importance to simplify the geometry of the building model. In our example, an extrusion is removed by using a combination of edge collapse and edge foreshortening operations (see section 5.2). Then, the algorithm checks the validity of constraints between affected faces and updates them according to their new condition.

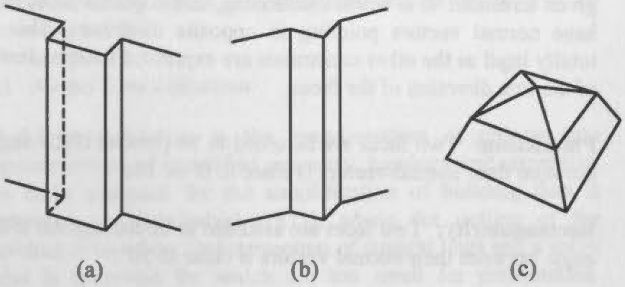


Figure 3. Feature detection distinguishes between face, edge and vertex based features: e.g. (a) extrusion, (b) notch and (c) tip.

5.1 Detecting Extrusions

Currently, our algorithm reliably detects and removes extrusions, which belong to the class of face based features types. Two examples for features that are based on edges and vertices are the notch and respectively the tip. The feature detection algorithm identifies an extrusion if the following two requirements are met for a face. First, the angle between the normal vector of the face and the normal vector of every neighbour face is within some given tolerance of 90° This tolerance angle should not be chosen too high as faces that are used to approximate curved elements of the building model might erroneously be identified as extrusions. Second, all edges of neighbour faces that start or end at the front face must lie behind that face.

5.2 Removing Extrusions

For the removal of an extrusion, we use a combination of two operators, namely the edge collapse operator and the edge foreshortening operator. The edge collapse, or edge contraction, operator deletes an edge and merges its two endpoints into a single vertex (Figure 4). After this operation, the number of edges of the adjacent faces are reduced by one. If the adjacent faces happen to be triangles, they are completely removed from the model. The edge foreshortening operator on the other hand preserves all edges and faces, but moves one of its endpoints along the edge towards the other endpoint in order to shorten the length of the edge.

The removal operator for extrusions performs either an edge collapse or edge foreshortening operation to all edges that emanate from the front face of the feature (as depicted in sketches of Figure 5). Edges that have approximately the same length as the shortest candidate are collapsed into the base vertices (Figure 5b+c). The foreshortening operator must be used for edges that do not completely belong to the extrusion themselves (Figure 5d+e). For this it follows that edges longer than the shortest edge are shortened by this length. If no foreshortening operations are used for long edges, too much of the model geometry gets removed (Figure 5f). In a final step, the original front face of the extrusion is tested for coplanarity to its new neighbour faces and invalidated constraints are removed from the constraint building model.

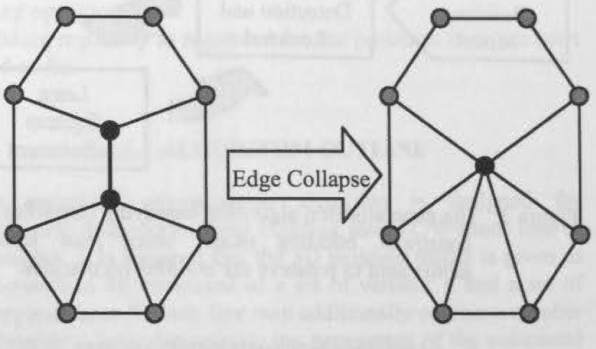


Figure 4. The edge collapse operation contracts the endpoints of the highlighted edge into a single vertex.

The original position of the front face vertices are not just discarded by the removal operator, but they are stored as additional interior points of the face. The next step of the generalisation algorithm, the least squares adjustment, uses these points to determine the new parameters of the planar faces of the building model.

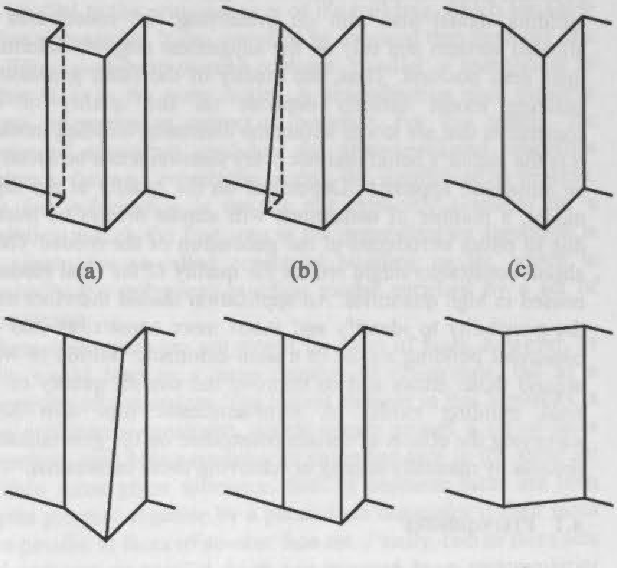


Figure 5. (a-e) Removal of an extrusion: The endpoints of the short edges are collapsed into their base vertices (b+c), whereas the longer edges are foreshortened by the same length (d+e). Collapsing all edges results in the removal of too much geometry (f).

6. LEAST SQUARES ADJUSTMENT

By feature removal, parts of the object are detected and completely eliminated from the dataset. The optimal shape of the reduced model, however, should still be determined by all original points, even though the number of planar faces is reduced by the preceding step. In order to resolve the final shape of the simplified model, a least squares adjustment is applied using the available constraints between the remaining faces as well as the points of the original model. Here the Gauss-Helmert model

$$\mathbf{Bv} + \mathbf{A}\hat{\mathbf{x}} + \mathbf{w} = 0 \qquad (3)$$

is used. During adjustment, the parameters of the buildings' faces as well as their points of intersection are determined. For parameter determination of the faces, the following constraints are applied.

6.1 Coplanarity

The parameters of coplanar faces F_i are determined using the x_k, y_k and z_k coordinates of the original and interior points. Just as 6.2 and 6.3, the coplanarity constraint follows from the constraint building model.

$$F_i : A_i x_k + B_i y_k + C_i z_k + D_i = 0 \qquad (4)$$

Since the definition of the plane parameters for face F_i in (4) is over-parameterised, the norm of the vector A_i, B_i, C_i is integrated as an additional constraint during least squares adjustment.

$$A_i^2 + B_i^2 + C_i^2 = 1 \qquad (5)$$

6.2 Parallelism

For parallel groups of coplanar faces, the parameters A_i, B_i, C_i of the coplanar face group F_i are shared with the other n coplanar face groups parallel to F_i.

$$\begin{aligned} F_i &: A_i x + B_i y + C_i z + D_i = 0 \\ F_n &: A_i x + B_i y + C_i z + D_n = 0 \end{aligned} \qquad (6)$$

6.3 Rectangularity

For two rectangular groups of coplanar or parallel faces, we use the constraint

$$\cos\phi = \frac{A_1 A_2 + B_1 B_2 + C_1 C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2}\sqrt{A_2^2 + B_2^2 + C_2^2}} \qquad (7)$$

As the cosine of 90° is 0, constraint (7) can be simplified to

$$0 = A_1 A_2 + B_1 B_2 + C_1 C_2 \qquad (8)$$

6.4 Point of Intersection

The new position of the remaining vertices of the model are determined by the intersection of three or more faces. In order to provide a complete solution, not only the planar surfaces, but also their points of intersection are integrated into the adjustment. This approach is additionally motivated by the fact, that the topological information about the intersection of the planar surfaces is not yet used. If this information is ignored, four or more planar surfaces are not guaranteed to intersect in one unique point after generalisation. Using a different weight value for each type of constraint, helps to exert influence on the adjustment, e.g. to favour unique intersection points over parallel planar surfaces.

The X_i, Y_i, and Z_i coordinates of the intersection point P_i are determined by using the following constraint for every plane F_k:

$$P_i : A_k X_i + B_k Y_i + C_k Z_i + D_k = 0 \qquad (9)$$

7. RESULTS AND FUTURE WORK

The algorithm above has been implemented and tested on polygonal building models of a 3D city dataset. In order to measure the complexity of each model, we used the number of triangles gained by triangulating the planar surfaces. The algorithm showed promising results on both complex and simple models. The complexity of the building models could in many cases be reduced by over 30%, in some cases, where the model exhibited a lot of extrusions, even by 50%. The model of the New Palace of Stuttgart (Figure 6), e.g., comprises of 721 planar surfaces, that make up a total number of 2730 triangles.

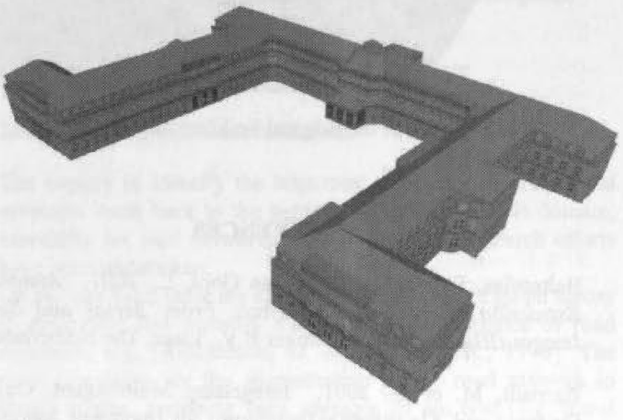


Figure 6. The New Palace of Stuttgart is used to show the results of our generalisation algorithm.

Our generalisation approach was able to detect 110 extrusions using three iterations. After removal of the extrusions, the model only comprised of 1837 triangles. The results are demonstrated in Figure 7 to Figure 10. Figure 7 shows part of the original model as it was captured from stereo imagery and an existing outline from the public Automated Real Estate Map, respectively. Figure 8 shows the result of the generalisation process. As it is visible, parallelism and rectangularity have been preserved for the remaining faces. Using textured models, as it is depicted in Figure 9 and Figure 10, this amount of detail is sufficient for visualisation in most cases.

As the general algorithm design proved to be correct, our future work mainly consists of defining more features types that can be detected and removed. Especially features that are based on edges and vertices have not yet been evaluated. More research has to be put into how complex features need to be dealt with and how curved elements of building models can be simplified.

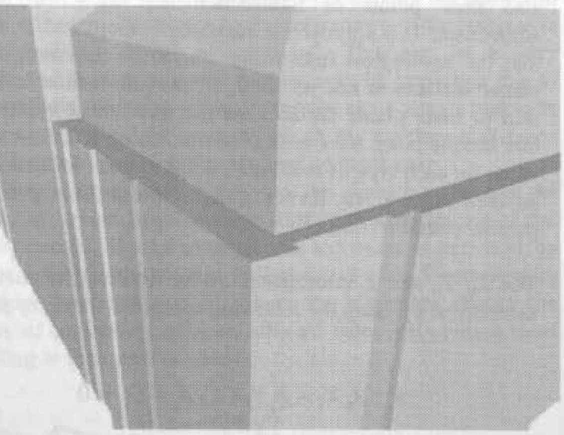


Figure 7. Part of the original building model

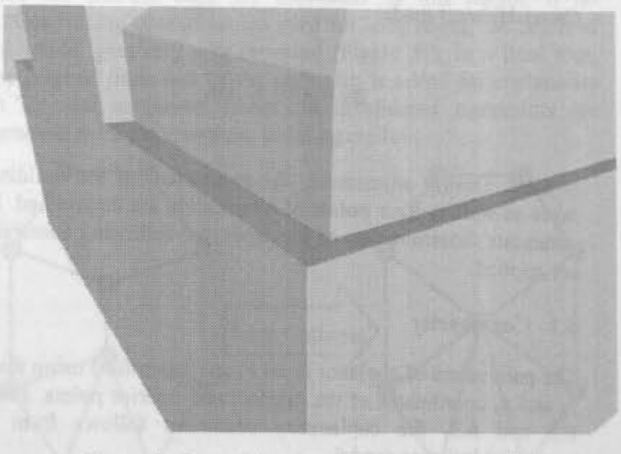


Figure 8. Part of the simplified building model.

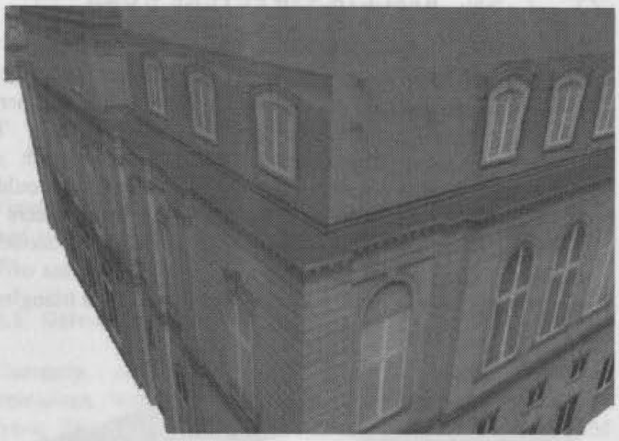


Figure 9. Part of the original building model (textured).

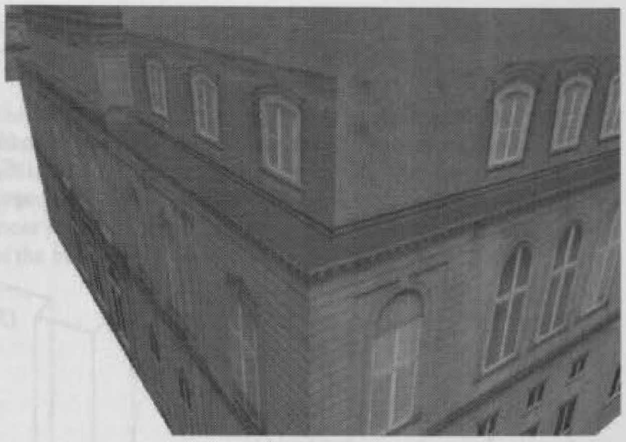


Figure 10. Part of the simplified building model (textured).

REFERENCES

Baltsavias, E., Grün, A. and van Gool, L., 2001. *Automatic Extraction of Man-Made Objects From Aerial and Space Images (III)*. Swets & Zeitlinger B.V., Lisse, The Netherlands.

Barrault, M. et al., 2001, Integrating Multi-Agent, Object-Oriented and Algorithmic Techniques for Improved Automatic Map Generalization. In: *Proceedings of the 20th International Cartographic Conference of the ICA*, BeiJing, China.

Coors, V., 2001, Feature-preserving Simplification in Web-based 3D-GIS, 1st International Symposium on Smart Graphics, 2001, Hawthorne, NY, USA.

Garland, M. and Heckbert, P., 1997, Surface Simplification using Quadric Error Metrics. In: *Proceedings of ACM SIGGRAPH 97*, pp. 206-216.

Grün, A. and Wang, X., 2001, *News from CyberCity-Modeler. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*. Swets & Zeitlinger B.V., Lisse, Netherlands, pp. 93-101.

Haala, N. and Brenner, C., 1999, Virtual City Models from Laser Altimeter and 2D Map Data. In: *Photogrammetric Engineering & Remote Sensing, Vol. 65, no 7*, pp. 787-795.

Heckbert, P. and Garland, M., 1997, Survey of Polygonal Simplification Algorithms. In: *Multiresolution Surface Modeling Course*, SIGGRAPH 97.

Hoppe, H., 1996, Progressive Meshes, In: *Proceedings of ACM SIGGRAPH 96*, pp. 325-334.

Mayer, H., 2000, Scale-Space Events for the Generalization of 3D-Building Data Adjustment. In: *International Archives of Photogrammetry and Remote Sensing, Amsterdam, Netherlands, Vol. XXXIII, Part B4*, pp. 639-646.

Ribelles, J. et al., 2001, Finding and Removing Features from Polyhedra, In: *Proceedings of DETC '01*, Pittsburgh, Pennsylvania, USA.

Rossignac, J. and Borrel, P., 1993, Multi-resolution 3D Approximations for Rendering Complex Scenes. In: *Modeling in Computer Graphics: Methods and Applications*, Springer-Verlag, Berling, pp. 455-465.

Sester, M., 2000, Generalization Based on Least Squares Adjustment. In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Netherlands, Vol. XXXIII, Part B4, pp. 931-938.

Staufenbiel, W., 1973, *Zur Automation der Generalisierung topographischer Karten mit besonderer Berücksichtigung großmaßstäbiger Gebäudedarstellungen*, Ph.D. thesis, Fachrichtung Vermessungswesen, Universität Hannover.

AUTOMATIC DERIVATION OF LOCATION MAPS

B. Elias
Institute for Cartography and Geoinformatics (ikg), University of Hannover,
Appelstr. 9 a, 30167 Hannover, Germany
Birgit.Elias@ikg.uni-hannover.de

Commission IV, WG IV/3

KEY WORDS: GIS, Generalization, Data Interpretation, Data Extraction, Navigation

ABSTRACT:

The paper presents the work in progress to use the digital data of the Authoritative Topographic Cartographic Information System (ATKIS), the geobase information system of the German national mapping agencies, to automatically deduce location maps. It is possible to render location maps in different ways. First of all there is to differentiate between a textual and a graphic kind of route description. The paper focuses on different graphic representations. One possible graphic illustration of the route description can be designed as a hierarchy of map clippings in different levels of details ranging from small to large scale. As another possibility the directions can be visualized as a route corridor from start- to endpoint, where the relevant streets, as well as the important landmarks along the streets are presented.

1. INTRODUCTION

ATKIS (Authoritative Topographic Cartographic Information System) is the digital topographic base data of the German national mapping agencies designed at three different scales. The so called base model has an associated scale of approximately 1:25.000. Besides deriving printed topographic maps from these data sets, they have a great potential for a wide range of applications. The national mapping agencies try to develop more useful applications to benefit from this digital data base in order to make it more profitable.

One possible application is the generation of individual location maps. Currently, there is an existing service, a mixture of using the digital data base with a major part of additional made by hand work to create an acceptable cartographic result. Because of the great financial expense it will be very useful to derive this product as a whole automatically from the ATKIS database.

Different possibilities to create a location map exist. A first schedule line is to differ between a textual and a graphic portrayal. The paper focuses on two different graphic representations. The graphic illustration of the route description can be portrayed as a hierarchy of map clippings in diverse levels of details, for example two different scales: a coarse level of detail to show an overview of the urban area and a fine scale map to portray the road network (similar to a city map) including the destination. Depending on the location an extension with a third level of scale is required and displays a very detailed clipping to outline the immediate neighbourhood. The coarsest level should only include the main traffic axes of the city, this requires that the classification of the roads has to be part of the ATKIS database.

In this paper the important structures of urban road networks are identified based on a geometric-topologic analysis. The content of the ATKIS data concept is investigated regarding this structures. Also methods for the automatic derivation of these relevant streets are studied.

As a second possibility the driving instruction can be interpreted as a route corridor from start- to endpoint, where the relevant streets as well as the important landmarks along the streets are presented. The automatic generation of such a presentation requires first of all the navigational information, then it is necessary to identify the important landmarks along the route, especially in the vicinity of turning points or junctions. Such points of reorientation are bridges, rails, prominent buildings or vegetation like a big tree or a park.

The paper describes the kinds of landmarks, which are extracted from the ATKIS database. Furthermore the integration of landmarks from other digital data bases will be investigated.

2. RELATED WORK

2.1 Road Network Generalization

The inquiry to identify the important structures of urban road networks leads back to the generalization task. In this domain, especially for road networks, a few number of research efforts have been undertaken.

On the one hand there are approaches dealing with graph theory to determine and quantify the functional importance of road segments, e.g. [Richardson, D. & Thomson, R., 1996]. The main principles are the segmentation of the road network in single graphs, assigning each segment an arc cost (e.g. travel time or distance), and the calculation of the minimum cost spanning tree for the network. The results are derived arc weights giving a partial ordering of the arcs which can be used as a basis for the attenuation of the network in generalization [Thomson, R. & Richardson, D., 1995]. Because of the usage of the minimum weight spanning tree, that means a connected graph links all nodes by using the least number of edges with the most important weight, the connectivity of the network is provided [Mackaness, W. & Beard, K. 1993].

A second kind of approach for network generalization is done by [Kreveld, M. van & Peschier, J., 1998]. The graph theory approach is criticised since the geometric aspects of coalescence and imperceptibility of the portrayed elements as well as the semantic aspects such as avoiding large detours are not taken into account explicitly. The outlined ideas are implemented partly in a data set and can be run via the World Wide Web [Peschier, J., 1997].

A further procedure is based on the theory of perceptual organization, particularly on one of the gestalt laws: the 'good continuation' grouping principle [Thomson, R. & Richardson, D., 1999]. For that purpose so called 'strokes' are built from the arc segments of the network presenting contiguous linear elements perceived from human cognition. It is asserted that an general correspondence between the perceptual salience of strokes and their functional importance in the network exists. A ranking of the stroke length information leads to an order of importance serving the attenuation of the network through selection.

The connection between an percental network attenuation and the required map scale is given through the principles of selection theory [Töpfer, F. & Pillewizer, W., 1966].

2.2 Route Corridors

The problem to give someone route instructions for getting from start point A to an endpoint B leads to the possibility to convey the navigational information by route directions. These guiding information can be given as a description (textual statement) or a depiction (as a kind of graphic represented route corridor), both seem adequate to convey sufficient information for arriving at a destination [Tversky, B. & Lee, P., 1998].

In [Tversky, B. & Lee, P., 1999] it is stated that the structure of route maps is essentially the same as the structure of route directions. Moreover, even the semantic content is similar. The components of both are landmarks, orientations and actions.

[Agrawala, M & Stolte, C., 2000] remark that the World Wide Web online mapping services typically provide directions as a set of maps complemented with text descriptions. But these portrays do not represent a plain route instruction, because they ignore important design goals for effective route maps: readability, clarity, completeness and convenience. The authors have designed and implemented some kind of computer-generated maps that mimic the style of hand-drawn route maps in order to achieve an effective compromise of the required design goals [Agrawala, M.& Stolte, C., 2001].

The approach lacks the existence of landmarks in the route maps, for example in long-distance directions even the geographic names of the cities are missing. Experiments have shown that people react to the absence of landmarks. The reason is that landmarks are essentially used as sub-goals along the route: people progress along a route by orientating themselves towards a landmark [Michon, P.-E. & Denis, M., 2001].

In a further experiment (see [Lovelace, K., Hegarty M. & Montello, D., 1999]) landmarks were classified in four different types: landmarks at a choice point, potential (but not used) landmarks at choice points, on route (along the path) landmarks and off-route landmarks (not neighboured to the followed path, but with some orientation value). The research indicates that the appearance of landmarks correlates significantly with quality of route directions. Especially for unfamiliar route directions landmarks at turning points and just on-route points are quite frequently used.

The usage of landmarks is so important because they serve multiple purposes in wayfinding: they help to organize space because they are reference points in the environment and they support the navigation by identifying choice points, where a navigational decision has to be make [Golledge, R., 1999]. Landmarks are characterized by being external to the observer and serving to define the location of other objects or locations. They may have particular visual characteristics, a unique purpose or meaning or may be in a central or prominent location that makes them effective as a landmark [Sorrows, M. & Hirtle S., 1999].

3. INTRODUCTION OF THE ATKIS – CONCEPT

The Authoritative Topographic Cartographic Information System called ATKIS is a joint project of all German national mapping agencies and can be described as a geobase information system. Within the framework of ATKIS several different data types exist: the digital topographic maps, digital terrain models, digital orthophotos and the digital landscape model, which is used here [ATKIS – product description, 2002].

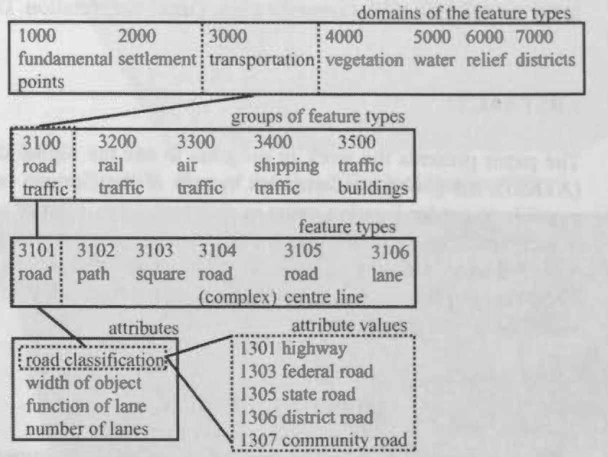


Figure 1. The ATKIS concept

The model illustrated in Figure 1 structures the space according to objects. It differentiates seven 'domains of feature types': fundamental points, settlement, transportation, vegetation, water, relief and districts. Each class is subdivided in 'groups of feature types' and further in 'feature types', marked by a number code. At the level of 'feature types' special thematic attributes are attached to the objects, e.g., for roads there are among other things the attribute 'road classification' with the possible attribute values: highway, federal road, state road, district road and community road.

The data are stored in vector data structure in the national grid reference system and have an accuracy of 3 metres for major feature types like roads. The feature types and attributes in the feature catalogue will be filled in continuously each (periodic) maintenance of the data.

A complete description of the concept is given in the object catalogue [ATKIS-OK, 2001].

4. LOCATION MAPS

The generation of location maps in form of a map hierarchy involves deriving different levels of detail from one single database. This study focuses on the representation of the road network, because it is the most important part of the data base to convey the navigational information needed in a location map. The used ATKIS data base contains the complete existing road network and bases on the scale 1:25.000. Therefore the fine scale of the location map hierarchy is a representation of the ATKIS data content. An extension in a larger scale representation (e.g. 1:10.000) for more clarity and readability is possible.

To provide a coarser scale the importance of the roads has to be modelled in the ATKIS concept in order to automatically select the elements, which are necessary to portray according to their outstanding functional meaning. Inspecting the ATKIS

catalogue reveals different possibilities to assign importance to a road: There is an attribute called 'function of traffic', with the distinctions 'long-distance traffic' or 'transit traffic' for example, but it is not included yet.
On the other hand for the feature type 'roads' there is the attribute 'road classification' (see Figure 1) and in fact, the attribute values are attached to each road segment. A visualization of all roads classified higher than a 'community road' and in addition all road segments with more than one lane for each driving direction shows an extremely reduced but in fact characteristic representation of the road network.
As an example the combination with a layer presenting the urban area of the small town Aurich produces a visualization of a coarser scale for the location map. This example is shown in Figure 2. The automatic upgrading with the classification names of the roads is possible. In general, in an overview map of a location map all important means of travel have to be regarded, that means arriving by car, by train or even by aeroplane have to be taken into account. In this example the town is too small for a train station or airport and so the road network is the most important navigational information for the map.
For comparison a topographic map of this area is given in Figure 3, in which the important transit roads are emphasized with cartographic methods.
If it is required to generate overview maps with a more detailed network, the task reduces to the generalization problem. The creation of an coarser scale from the ATKIS data leads to the idea to attenuate the network relying on geometric-topologic properties in order to extract the functional important roads only.

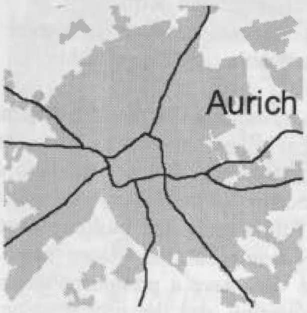


Figure 2: ATKIS data, higher classified roads only and urban area layer (in grey)

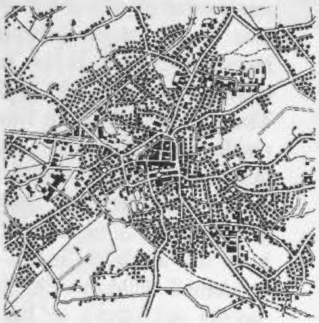


Figure 3: Small town Aurich, topographic map, scale 1:100.000

The approach tested here is the 'good continuation' grouping principle. The implementation from [Sester, M., 2000] is used to calculate the data. As described above in this approach a importance ranking of strokes is established and used to select the required segments by cutting off a percental part. Figure 4 shows the complete ATKIS data network for Aurich. In Figure 5–7 the results of several attenuated versions, that are calculated each with a different percent of data, are given. They show the continuous reduction of the network in 10%-steps. A good characteristic of the approach is the elimination of short road segments to simplify the whole network. Though not specially weighted in the calculation the higher classified roads remain in the attenuated networks as the underlying principle prioritizes long and straight lines. Besides clearly visible structures like round shapes (see the arrows in Figure 7, for example the semi arc in the down right quarter and the elliptical curve left from the transit road directing north) are retained.

Disadvantages of the approach (and implementation) are that some small irregular structures are formed, for example in the top right quarter a small angled shape is visualized (Figure 7, the arrow directing left). As a general deficiency of the approach, long roads that are very close to each other are retained and form unintentional parallel structures. Furthermore disconnections within the reduced network are possible. Therefore a completely automatic generalization with the aim of a high quality cartographic result is not achievable.

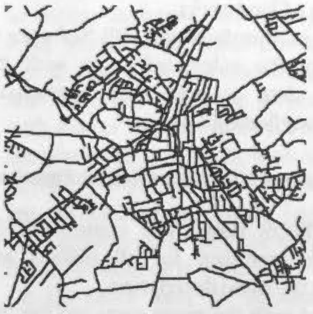


Figure 4: Complete ATKIS road network

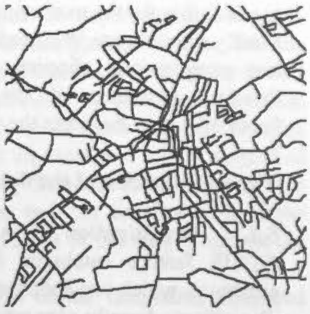


Figure 5: Road network, 10% reduced

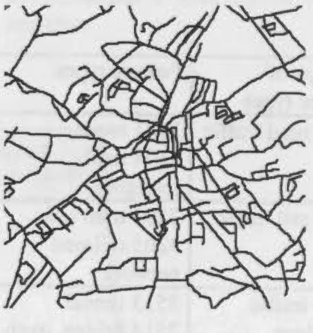


Figure 6: Road network, 20% reduced

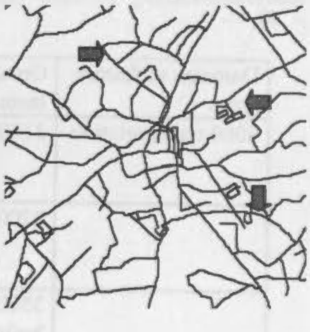


Figure 7: Road network, 30% reduced

But the comparison of Figure 7 with the topographic map in scale 1:100.000 (see Figure 2) shows on the one hand that a 30% percent attenuation of the network corresponds with the content of a map scaled in 1:100.000 and on the other hand that the approach works at least satisfying: nearly the same selection of streets is visualized in both maps. So it is to state that the approach of the 'good continuation' principle leads to an acceptable result to create location maps.

5. ROUTE CORRIDORS

The second way to give route directions is to generate route corridors. In these versions only the desired navigational information including additional landmarks are given. Again the data used are ATKIS data.

5.1 The approach for generating route corridors

For automatic derivation of a route corridor from the ATKIS data base the data structure must be suited for routing tasks. First of all a road line network dataset is required. Then special traffic information like one-way streets or allowed driving directions on intersections have to be given. Unfortunately ATKIS data do not meet these criteria. For example the structure of roads is not closed: in general the roads are stored as linear features, but it can occur that squares are defined as polygons and thus destroy the topology of the network. Furthermore actual traffic sign information are not taken into account. So the investigation has to start with a routing result, which is corrected by hand.

The next step is to identify the potential landmarks and to extract them from the data. Before this, the ATKIS concept is examined with respect to the object classes and their attributes to identify all information that can be used as landmarks. Further data sources are investigated concerning their applicability for the extraction of landmarks.

Finally the data are imported in a professional GIS-Software to test merging and clipping of the calculated route with the thematic attributes of the data to designate appropriate landmarks, that facilitate the wayfinding task.

5.2 Examination of the data bases for potential landmarks

5.2.1 Investigation of ATKIS data: The content of the ATKIS feature catalogue is examined for all objects and attributes, which can enhance the route description.

The objects directly connected with the given route are the first group to recover, that means feature types crossing the roads of the route. Analysing the data reveals some typical classes of objects, which often intersect with roads and paths (see Table 8).

Domains of feature types	Groups of feature types	Feature types
3000 transportation	3100 road traffic	3101 roads 3104 roads (complex)
	3200 rail traffic	3201 tram rails 3205 railroad network
	3500 traffic buildings	3513 tunnel 3514 bridge, over-/underpass
5000 water	5100 water area	5101 stream, river, creek 5102 canal

Table 8. Route crossing objects

Domains of feature types	Groups of feature types	Feature types
2000 settlement	2100 built-up area	2122 landfill 2126 powerhouse 2127 transformer stat. 2133 heat plant
	2200 clearway	2201 sports facilities 2213 cemetery 2221 stadium 2224 swimming pool 2225 zoo 2226 public park
	2300 buildings and further facilities	2316 tower 2317 chimney 2327 pin wheel 2332 monument 2351 wall
5000 water	5100 water area	5111 sea 5112 lake, pond

Table 9. Selection of landmarks besides the route

The next group of potential landmarks comprises all objects, in ATKIS mostly area features, which lie directly next to the route. After a first investigation of the feature catalogue the listed feature types in Table 9 were chosen to be potential candidates for landmarks.

In ATKIS the features are assigned to the different feature types by use of the four digit number code, see Table 8 and 9. The substitution of the feature number code with the text notation is possible. For each feature type an attribute 'geographic name' and where appropriate 'shortcut name' is planned. (Whether the attribute value is set or not depends on the acquisition date of the data set.) But for every record with a geographic name (and/or shortcut name) an automatic labelling is feasible.

A striking characteristics of the ATKIS data is the lack of the representation of single buildings. Although a feature type 'buildings' is planned, until now this information is not yet included. To overcome this difficulty and to extract buildings for supplying landmarks a second authoritative data source is used: the cadastral map.

5.2.2 Investigation of cadastral data: The digital cadastral map (in Germany called ALK) gives evidence about all parcels of land, the proof of ownership and administrative units. Furthermore the illustrative part of the real estate cadastre serves the purposes of all kind of planning, so it is constituted by law that the buildings belong to it. The digital version of the cadastre map (in former times on paper) is under construction, but in most parts of Germany the map in scale 1:1000 is already completed. Similar to the ATKIS concept the content of the cadastral map is based on a feature catalogue in which the represented objects and their signatures are defined.

Due to the object oriented representation an extraction of individual buildings is practicable and is possible to be combined with the ATKIS data.

There are four different types of buildings in the cadastral map: residential buildings, outbuildings (including industrial buildings), underground buildings (e.g. subway stations) and public buildings. This classification of buildings alone will not generate useful landmarks. Additionally the building objects have the attribute 'name' standing for the proper name of residential buildings, in case they have a prominent meaning (occurs not often), or the function and the proper name of public buildings.

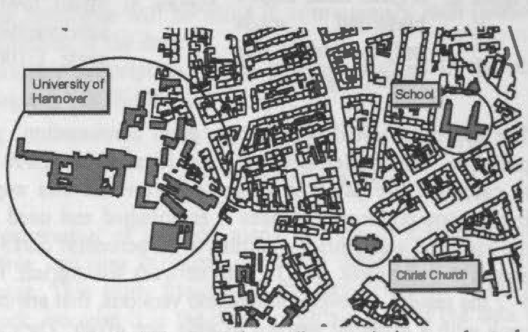


Figure 10. Public buildings (in grey) in cadastral map; clipping in scale 1:5000

Figure 10. visualizes a clipping from the cadastral map in order to give an impression of the frequency of public buildings. For some instances the name is added to point out what kind of buildings and names can be extracted. The examples given here are appropriate as landmarks: they are outstanding in their surroundings because of their size and architecture. Possibly, they also appear on traffic signs or decal information because of their functional meaning, so they can be easily identified on the route.

5.3 Exemplary Presentation

To gain insight of such a route corridor enriched with landmarks all data sets are imported in a commercial GIS-Software (here: ArcView 3.2 from ESRI).

In the first step the route is generated together with the street names and distances between the turning points. The start- and the endpoint are marked with labels and the route is depicted as a simple linear feature.

After that a buffer around the complete route is built. In this case study a 20 metres distance buffer is chosen as a practical compromise and then clipped step by step with the selected feature types. The buffer is defined by empiric tests: the width of the buffer has to be wide enough to reach over the whole street and contact the first row of objects next to the route. It must be pointed out that it has to work not only for narrow streets in housing areas but also for main traffic axis with more than one lane for each driving direction. On the other hand the distance buffer has to be narrow enough to avoid the clipping of objects which are not lying next to the route but in second row and so possibly are not directly visible from the route.

The chosen 20 metres buffer seems to be adequate for this case study. An approach for future work is the investigation of how to calculate the buffer width depending on the classification of the underlying road. Another option is to use a Delaunay triangulation and thresholds for the distances to determine the neighbourhood of a road

After buffering the extracted objects are automatically labelled by their names. The result is shown in Figure 12. For comparison in Figure 11 the complete topographic map of that area overlaid with the route is pictured.

Because of the human cognition abilities the simplification of the calculated route corridor is possible and makes the visualization more precise. For the navigation it is not necessary to portray authentic turning angles. To depict them as a nearly right angle is sufficient for the navigation because for the human cognition it is interesting if there is a obvious turn at all and remember it later as perpendicular (in case it is a junction with four or less directions). Vice versa, roads with small curves are perceived as straight without any turn and so are simplified in the portrayal. Furthermore long distances following one road are experienced as straight lines and can be shortened in the visualization since only the proportion of the ranges but not the absolute distance is realized.

On the other hand the signatures of the landmarks produced automatically have to be revised cartographically. Shape, size and position of the landmarks themselves as well their text labels have to be generalized and made clearly legible. Figure 13 visualizes a first try of this kind of generalized route corridor.

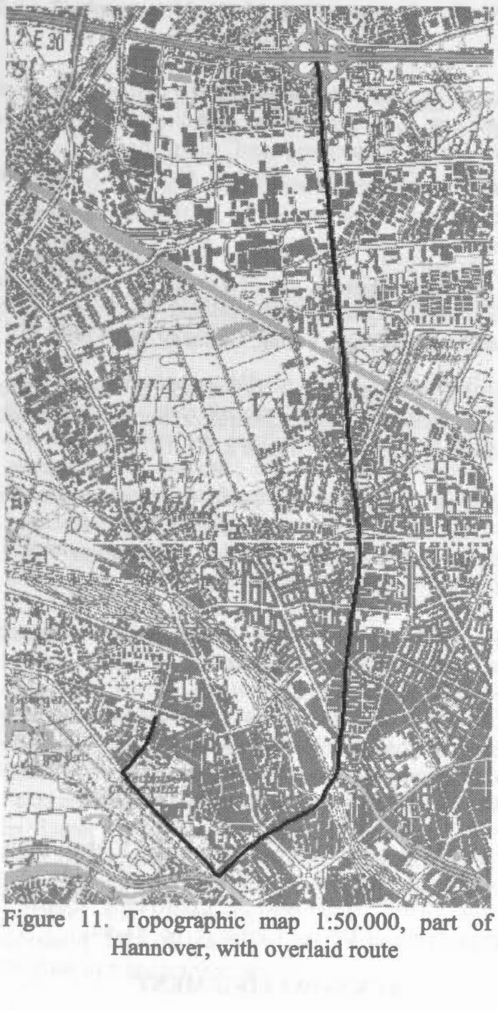


Figure 11. Topographic map 1:50.000, part of Hannover, with overlaid route

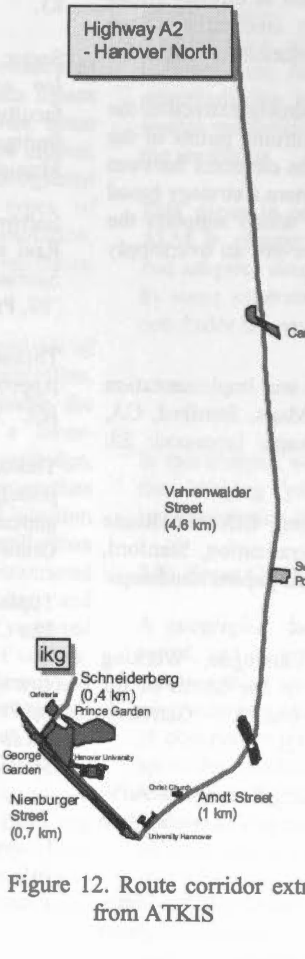

Highway A2
- Hanover North
Canal
Vahrenwalder Street (4,6 km)
Swimming Pool
ikg
Schneiderberg (0,4 km)
Cafeteria
Prince Garden
George Garden
Hanover University
Nienburger Street (0,7 km)
Christ Church
Arndt Street (1 km)
University Hannover

Figure 12. Route corridor extracted from ATKIS

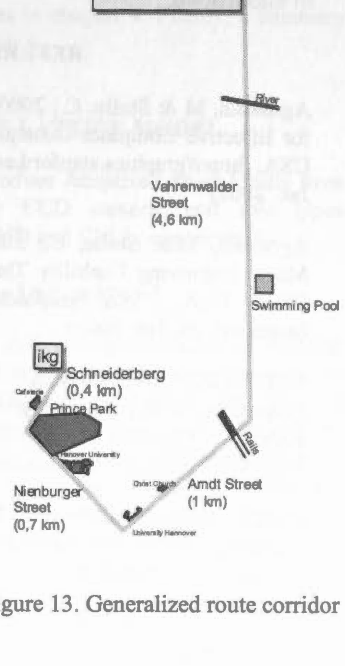

Highway A2
- Hanover North
River
Vahrenwalder Street (4,6 km)
Swimming Pool
ikg
Schneiderberg (0,4 km)
Cafeteria
Prince Park
Hanover University
Rails
Nienburger Street (0,7 km)
Christ Church
Arndt Street (1 km)
University Hannover

Figure 13. Generalized route corridor

6. CONCLUSIONS AND FUTURE WORK

This paper demonstrates two different possibilities to generate location maps using the ATKIS data base.

The first way generates a hierarchy of maps consisting of two or more levels of detail and scale. The finer scale including the location and its surrounding streets is a representation of the ATKIS content. The coarser scale for the overview map has to be an extract from the ATKIS data. Therefore an attenuation of the network is needed and leads back to the general problem of network generalization. In a first test the 'good continuation' grouping principle is used and the results show that it works satisfying. The automated generation of a coarser scale consisting of the important streets and most of the characteristic structures of the network is practicable.

As a second way to generate a location map the creation of a route corridor in combination with landmarks as a guiding assistance is introduced. Therefore the extraction of landmarks from the ATKIS data base is tested and evaluated to be effective. It is shown that the numerous feature types and attributes pertaining to the data are adequate for producing landmarks. Data integration of additional data sets (here the cadastral map) is a promising chance to successfully enrich route maps with adequate landmarks. An automatic derivation is technical practicable and part of future work.

Next steps are the detailed segmentation of the feature catalogue to derive an ordering of data clipping. Then it has to be investigated, if different classes of landmarks can be distinguished. A possibility is the distinction in directly (from the route) visible landmarks and indirect landmarks whose existence is reasoned from warning and information signs for example.

To control the quantity of results of the landmark extraction the idea of buffering first the area around the turning points of the routes and then if still necessary the straight elements between them will be investigated in detail. Furthermore a strategy based on information theory will be developed, which supports the selection of the important landmarks to prevent an oversupply of information.

REFERENCES

Agrawala, M & Stolte, C., 2000. A Design and Implementation for Effective Computer-Generated Route Maps, Stanford, CA, USA, http://graphics.stanford.edu/papers/maps/ (accessed: 28. Jan. 2002)

Agrawala, M.& Stolte, C., 2001. Rendering Effective Route Maps: Improving Usability Through Generalization, Stanford, CA, USA, http://graphics.stanford.edu/papers/routemaps/ (accessed: 28. Jan. 2002)

ATKIS-OK, 2001. ATKIS – Feature Catalogue, Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV), Germany, http://www.atkis.de (accessed: 20. Feb. 2002)

ATKIS – product description, 2002. Working Committee AdV, Germany, http://www.adv-online.de/english/products/atkis.htm (accessed: 4. March 2002)

Golledge, R., 1999. Human Wayfinding and Cognitive Maps, In: Golledge, R., Edt., *Wayfinding Behavior*, The John Hopkins University Press, Baltimore and London, pp. 5-46.

Kreveld, M. van & Peschier, J., 1998. On the Automated Generalization of Road Network Maps. In: *Proceedings 3rd International Conference on GeoComputation* (CD-Rom). http://divcom.otago.ac.nz/sirc/webpages/Conferences/GeoComp/GeoComp98/geocomp98.htm (accessed 28. Feb. 2002)

Lovelace, K., Hegarty M. & Montello, D., 1999. Elements of Good Route Directions in Familiar and Unfamiliar Environments, In: Freksa, C. & Mark, D., Eds., *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science*, International Conference COSIT '99, Proceedings, Springer Verlag, Germany, pp. 65-82.

Mackaness, W. & Beard, K. 1993. Use of Graph Theory to Support Map Generalization. *Cartographic and Geographic Information Systems*, 20 (4), pp. 210-221.

Michon, P.-E. & Denis, M., 2001. When and Why Are Visual Landmarks Used in Giving Directions, In: Montello, D., Ed., *Spatial Information Theory*, International Conference COSIT 2001, Proceedings, Springer Verlag, Germany, pp. 292-205.

Peschier, J., 1997. Computer Aided Generalization of Road Network Maps, M. Sc. Thesis, Utrecht University, Department of Computer Science, http://www.jarno.demon.nl/thesis.htm (accessed 28. Feb. 2002).

Richardson, D. & Thomson, R., 1996. Integrating Thematic, Geometric, and Topologic Information in the Generalization of Road Networks, *Cartographica*, 33 (1), Monograph 47, pp.75-83.

Sester, M., 2000. *Maßstabsabhängige Darstellungen in digitalen räumlichen Datenbeständen*, Professorial Dissertation, faculty of civil engineering and surveying, University of Stuttgart, series C, Nr. 544, German Geodetic Commission, Munich.

Sorrows, M. & Hirtle S., 1999. The Nature of Landmarks for Real and Electronic Spaces, In: Freksa, C. & Mark, D., Eds., *Spatial Information Theory*, International Conference COSIT '99, Proceedings, Springer Verlag, Germany, pp. 37-50.

Thomson, R. & Richardson, D., 1995. A Graph Theory Approach to Road Network Generalization, In: *Proceedings ICC '95 Barcelona*, Spain, Volume 2, pp. 1871-1880.

Thomson, R. & Richardson, D., 1999. The 'good continuation' principle of perceptual organization applied to the generalization of road networks. In: *Proceedings ICA '99*, Ottawa, Canada, Session 47 B.

Töpfer, F. & Pillewizer, W., 1966. The Principles of Selection, *The Cartographic Journal*, 3, pp. 10-16.

Tversky, B. & Lee, P., 1998. How Language Structures Space, In: Freksa, C., Habel, C.& Wender, K., Eds., *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, Springer Verlag, pp. 157-175.

Tversky, B. & Lee, P., 1999. Pictorial and Verbal Tools for Conveying Routes, In: Freksa, C. & Mark, D., Eds., *Spatial Information Theory*, International Conference COSIT '99, Proceedings, Springer Verlag, Germany, pp. 51-64.

ACKNOWLEDGEMENT

Many thanks to the National Mapping Agency of the state of Lower Saxony (called LGN) for providing the ATKIS data.

SPATIAL GENERALIZATION: AN ADAPTIVE LATTICE MODEL BASED ON SPATIAL RESOLUTION

P. Wang [a],*, T. Doihara [a], W. Lu [a]
[a] Research Institute, Asia Air Survey, Tamura 8-10, Atsugi, Kanagawa, JAPAN
(pt.wang, ta.Doihara, luwei)@ajiko.co.jp

Commission IV, Working Group IV/3

KEY WORDS: Generalization, Clustering, Aggregation, GIS - Geographic Data, Simplification, Adaptive Lattice Model

ABSTRACT:

The paper presents an Adaptive Lattice Model (ALM) based on spatial constraints or resolution such as minimum object size and minimum spatial interval. An ALM represents GIS data by latticing adaptively the spatial space at a certain abstraction level or spatial resolution and omitting all details smaller or finer than the resolution. One implementation of ALM is called CELL, which considers each lattice as a cell whose characteristics are statistically measured as a whole. And another one is called GRID, which moves all vertexes of the objects to the nearest point or one of the corners of lattices in ALM. This paper concentrates on the CELL based generalization including adaptive clustering, adaptive aggregation, and adaptive simplification. Adaptive clustering uses the target resolution to decide which features of the same type can be clustered or not. Adaptive aggregation, following the adaptive clustering, aggregates features in a cluster to form a new feature with the same type. And, adaptive simplification merges adjoining vertexes and deletes small loops and kickbacks by rounding up all vertexes of linear features in the input dataset. Experimental results are also presented to verify the effectiveness of the proposed model.

1. INTRODUCTION

A geographic dataset is a resolution-dependant representation of the original data, and generalization of geographic data means changing the representation resolution. In other words, some features or their attributes in the original dataset may be hidden, aggregated, simplified, or deleted, depending on the designated or target resolution. Although there are three types of resolutions, including spatial resolution, thematic resolution, and temporal resolution (Peng 2000), only spatial resolution will be discussed in this paper.

Map generalization usually involves a great deal of analysis of the geographic data for deciding what and how to generalize, and how to resolve conflicts that might occur during the generalization process. Manual generalization is a labor-intensive process and depends mainly on operator's knowledge, experience and skills. Recently, there many researches regarding general problems about generalization (Johnston 1999; Peng 2000) and concrete operations of generalization (Kreveld 1998; Sester 2000) for building automated generalizing algorithms to shorten generalization time and relieve operators from hard decision-making and repeated operations. For geographic dataset mainly consisted of vectors, polyline generalization is one of the most widely studied problems in the cartographic literature (Johnston 1999). An effective polyline generalization method, called Douglas-Peucker-Ramer algorithm (DPR) and proposed independently by Ramer (1972) and Douglas and Peucker (1973), is widely used. DPR creates a new generalized polyline by calculating the perpendicular distance between any vertex and the result polyline, and comparing the distance with a predefined threshold. DPR is based on the geometric distortion of every polyline in a geographic dataset.

This paper proposes an Adaptive Lattice Model (ALM) based on spatial constraints or resolution, which focuses on spatial generalization including clustering, aggregation, and simplification of geographic data such as roads, rivers and buildings, etc. An ALM represents geographic data by latticing adaptively the spatial space at a certain abstraction level or spatial resolution and omitting all details smaller or finer than the resolution.

This paper is organized as follows: after the explanation of ALM in chapter 2, adaptive clustering, adaptive aggregation, and adaptive simplification are discussed in chapter 3, followed by some experimental results in chapter 4. Finally, a summary concludes the paper in chapter 5.

2. ADAPTIVE LATTICE MODEL

In this chapter, we shall introduce Adaptive Lattice Model from the imaging principal by CCD camera, and two types implementation of ALM, GRID and CELL, consequently.

2.1 From CCD to Adaptive Lattice Model

A geographic dataset is a kind of representation of the real world, and its generalization is nothing more than changing its representation resolution. Although geographic data may be represented on maps with a range of different scales, the validity of observations or inferences drawn from such maps depends upon the resolution of the representation and the observer's resolution. Considering the representation of a geographic dataset as a imaging system with a CCD camera, what the observer can see are pixels arranged in two-dimension, where each pixel corresponds to and represents statistically a region of the geographic data. Figure 1 shows a typical example. Here, the left graph is the original data latticed with a higher resolution (smaller interval), and the right one shows the same data latticed with a lower resolution (larger interval). Considering each rectangle as a cell of a CCD, an observer cannot identify the details within a rectangle. That is to say, an observer can only take a cell as the overall representation of the corresponding region of the geographic data. Similar to the intensity of cells of the CCD device, the overall representation may be whether there are any vertexes in the cell, or whether there are any segments crossing the cell.

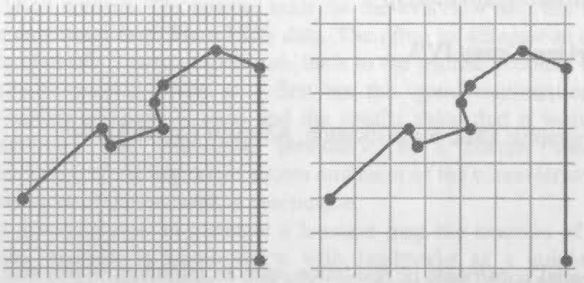


Figure 1. Lattice model for generalization

From the similarity between a CCD camera and the latticed geographic data, we define Adaptive Lattice Model (ALM) as follows: when a geographic dataset is confined to a predefined lattices where every edge's length of cells is equal to the target resolution (or spatial interval), a generalized result of the geographic dataset will be obtained by calculating the overall representation of the corresponding region of the data for every cell. For example, we can check if there are any vertexes of features in a rectangle, or calculate the overlapping area's percentage between a cell and a face feature in the geographic dataset. Different calculating methods may be used to generalize various types of geographic data. Here, two kinds of implementations of ALM are discussed below.

2.2 Implementation of ALM: GRID

In ALM, we can measure a cell by counting the number of vertexes in the cell. All vertexes in the same cell cannot be distinguished from each other, because the distance between any two of those vertexes is smaller than the target resolution, and should be represented by only one vertex in accordance with the principle of generalization. To merge all vertexes in the same cell to one point, moving those vertexes to the centre of the cell is a straightforward solution. In this way, the coordinates of all vertexes of features in the input dataset are rounded up with the cell's size or target resolution, similar to positioning graphs onto grids in a word processor. For this reason, the implementation of ALM mentioned here is called GRID. Since the translation of vertexes is within the cell, the statistic characteristics of the cell are not changed.

With GRID, the generalization process can be simplified to rounding up all vertexes' coordinates to the integral multiple of the target resolution and analysing the relationship among all the resulted vertexes. For example, a pair of adjacent vertexes will be merged if their rounded coordinates are the same as each other; a linear feature may be split if a middle vertex of the feature is merged with a vertex of another feature within the rounding procedure; etc. The algorithms based on GRID are discussed in Doihara 2002.

2.3 Implementation of ALM: CELL

In comparison to merging all vertexes in the same cell by GRID implementation, calculating the overlapping ratio between a cell and the portions of features that fall within the cell is another representation of the cell. GRID considers only vertexes of features and is effective for simplifying linear features. On the other hand, overlapping ratio should be defined as the percentage of the overlapping area between the cell and the portion of the features contained in the cell's area. That is to say, the representation of overlapping ratio may be used to generalize face features, and is called CELL to be distinguished with GRID. CELL can be used to aggregate face features by analysing the relationship of adjoined features. More details are discussed in the following chapter.

3. GENERALIZATION BASED ON ALM

Automated generalization usually focuses on geometrical features such as buildings, rivers, roads, etc. These features are commonly represented by polylines and/or polygons. Figure 2 shows a processing flow for generalizing geographic data by using ALM. Firstly, Selection is used to select features to be generalized. Continuously, we have two processing routes which are designed for selected areal and linear features separately. One is simplification based on GRID for generalizing linear features, which is discussed in another paper (Doihara 2002). Another is designed for generalizing areal features, including clustering and aggregation. We shall show the details of every procedure below.

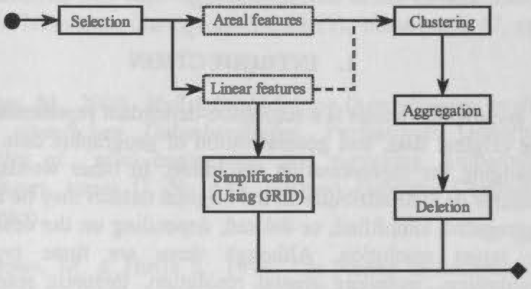


Figure 2. Generalizing Flow with ALM

3.1 Selection

Selection procedure is introduced to filter features to be generalized or not. With the decrease of the resolution, selection procedure ignores features to be omitted and reserves features to be generalized. Selection procedure can also be used to restrict the features to be generalized at the same time for avoiding erroneous data. For example, roads and buildings may be generalized simultaneously, because roads cannot cross with buildings at any target resolution. On the other hand, land boundaries may cross with some roads. In this case, selection procedure let the roads and buildings be generalized in the first phase, and the land boundaries at the next phase. Selection is commonly a manual procedure or a procedure based on predefined database, and will not be discussed more in this paper.

3.2 Adaptive Clustering

Adaptive clustering uses the minimum spatial interval or target resolution to decide which features of the same type can be clustered or not. When linear and face features are selected by the selection procedure, and need to be generalized, the clustering procedure is shown in figure 3. Here, linear features are used as constraints for separating face features compulsorily when necessary, e.g. when two buildings located at the different sides of a main road cannot be merged together. And face features are the targets of clustering, which results in lists of face feature. Some details of figure 3 are given below.

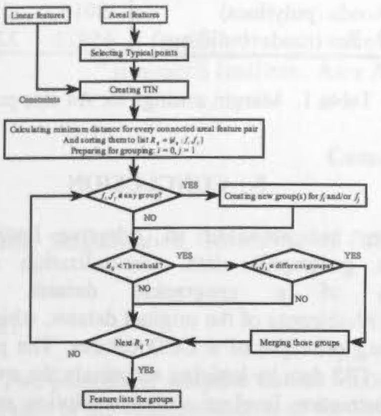

Linear features
Areal features
Selecting Typical points
Creating TIN
Constraints
Calculating minimum distance for every connected areal feature pair
And sorting them to list $R_{ij} = \{d_{ij}, f_i, f_j\}$
Preparing for grouping: $i = 0, j = 1$
$f_i, f_j \notin$ any group?
YES
Creating new group(s) for f_i and/or f_j
NO
d_{ij} < Threshold?
YES
$f_i, f_j \in$ different groups?
YES
NO
NO
Next R_{ij} ?
Merging those groups
NO
YES
Feature lists for groups

Figure 3. Adaptive clustering diagram

a) Selecting representative points for all face features includes several instances. One is using the centre point of a face feature. The next one is treating all vertexes of the outline as a face feature's representative points. The third one is to split a face feature to Triangulated Irregular Network (TIN) and use the centres of all triangles as the feature's representative points.
b) When creating TIN from all representative points, the outlines of face features and all linear features are used as the constraints or break lines to avoid unfavorable clustering.
c) A link is an edge of a triangle in TIN, and must connect two different face features. If there are multiple links between two face features, only one is reserved for later process and others are ignored, when calculating the minimum distances for all links. Links connecting the same features are ignored in this implementation.
d) After sorting all links with their minimum distance, the shortest link should be at the front, and the longest one at the rear. The clustering procedure must begin from the first (shortest) link.

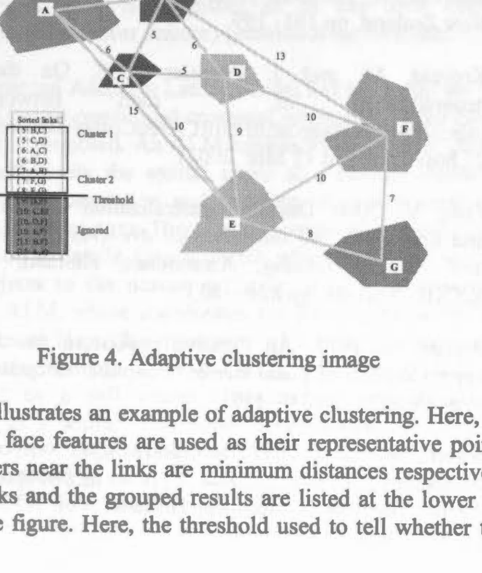

A
B
C
D
E
F
G
7
5
6
6
5
13
15
10
9
10
10
7
8
Sorted links
(5: B,C)
(5: C,D)
(6: A,C)
(6: B,D)
(7: A,B)
(7: F,G)
(8: E,G)
Cluster 1
Cluster 2
Threshold
Ignored

Figure 4. Adaptive clustering image

Figure 4 illustrates an example of adaptive clustering. Here, the centres of face features are used as their representative points; the numbers near the links are minimum distances respectively. Sorted links and the grouped results are listed at the lower left part of the figure. Here, the threshold used to tell whether two face features belong to the same cluster or not is 8.5. Of course, all links which minimum distances are larger than the threshold can be ignored in the clustering procedure.

Adaptive clustering may be used for generalizing buildings, lakes, or other face features.

3.3 Adaptive Aggregation

Adaptive aggregation, following the adaptive clustering procedure mentioned above, aggregates face features in a cluster to form a new face feature with the same type. Figure 5 shows a diagram of adaptive aggregation.

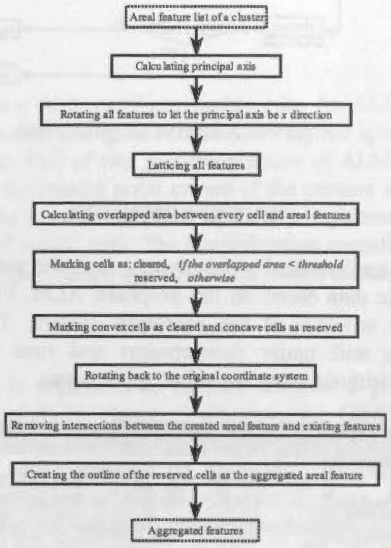

Areal feature list of a cluster
Calculating principal axis
Rotating all features to let the principal axis be x direction
Latticing all features
Calculating overlapped area between every cell and areal features
Marking cells as: cleared, if the overlapped area < threshold
reserved, otherwise
Marking convex cells as cleared and concave cells as reserved
Rotating back to the original coordinate system
Removing intersections between the created areal feature and existing features
Creating the outline of the reserved cells as the aggregated areal feature
Aggregated features

Figure 5. Aggregating diagram

In figure 5, the input data are the face feature lists obtained by clustering procedure mentioned above, and the results are aggregated face features. Some steps of the aggregation procedure in figure 5 are described below.

a) Defining the principal axis of a cluster as the maximum likelihood direction of the cluster. In other words, most segments of features in a cluster are parallel to the principal axis. The principal axis is calculated as follows: first define a histogram of direction from 0 to 180 degrees; then calculate the directions of all segments and vote the obtained direction with the segment's length consequently; lastly find the direction with the peak value in the voted histogram.
b) Rotating all features to let the x-axis parallel to the principal axis. This makes most segments of the features in the cluster being parallel or perpendicular to the x-axis.
c) Removing intersections between created face features and existing features is used to check if there are any intersections and separate them if they are crossed.
d) Creating the outline of all reserved cells as the aggregated face feature was accompanied by representing stair-like cells with slant segments to decrease the vertexes of the aggregated result. Furthermore, holes may be created if there are any cleared cells in the resulted face feature.

3.4 Adaptive Simplification

Adaptive simplification is completed by rounding up all vertexes of linear features in ALM. Figure 6 shows the diagram of adaptive simplification, including elimination, separation, collapse, division, and connection, etc. Adaptive simplification is described in detail in Doihara 2002.

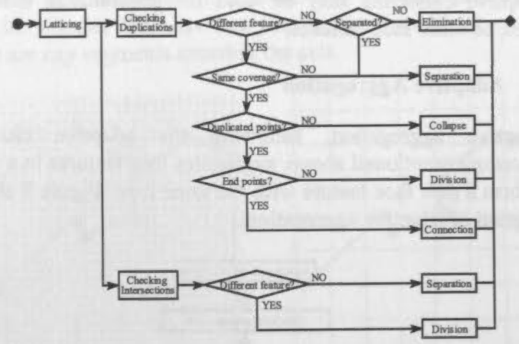


Figure 6. Adaptive simplification

4. EXPERIMENTS

We have implemented a prototype system for generalization of geographic data based on the proposed ALM. Figure 7 shows the main window of the prototype system. This prototype system is still under development and runs on Microsoft windows without using any other applications.

Figure 7. Running window

Figure 8 shows an experimental result, where the input geographic data includes buildings and roads. The statistics about the experiment are listed in Table 1.

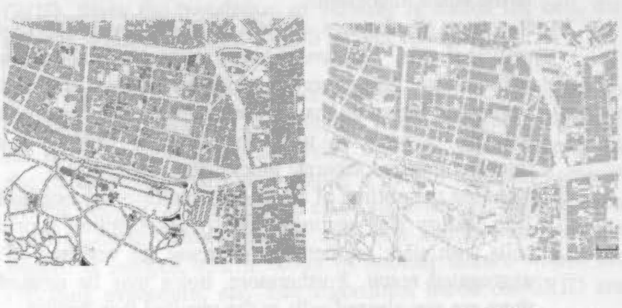


Figure 8. Generalized result: Original buildings and roads (up: 1/2500) and aggregated buildings and simplified roads (down: 1/25000)

Features / Points	Generalization	
	Before	After
Buildings (polygons)	5255	1183
Roads (polylines)	8016	7119
Points (roads+buildings)	65922	32034

Table 1. Margin settings for A4 size paper

5. CONCLUSION

This paper has proposed an Adaptive Lattice Model to generalize geographic data. Generalization decreases the resolution of a geographic dataset, and causes indistinguishableness of the original dataset, which is similar to the imaging principle of a CCD camera. The proposed ALM represents GIS data by latticing adaptively the spatial space at a certain abstraction level or spatial resolution and omitting all details smaller or finer than the resolution. The generalization based on ALM mainly includes selection, adaptive clustering, adaptive aggregation and adaptive simplification, etc. A prototype system has also been constructed based on the proposed model, and experimental results show the effectiveness of the proposed model.

ALM is still under development. Both precise aggregation and improvements for exterminating the intersection between generalized features still remain as our future works.

REFERENCES

Doihara T., 2002. An adaptive lattice model and its application on map simplification. In: *Joint International Symposium on Geospatial Theory, Processing and Application*, Ottawa, Canada, ISPRS Vol.34 ().

Douglas, D.H. and T.K. Peucker , 1973. Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. *Canadian Cartographer*, 10(2), pp.112-122.

Johnston, M.R., Scott, C.D., and Gibb, R.G. 1999. Some problems arising from a simple GIS generalisation algorithm. In: *Proceedings of the 11th Annual Colloquium of the Spatial Information*, Research Centre, University of Otago, Dunedin, New Zealand, pp.191- 199.

Kreveld, M. and J. Peschier 1998. On the automated generalization of road network maps. http://divcom.otago.ac.nz/SIRC/GeoComp/GeoComp98/21/gc_21.htm (accessed 12 Mar. 2002)

Peng W., 2000. Database generalization: conflicts, problems, and operations. In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Holland, ISPRS Vol. XXXIII, Part B4, pp.826—833.

Rumer, U. 1972. An iterative procedure fro the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1, pp.244-256.

Sester, M. 2000, Generalization Based on Least Squares Adjustment, In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Holland, Vol. XXXIII, Part B4, pp.931—938.

AN ADAPTIVE LATTICE MODEL AND ITS APPLICATION TO MAP SIMPLIFICATION

T. Doihara [a], P. Wang [a], W. Lu [a]
[a] Research Institute, Asia Air Survey, Tamura 8-10, Atsugi, Kanagawa, JAPAN
ta.doihara @ajiko.co.jp

Commission IV, Working Group IV/3

KEY WORDS: Simplification, Generalization, GIS – Geographic Data, Resolution, Adaptive Lattice Model

ABSTRACT:

In this paper, we propose an Adaptive Lattice Model (ALM) for simplifying linear features of geographic data. An ALM, based on spatial constraints or the target resolution of the data to be created, represents spatial data by latticing adaptively the spatial space at the target resolution and omitting all details smaller or finer than the resolution. Out of two implementations of ALM, this paper presents the GRID implementation, which moves all vertexes of the features to the nearest point or one of the corners of lattices in ALM. By GRID, all coordinates are rounded up to the integral multiple of the resolution. Since by GRID implement, different vertexes may be translated to the same grid, it can be used to simplify features of spatial data. The simplification operations derived from GRID include connection, division, collapse, elimination, and separation. Experimental simplification results are also presented to show the validity of the present model.

1. INTRODUCTION

Map is an important way of representing geographic data, and is usually measured with a scale, which is defined as "the ratio between the size of an object on the map and its real size on the ground" (Johnston 1999). Changing a map's scale to a smaller one requires the process of generalization, which transforms the original spatial dataset to a coarser representation of the real world. The purpose of generalization is to preserve as much valid information at the target resolution as possible, while reducing as much the amount of data as possible. In other words, some features or details in the original data may be hidden, aggregated, simplified, or deleted, depending on the target resolution.

There are many researches on the generalization of geographic data (Kreveld 1998; Johnston 1999; Peng 2000; Harrie 2000). And the Douglas-Peucker-Ramer algorithm, proposed independently by Ramer (1972) and Douglas and Peucker (1973) respectively, is suggested to be the most visually effective polyline (linear feature) generalization heuristic.

In this paper, an Adaptive Lattice Model (ALM) based on final resolution (target resolution) or spatial constraints of the data to be created is proposed. An ALM represents geographic data by latticing adaptively the spatial space at a certain abstraction level or spatial resolution and omitting all details smaller or finer than the resolution. There are two types of implements for the omission of details. One is GRID, which moves all vertexes of the objects to the nearest point or one of the corners of lattices in ALM, whose coordinates are the integral multiple of the resolution. Here, the movement of a vertex is also called rounding up. The other one is CELL, where every lattice is considered as a cell whose characteristics are statistically measured as a whole. This paper concentrates on the GRID implement of ALM, leaving the CELL implement to another paper (Wang, 2002).

This paper is organized as follows: chapter 2 gives detailed description of ALM, chapter 3 discusses the GRID implement and its related operations such as connection, division, collapse, elimination, and separation are discussed. Chapter 4 presents the implementation of GRID. Chapter 5 gives experimental results to show the validity of the present model.

2. ADAPTIVE LATTICE MODEL

Since the detailed description of ALM has been presented in another paper (Wang, 2002), this paper only gives its brief summary.

A geographic dataset is a kind of representation of the real world, and the generalization of geographic data is only for generating a variety of representation. Although geographic data may be represented on maps with a range of different scales, the validity of observations or inferences drawn from such maps depends upon the resolution of the representation and the viewer's resolution. Considering the observation of a geographic dataset as an imaging system with a CCD camera, what the observer sees are pixels arranged in two-dimension, where each pixel corresponds to and represents statistically a region of the geographic data. Figure 1 shows a typical example.

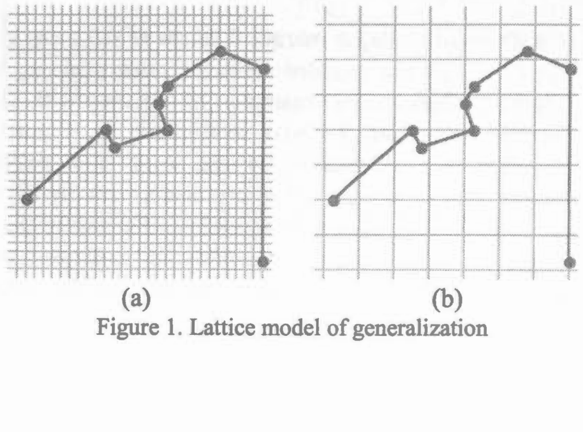

(a) (b)

Figure 1. Lattice model of generalization

Figure 1(a) shows the original data latticed with a higher resolution, and figure 1(b) shows the same data with a lower resolution. Every rectangle or cell of both left and right images works in the same principal as a CCD cell. That is to say, we cannot identify the details within a cell. In other words, an observer can only take a rectangle as the total representation of the corresponding region. The total representation may be whether there are any vertexes in the region, or whether where are any segments crossing with the region, etc.

An Adaptive Lattice Model (ALM) is defined as follows: when geographic data is confined to a predefined lattice where every edge's length is equal to the target resolution, a generalized result of the geographic data will be obtained by calculating the total representation of the corresponding region of the data for every cell. For example, we can check if there are any vertexes of polylines in a rectangle, or calculate the overlapping area's percentage between a rectangle and a polygon in the geographic data. Different calculating methods may be used to generalize various geographic data. In our research, there are two kinds of methods for estimating the features in every cell. One is a statistic method by which the feature is estimated statistically for every rectangle. Another is a direct method where all vertexes' coordinates are rounded up to the integral multiple of the target resolution. The former method, which is based on the statistic characteristics of every cell and called CELL implement of ALM, is useful for generalizing face features or polygon data. The later one, called GRID implement of ALM because it is equal to moving all vertexes to the nearest GRID, is suitable for simplifying linear features or polyline data. The following chapters will concentrate on the GRID implement and its application to the simplification of geographic data.

3. GRID AND ITS OPERATIONS

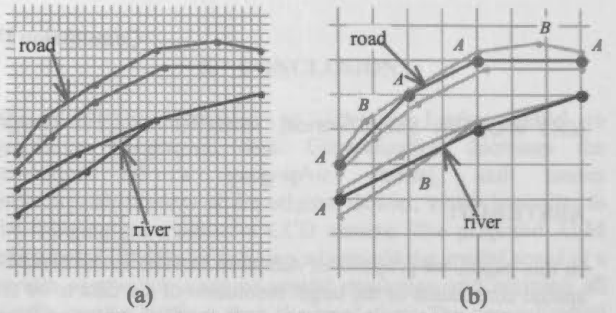
As stated above, GRID is an implementation of ALM by moving all vertexes of the objects to the nearest point or one of the corners of lattices in ALM. As to the translation of vertexes are within the nearest rectangle of each vertex, the statistics of the region data in every cell will not be changed. With GRID, the generalization process is simplified to rounding up all vertexes' coordinates to the integral times of the target resolution and analysing the relationship among all rounded vertexes. For example, a pair of adjoining vertexes will be merged if their rounded coordinates are equal to each other; a polyline may be split if a middle vertex of the polyline is going to be merged with a point of another polyline after the latticing procedure; etc. The following sections will show some typical operations based on GRID for simplifying geographic data.

3.1 Collapse

GRID based Collapse merges features of the same type if their vertexes are rounded up to the same grids in ALM. Figure 2 shows some examples of collapse with GRID, marked as *A* (see the next section for B). Figure 2(a) shows a pair of road edges and some river polylines. Figure 2(b) shows the collapsed results with the decrease in resolution. As shown in figure 2(b), the pair of the road edges are shifted to the same grids and collapsed to one polyline. Similar result is also obtained for the polylines of the river.

3.2 Elimination

Elimination includes deleting duplicated vertexes when adjoining vertexes are rounded up to the same grid and abandoning the middle points when its neighbouring points are colinear. Some examples of elimination are marked as *B* in Figure 2.

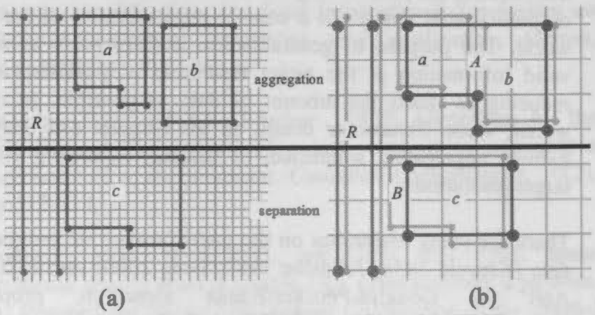


(a) (b)

Figure 2. Collapse/elimination with GRID

3.3 Connection / Aggregation

Connection merges two or more different features in the case when one or more of their vertexes are transformed to the same grids during latticing. The upper half of figure 3 shows an example of connection (marked as *A* in the right graph) between two buildings *a* and *b*, when they are simplified by using GRID implement of ALM. Here, the adjoining edges of buildings *a* and *b* become partially overlapping after latticing. After deleting the overlapping part, the reserved edges are connected to create a large building. The connection of face features is also called aggregation in the generalization of geographic data.



(a) (b)

Figure 3. Connection/separation with GRID

3.4 Separation / Displacement

Separation or displacement appears when points consisting different types of features are latticed to the same grid. Since they belongs to different coverage, they cannot be merged or connected to create a new feature, like the connection of two building mentioned above. In such case, the feature with less importance must be displaced with a certain distance away from the more important feature. The lower half of figure 3 shows another example, where one edge of building *c* overlaps with road *R* when they are latticed, and the overlapped edge of building *c* has been shifted to separate it from road *R*.

3.5 Division

Division occurs in the case of intersection between two different features. Consider a river, which is represented by a pair of polylines, and its width varies. When the river is simplified by using a given resolution, the narrower parts of the river may be collapsed to a single polyline with width 0, but the left should be still represented with two polylines. Then, the collapsed parts and the left parts must be divided first. Figure 4 shows one example of the division based on GRID.

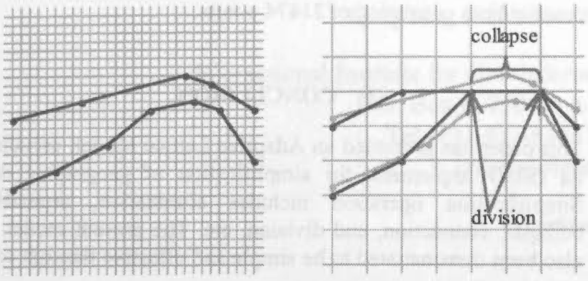


Figure 4. Division with GRID

4. SIMPLIFICATION WITH GRID

We propose a new simplification model based on GRID, including collapse, elimination, connection, division, and their combination. The processing flow is shown in Figure 5.

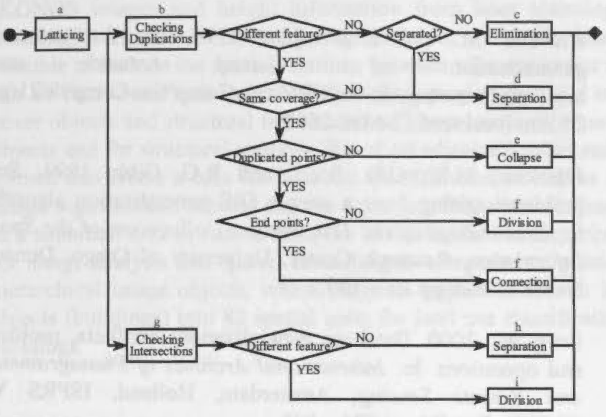


Figure 5. Processing Diagram

The present simplification mode includes the following steps:

a) Latticing all vertexes of features with the following equation.

$$x_r = r \cdot \text{round}\left(\frac{x}{r} + 0.5\right) \qquad (1)$$

$$y_r = r \cdot \text{round}\left(\frac{y}{r} + 0.5\right)$$

where, r = target resolution

x, y = coordinates before latticing

x_r, y_r = coordinates after latticing

round(z): return the largest integer which is less than or equal to z.

b) Checking duplicated vertexes and segments.
Duplicated vertexes are those points which are originally different but locates at the same grid after being latticed. If a segment of one feature overlaps partly or entirely with a segment of another feature after being latticed, the overlapping segments are called duplicated segments.

c) Eliminating duplicated vertexes and segments.
If the duplicated vertexes and segments are from the same feature and adjoining with each other, the duplicated one should be deleted to keep the uniqueness of the simplified data.

d) Separating overlapped segments and duplicated vertexes.
There are two cases for separation for duplications. One case is the separation of overlapping segments and duplicated vertexes, if they belong to the same feature but do not adjoin each other. The other case is the separation of duplications that belong to different coverage.

e) Collapsing overlapped segments.
If the overlapped segments are the parts of same road edges or the same riverbanks, the overlapping segments will be merged to one segment, and the original features, road edges or riverbanks, are collapsed from areas or pairs of polylines to single polylines.

f) Dividing features at the duplicated vertexes.
If one of the duplicated vertexes is not the end-point of the original feature, the original feature should be divided into two features to convert the duplicated vertexes to nodes. Otherwise, the two features should be connected with each other to create a new large feature.

g) Checking intersections.
Latticing vertexes of all features may cause intersections between segments. Checking of intersections is for identifying such intersections.

h) Separating the intersections of the same feature.
An intersection between two different segments of the same feature indicates a loop error. To avoid loop generation, the intersection should be settled by separating them with each other or deleting the loop if the area of the loop is small enough.

i) Dividing crossing features.
If the intersection happened between two different features, one or both features may be divided to make the cross point be a new node, which is the end point of each divided feature.

5. EXPERIMENTS

Based GRID implement and its operations, we constructed a prototype system for simplification of geographic data. Figure 6 shows the main window of the prototype system.

Figure 6. Main window of the prototype system

We applied the prototype system to some sample data of real world and have obtained some promising results. The upper image in figure 7 is the original data, in which filled areas represent roads. The lower image is the simplified result, where narrow roads are collapsed to single polylines and the wide roads still remain as areas.

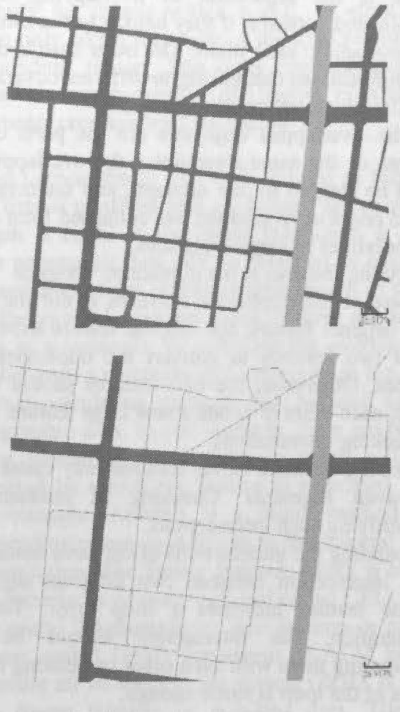


Figure 7. Simplified result. Upper image: original roads (1/2500), lower image: simplified results (1/25000)

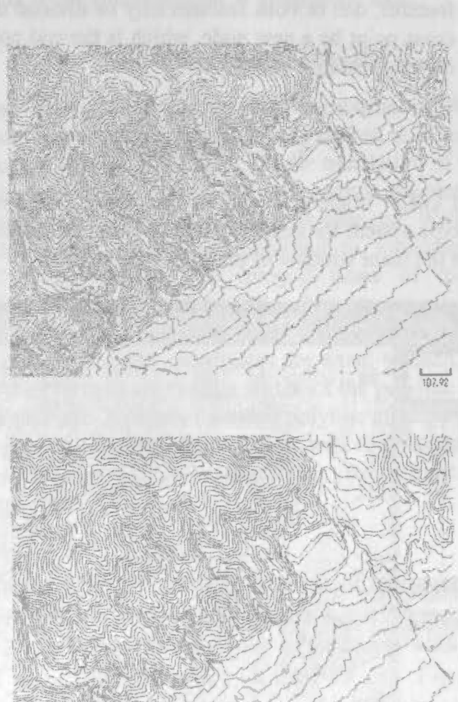


Figure 8. Simplified result. Upper image: original contour lines (1/2500: 44304 points / 1370 objects), lower image: simplified contour lines (1/10000: 21474 points / 1315 objects)

Figure 8 shows a result of simplified contour lines. Here, the upper image is the input data with 1370 contour lines consisting of 44304 points. After simplification with the proposed ALM model, the result, shown as the lower image, becomes 1311 contour lines consisting of 21474 points.

6. CONCLUSION

This paper has presented an Adaptive Lattice Model, especially its GRID implement, for simplification of geographic data. Simplification operation includes elimination, separation, collapse, connection, and division, etc. The present model has also been demonstrated to be simple and effective through some experiments.

As future works, we are going to improve our model to preserve the non-intersection property of the simplified linear features. We are also going to further integrate the two implements, GRID and CELL, to implement other operations which are necessary for generalizing geographic data.

REFERENCES

Kreveld, M. and J. Peschier 1998. On the automated generalization of road network maps. http://divcom.otago.ac.nz/SIRC/GeoComp/GeoComp98/21/gc_21.htm (accessed 12 Mar. 2002)

Johnston, M.R., C.D. Scott, and R.G. Gibb, 1999. Some problems arising from a simple GIS generalisation algorithm. In: *Proceedings of the 11th Annual Colloquium of the Spatial Information*, Research Centre, University of Otago, Dunedin, New Zealand, pp.191- 199.

Peng W., 2000. Database generalization: conflicts, problems, and operations. In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Holland, ISPRS Vol. XXXIII, Part B4, pp.826—833.

Harrie, L. and T. Sarjakoski 2000. Generalisation of vector data sets by simultaneous least squares adjustment. In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Holland, ISPRS Vol. XXXIII, Part B4, pp.348—355.

Wang P.T., T. Doihara and W. Lu, 2002. Spatial generalization: an adaptive lattice model based on spatial resolution. In: *Joint International Symposium on Geospatial Theory, Processing and Application*, Ottawa, Canada, ISPRS Vol.34.

FINDING SPATIAL UNITS FOR LAND USE CLASSIFICATION BASED ON HIERARCHICAL IMAGE OBJECTS

Qingming Zhan, Martien Molenaar, Klaus Tempfli

International Institute for Geo-Information Science and Earth Observation (ITC)
P.O. Box 6, 7500 AA Enschede, The Netherlands
Tel: +31 53 4874374, Fax: +31 53 4874335
{zhan|molenaar|tempfli}@itc.nl

Working Group IV/3

KEY WORDS: Spatial clustering, Urban land use, Triangulation, Hierarchical image segmentation, Image object

ABSTRACT:

Remote sensing in urban areas has been a challenger for quite some time due to their complexity and fragment with combination of man-made features and natural features. High-resolution satellite images and airborne laser altimetry data offered potential possibilities for feature extraction and spatial modelling in urban areas. Land use classification of urban areas may become possible by exploiting current high-resolution sensor data. The proposed approach incorporates spectral information from multi-spectral IKONOS images and height information from laser scanning data in hierarchical image segmentation based on semantically meaningful thresholds. By image segmentation, we obtain image objects at several levels with certain properties, which make it possible to include the spatial relations between adjacent image objects. Land cover classification and identification of image objects can be carried out mainly according to their properties. Land use classification at a higher level need to be inferred based on land cover objects and structural information at lower levels. We use Delaunay triangulation for deriving spatial relations between image objects and for structural analysis. Based on adjacency relationships of image objects, human settlements and other urban spaces are formed that create a base for land use identification as well as for structural analysis of urban areas. In this paper, the hierarchical image segmentation schema and the corresponding semantic-based thresholds are presented. To test the approach, we selected a site in a suburban area in Amsterdam, the Netherlands. The experiments show that hierarchically formed image objects are useful tools for image analysis and spatial modelling as compared to pixel-based approaches. Structural information can be derived based on hierarchical image objects, which plays an important role in land use classification in urban areas. We could cluster 704 image objects (buildings) into 82 spatial units for land use classification based on the measurement of shortest distance between adjacent buildings.

1. INTRODUCTION

Increasing availability of high spatial-resolution remote sensing images and airborne laser altimetry data offers new opportunities, especially for urban planning. However, it seems desirable to make progress in automating high-level information extraction on urban land use (Barr and Barnsley, 1995, 1997). Urban land use in urban planning context refers to certain functions with related social economic characteristics. For instance, a residential area consists of a number of physical features such as residential buildings, parking space, footpaths, green space, and maybe canals. Quite often, these features are targets of land cover classification. Physical features in general have certain associations with spectral features, so they can be identified by using multi-spectral information of remote sensing images. However, land use cannot be determined by land cover information directly. For instance, a building can be a residential house, a shop or a warehouse. Green space can be found in a park, a residential area or even in a commercial area. Therefore, in land use identification, additional information or evidences have to be found on top of land cover information. Furthermore, a reasonable spatial unit has to be delineated to represent the spatial extend of certain land use although this type of boundary is often vague. This research is intended to find the way to delineate areas of different land use and identify the land use type in every delineated area. Delaunay triangulation is deployed in creating spatial associations and structural analysis toward spatial clustering of physical features in image space with the aim of identifying land use. Delaunay triangulation has been widely used in spatial analysis and spatial modelling (Bundy *et al.*, 1995). However, most applications are based on vector data. In this research, we deploy Delaunay triangulation on lattice (image) data. The test image we used in our triangulation is a binary image; showing all the buildings with value 1 against everything else 0. The approach of deriving such building segments is presented in section 2. A detail description for finding spatial units can then be found in section 3. The experimental results in finding spatial units for land use identification are presented in section 4.

2. IMAGE SEGMENTATION AND IMAGE OBJECT

Airborne lasers scanning data and high-resolution multi-spectral remote sensing images are expected to have great potential use in feature extraction, especially for urban areas. Semantic-based image segmentation introduced by Zhan shows that many man made objects can effectively be derived from laser scanning data and multi-spectral remote sensing images. Detailed description of the image segmentation schema applied in this research can be found in Zhan *et al.* (2001). Some related approaches can be found in other literatures as well (Brunn and

Weidner, 1997; Haala and Brenner, 1999; Hug and Wehr, 1997; Lemmens *et al.*, 1997).
We used high-resolution IKONOS imagery and airborne laser scanning data in this case study presented here. The laser scanning data we took, the 5-meter resolution raster product of AHN (Actual Height of the Netherlands), were resampled to 4-meter grid image to match the resolution of IKONOS multi-spectral image. The NDVI (the Normalised Difference Vegetation Index) image was created based on Band 3 and Band 4 of the IKONOS image in order to get semantically meaningful data (i.e. vegetation etc.) and reduce the amount of data.

Based on knowledge of the test site and careful study of related data, we segmented the AHN data with a threshold of 0-meter in altitude (i.e. mean sea level). The study area and its surroundings are reclaimed from a lake (Bijlmermeer). Most of ground floors are around 3 meters below sea level and the area is flat in general. 3 meters above ground floors is also a good threshold for finding buildings and other high objects such as trees. Therefore, the first segmentation was applied with threshold of 0-meter on AHN data. The result shows that not only buildings and trees are segmented but also most of main roads and parking garages. Hence, additional separation has to be made by using size of objects to separate raised roads from buildings. Subsequent segmentations are mainly based on the NDVI data for each branch of the hierarchical tree structure according to Zhan *et al.* (2001). Thresholds of 0.75 and 0.65 are based on the sample study of built-up area, green space and water as well as efforts in finding lowest possible valleys in the histogram space of NDVI data. This resulted in separating trees from buildings.

Urban space has been planned and constructed in certain ways that enable serving certain functions of usage. Some spatial arrangements have been applied in spatial organization. Therefore, we believe that structural information and spatial pattern can play an important role in finding out the spatial extends of certain land use. In this stage, only building segments ($S_{Building}$) were used in this experiment. Fig. 1 shows the building segments of a sub area of the total test site. The building segments were derived based on following thresholds:

$$S_{Building}\{AHN(0,\max), Size(\min,\ 1\ ha.), NDVI(\min,0.65),\ \ \forall\ s_i \in I\}$$

where image space $I\ \{AHN,\ NDVI,\ ...,\forall\ pixel\ p_i \in I\}$

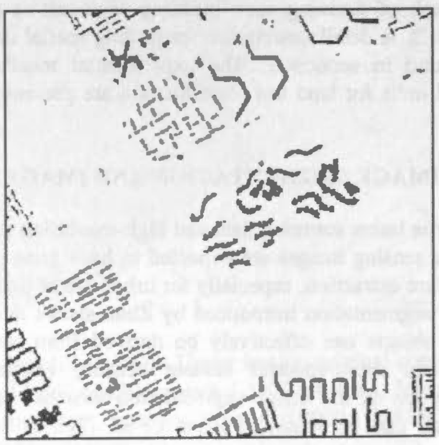


Fig. 1 Image objects (buildings) derived from image analysis

3. SPATIAL CLUSTERING

3.1 Spatial Clustering Based on Distance between Adjacent Buildings

Based on the assumption that buildings, which are close to each other, are used for similar function, we use the shortest distance between adjacent buildings as a measurement in spatial clustering. A threshold is needed for distance as criterion for similar type of land use. This threshold is to be inferred in an optimisation process based on number of spatial clusters and other indicators.

3.2 Delaunay Triangulation

In order to find out possible links between adjacent buildings and the shortest distance, we developed a Delaunay triangulation algorithm based on the Quickhull Algorithm (Barber at al 1996).

The binary image is converted to a 'labelled image'. By checking the 4-connection, individual labelling are identified and uniquely labelled. The next step in raster to vector conversion of the labelled image, the centre of building pixel becomes a point.

Delaunay triangulation is applied then to all points (building pixels). Since we are interested in find relations between adjacent buildings, we delete all the triangles that have the same label value. From the remaining triangles, we create a matrix, which indicates the links between adjacent buildings and the shortest distance between them.

3.3 Reasoning for Spatial Clustering

Threshold (MaxDist) is defined as the largest possible distance between adjacent buildings to be considered as in the same cluster (spatial unit of a land use type candidate). All links between adjacent buildings will be broken if they are apart by more than the threshold.

In order to determine the threshold, spatial clustering is done in a loop using trial thresholds from 5 meters to 250 meters. A threshold of 5 meters will generate a cluster for each building. A threshold of 250 meters will group all buildings in only one cluster. An optimised threshold can be determined based on checking several cluster-related indicators such as number of spatial clusters, maximum number of buildings in a cluster, minimum number of buildings in a cluster, number of isolated buildings, and average number of buildings per cluster. The definitions of these indicators with short explanations are:

- Number of spatial clusters (NoCluster)

The umber of spatial clusters is counted according to number of separated clusters (no link between them) when a threshold was applied. When the threshold increases the number of spatial clusters decreases.

- Maximum number of buildings in a cluster (MaxNoObjects)

The maximum number of buildings in a cluster indicates highest number of buildings among all clusters. When the threshold increases the maximum number of buildings in a cluster increases.

- Minimum number of buildings in a cluster (MinNoObjects)

The minimum number of buildings in a cluster indicates the lowest number of buildings among all clusters. When the threshold increases, the minimum number of buildings in a cluster increases as well but with different rate as compared to the maximum number of buildings in a cluster.

- Number of isolated buildings (NoIsolatedObjects)

The number of isolated buildings is taken from the number of clusters that consist of only one building. When the threshold increases the number of isolated buildings decreases.

- Average number of buildings per cluster (AvNoObjectsPerCluster)

The average number of buildings per cluster is calculated as the total number of buildings divided by the number of clusters. When the threshold increases, the average number of buildings per cluster increases. The rate of increase is in a range between maximum and minimum number of buildings in a cluster.

3.4 Creating Convex Hull of Land Use Clusters

The next step is to assign each building to one of clusters found in previous steps and to delineate the convex hull for each cluster using the convex hull algorithm (Barber *at al.*, 1996). The derived convex hulls will be used in identifying the land use type at a higher level of reasoning.

4. CASE STUDY

4.1 Study Area and Data used in this Study

A 9 km^2 (3 km × 3 km) area, Southeast of Amsterdam, was selected for the experiment (see Fig. 2). Approximately 200,000 people live in this sub-urban district. Several types of residential as well as commercial areas, parks, lakes and canals can be found in the study area. Built-up area, green space and water are the three end-member land cover classes in this study. However, to avoid additional complexity, a test image was prepared that consists of image objects, which are exclusively buildings (derived from image analysis and edited with reference to the 1:1000 scale cadastral maps (see Fig. 4)). Other image objects would have to be used in other stages of the total process.

Data used in case study are listed in Table 1. The laser scanning data is shown in Fig. 3.

Type of Data	Date	Band/Colour	Resolution or Scale
Laser scanning data	March 1998	Height in real value	5 m
IKONOS image	June 2000	Multi-spectral	4 m
Topographic map	May 1999	Black & White	1:1,000

Table 1. List of data used in the experiments

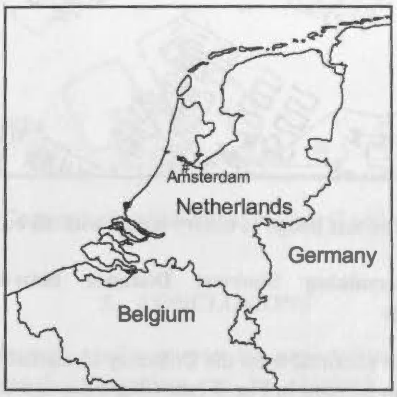


Fig. 2 Location of the study area (Southeast, Amsterdam)

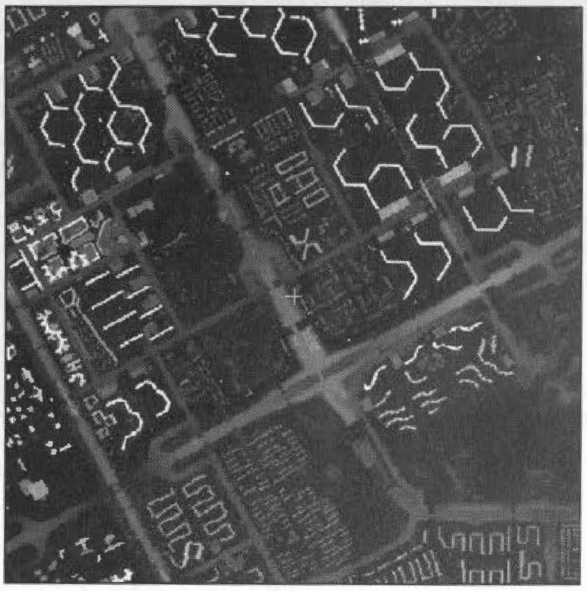


Fig. 3 The AHN image (Light tone refers to higher objects)

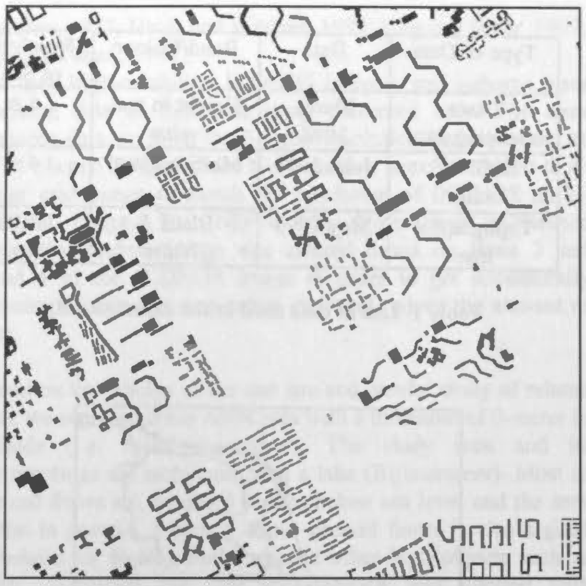


Fig. 4 The test image (a binary image with all building pixels)

4.2 Determining Shortest Distance between Adjacent Buildings

The result obtained from the Delaunay triangulation on building pixels can be seen in Fig. 5 (showing a portion of the total area of Fig. 4 and it corresponding to a area in Fig. 10 where cluster 7, 26 and 30 are located). Removing the links based on the shortest distance between adjacent buildings produces a result as shown in Fig. 6. An output shown in Fig. 7 is a result of using 40 meters as a threshold.

Based on the reasoning approach outlined in the previous section, a test run was initiated in finding a reasonable number of spatial clusters using the following indicators. The shortest distance between adjacent buildings was used as criterion to link adjacent buildings. The shortest distances between buildings were checked against the threshold. Using the indicators described in the previous section lead to the results plotted out in Fig. 8.

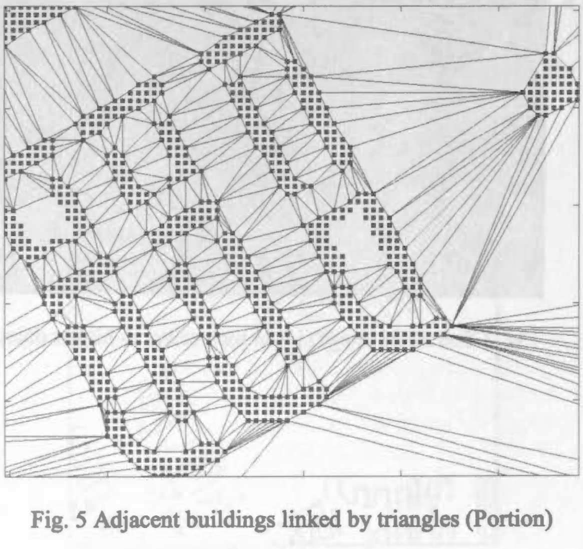


Fig. 5 Adjacent buildings linked by triangles (Portion)

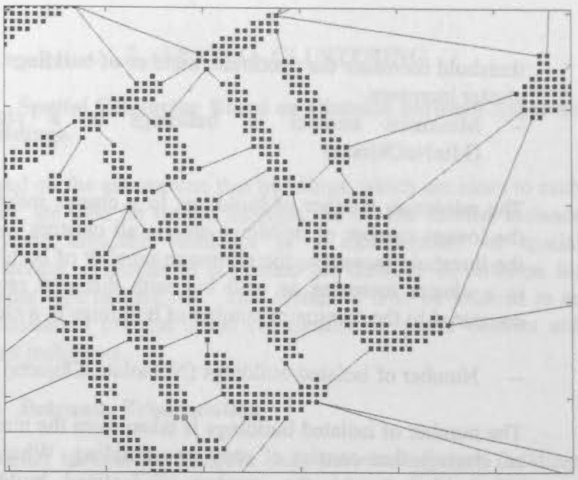


Fig. 6 Adjacent buildings linked with shortest distances between them (Portion)

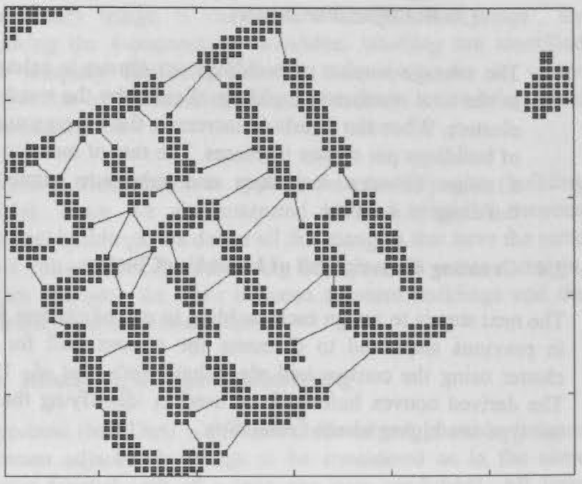


Fig. 7 Linked buildings with the threshold value of 40 meters (Portion)

4.3 Determining the Number of Clusters

Fig.8 shows that the number of spatial clusters (NoCluster) and the number of isolated buildings (NoIsolatedObjects) decrease sharply when the threshold increases in the range of 5-meter to 30-meter. The rate of decrease becomes stable when the threshold follow increased.

On the other hand, the maximum number of buildings in a cluster (MaxNoObjects), the minimum number of buildings in a cluster (MinNoObjects) and the average number of buildings per cluster (AvNoObjectsPerCluster) increase when the threshold increases but with different rates of increase. MinNoObjects increases sharply when there is no isolated building found. This is not a good indicator since there are buildings such as gasoline stations etc. that should not be included in clusters with many buildings surrounding. However, we consider that AvNoObjectsPerCluster is not a proper indicator either, since it mixes up many different types of spatial clusters such as clusters with many buildings inside and clusters containing only one building.

The test results show that number of spatial clusters (NoCluster) and maximum number of buildings in a cluster (MaxNoObjects) gives the best indication in finding number of spatial clusters in this case study (see Fig. 8). Since MaxNoObjects shows as a continuous curve in Fig. 8, but it is not a smooth curve and it may behaves differently in different types of urban areas, further investigations have to be made in finding better indicators. However, clustered regions created by the proposed approach can be used as a base for additional analysis such as removing outliers in each spatial unit and combining similar clusters using other measurement for instance size, shape, height etc.

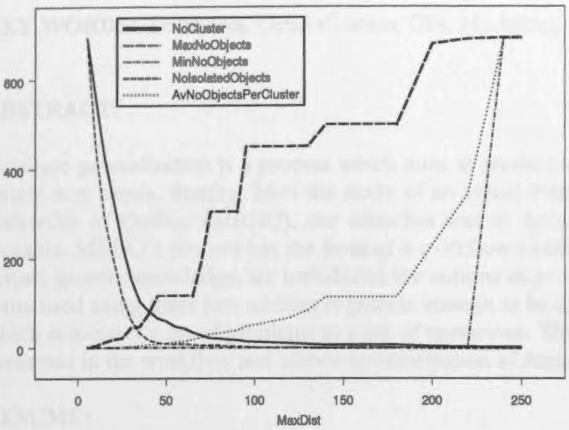


Fig. 8 Proposed indicators change in corresponding to different thresholds used for determining reasonable number of clusters

4.4 Classifying Buildings into Spatial Clusters and Delineation of Cluster Boundaries

Once the best threshold is determined, individual buildings are grouped based on the links that survive (see Fig. 7). Clustered buildings can be represented in image space again assigning the spatial cluster numbers to building pixels, replacing the label numbers. A part of the respective output image is shown in Fig. 9. The clusters shown in Fig. 9 correspond to the clusters 65, 67, 69 and 74 in Fig. 10. The cluster boundaries as shown in Fig. 10 for the entire test site were delineated by using the convex hull algorithm.

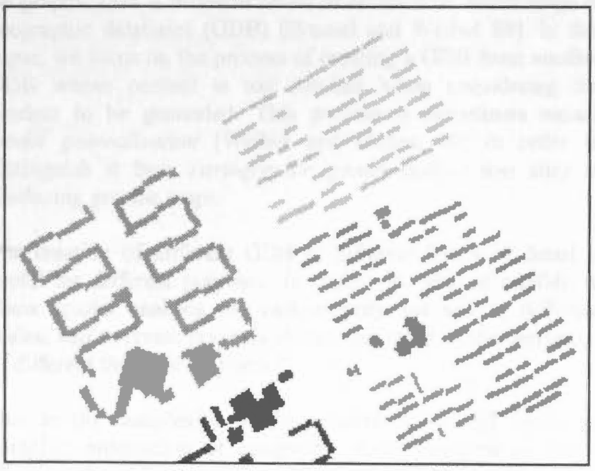


Fig. 9 Clustered buildings with threshold of 40 meters (Portion)

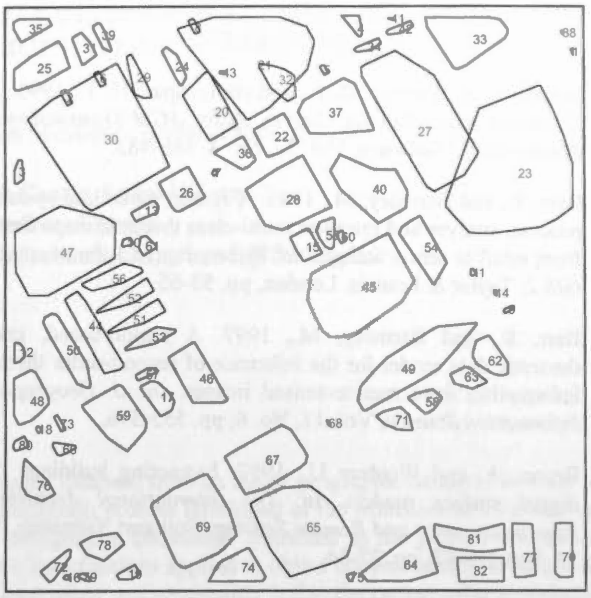


Fig. 10 Convex hulls of clustered regions (Land use units)

5. CONCLUSIONS

The experimental results support the expectation that higher-level structural information can be obtained from spatial reasoning on segmented data. Structural information is useful in image understanding as well.

Hierarchical image segmentation and hierarchical image objects are useful tools for image analysis and classification especially in urban areas. This opens new possibilities to incorporate rules in image understanding, classification and aggregation. The results presented in this paper are just a beginning of referencing higher level of structural information. Only building pixels were used at this stage. More complicated structural information are expected to be derived when using additional information such as size, shape, height, etc of buildings and including vegetation and water pixels. These types of information have been derived already during image analysis and hierarchical image segmentation. We expect that the current approach is quite suitable in finding different types of residential and commercial areas.

There are a number of issues for further investigation such as finding better indicators to replace the maximum number of buildings in clusters, i.e. to include size, height of buildings, etc. A check on outliers in each clustered region will lead to further improvement.

ACKNOWLEDGEMENT

This research was funded by the Ministry of Development Cooperation (DGIS), the Netherlands as part of DSO Project implemented by ITC, the Netherlands and Wuhan University, China. Thanks will go to Survey Department (Rijkwaterstaat, Meetkundige Dienst), Ministry of Transport and Public Works, The Netherlands for providing the laser scanning data (AHN) of the study area.

REFERENCE

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. T., 1996. The Quickhull Algorithm for Convex Hulls. *ACM Transactions on Mathematical Software*, Vol. 22, No. 4, 469-483.

Barr, S., and Barnsley, M., 1995. A spatial modelling system to process, analyse and interpret multi-class thematic maps derived from satellite sensor images. In: Fisher, P., (ed.), *Innovations in GIS 2*. Taylor & Francis, London, pp. 53-65.

Barr, S., and Barnsley, M., 1997. A region-based, graph-theoretic data model for the inference of second-order thematic information from remote-sensed images. *Int. J. Geographical Information Science*, Vol.11, No. 6, pp. 555-576.

Brunn, A. and Weidner U., 1997. Extracting buildings from digital surface models. In: *The International Archives of Photogrammetry and Remote Sensing*, Stuttgart, Germany, Vol. 32, Part 3-4W2, pp. 27-34.

Bundy, G. L., Jones, C. B. and Furse, E., 1995. Holistic generalizetion of large-scale cartographic data. In: J. C. Muller, J. P. Lagrange and R. Weibel (eds.), *GIS and Generalization*. London etc., Taylor & Francis: pp.106-119.

Haala, N. and Brenner, C., 1999. Extraction of buildings and trees in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, No. 2-3, pp. 130-137.

Hug, C. and Wehr A., 1997. Detecting and Identifying Topographic Objects in Imaging Laser Altimeter Data. In: *The International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3-4W2, pp. 19-26.

Lemmens, M. J. P. M., Deijkers, H. and Looman, P. A. M., 1997. Building detecting fusing airborne laser-altimeter DEMs and 2D digital maps. In: *The International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3-4W2, pp. 42-49.

Zhan, Q., Molenaar, M., and Gorte, B., 2000. Urban land use classes with fuzzy membership and classification based on integration of remote sensing and GIS. In: *The International Archives of Photogrammetry and Remote Sensing*, Vol. 33, Part B7, pp. 1751-1758.

Zhan, Q., Molenaar, M., and Xiao, Y., 2001. Hierarchical object-based image analysis of high-resolution imagery for urban land use classification. In: *IEEE - ISPRS joint workshop on remote sensing and data fusion over urban areas*, Rome, Italy, pp.35-39.

FORMALISING THE GEOGRAPHIC DATABASE GENERALISATION PROCESS BY MEANS OF A CONFLICTS/OPERATIONS GRAPH

Dominique Han-Sze-Chuen, Sébastien Mustière , Bernard Moulin

Computer Science Department and Geomatics Research Centre, Laval University,
G1K 7P4, Ste-Foy, Québec, Canada
dominiquehan@yahoo.com, {bernard.moulin, sebastien.mustiere}@ift.ulaval.ca

KEY WORDS: Databases, Generalisation, GIS, Modelling.

ABSTRACT:

Database generalisation is a process which aims at producing a geographic database from an initial geographic database in order to satisfy new needs. Starting from the study of an actual database generalisation process developed at the Ministère des Ressources Naturelles of Québec (MRNQ), our intention was to formalise the cartographic knowledge contained in the process to make it reusable. MRNQ's process has the form of a workflow which contains a list of actions applied to object classes. Since we wanted to extract generic knowledge, we introduced the notions of *problem* and *operation*: our assumption is that the cartographic knowledge formalised using these two notions is generic enough to be applied to other cases of generalisation. Thus, we propose a causal graph which associates a set of problems to a set of operations. The utility of such a model is that it offers an explanation for each actions contained in the workflow and allows an anticipation of future problems when creating a new workflow.

RÉSUMÉ:

La généralisation de base de données est un processus qui vise à produire une base de données géographiques, à partir d'une base de données géographiques initiale en vue de répondre à de nouveaux besoins. En nous appuyant sur l'étude d'un cas réel de généralisation développé au Ministère des Ressources Naturelles du Québec (MRNQ), notre objectif a été de formaliser les connaissances cartographiques contenues dans ce processus afin de pouvoir les réutiliser par la suite. Le processus du MRNQ se présente sous la forme d'une chaîne de production contenant une liste d'actions qui s'appliquent sur des classes d'objets. Du fait que nous voulions extraire des connaissances génériques, nous avons introduit les notions de *problème* et d'*opération* : notre hypothèse est que les connaissances cartographiques, formalisées à l'aide de ces deux notions, sont assez génériques pour être utiles à d'autres cas de généralisation. Nous proposons ainsi un graphe causal qui associe un ensemble de problèmes à un ensemble d'opérations. L'intérêt d'un tel modèle est qu'il fournit une explication pour chaque action de la chaîne de production et qu'il permet d'anticiper d'éventuels problèmes lors de la création d'une nouvelle chaîne de production.

1. INTRODUCTION

Geographic data generalisation is a key process to produce geographic data at different levels of abstraction, either maps or geographic databases (GDB) [Brassel and Weibel 88]. In this paper, we focus on the process of creating a GDB from another GDB whose content is too detailed when considering the product to be generated. This process is sometimes named *model generalisation* [Weibel and Dutton 99] in order to distinguish it from *cartographic generalisation* that aims at producing graphic maps.

The creation of different GDB at different levels of detail is useful for different purposes: to reduce the size of a GDB, to allow spatial analysis for various purposes and at different scales, and to create pivot databases dedicated to the derivation of different thematic products.

Due to the complexity of the process, there still exists no complete automation of geographic data generalisation [Joao 98; Weibel and Dutton 99]. To date, no GIS provides the means to automatically generalise data. Few of them provide a toolbox of generalisation algorithms, but the cartographer's intervention is still needed at two levels. First, s/he must define how to use the toolbox algorithms according to the target product, either to develop a global process or to interactively select algorithms during the generalisation process. Then, s/he must perform manual operations, either to correct the results provided by the algorithms or to manually perform operations for which no algorithm is provided.

Our work starts from the study of an actual generalisation process developed at the Ministère des Ressources Naturelles (MRN) of Quebec. Several meetings with the cartographers of the MRN and several documents describing this particular generalisation process enable us to collect a large amount of cartographic knowledge about model generalisation. The purpose of our work presented in this paper is to study how to formalise the knowledge collected from this particular generalisation process in order to make it reusable for others cases of generalisation.

In the second section of this paper, the MRN's generalisation process is described. Then, in the third section, a model of representation of cartographic knowledge in generalisation is proposed by means of a causal graph. Finally, in the fourth section, we discuss the advantages and limits of the proposed model.

2. ANALYSIS OF AN ACTUAL GENERALISATION PROCESS

2.1 The new database context production

Our study dealt with an actual generalisation process developed at the *Ministère des Ressources Naturelles (MRN)* of Quebec. This process aims at producing a 1/100,000 scale topographic database (namely the BDAT) from a 1/20,000 scale topographic database (namely the BDTQ). This work is supported by the Intergraph's DynaGen software, which is a platform dedicated to model generalisation.

This workflow have several characteristics. First, a good quality of generalisation is needed because the BDAT is a pivot database that will be used for several purposes. Second, because of cost and time constraints, the MRN's generalisation process has been automated as much as possible. Then, due to the fact that the production of the BDAT will be subcontracted to private companies, the generalisation process is very detailed in order to insure a good and homogeneous quality of the results.

Consequently, in a reverse engineering point of view, this workflow is a great opportunity to collect a large amount of cartographic knowledge about the generalisation process.

2.2 The workflow

In order to specify how to produce the new intended database (called the final database in this paper), the MRN's cartographers developed a workflow decomposing the generalisation process into a sequence of basic steps associating certain groups of objects to specific actions to be applied on them (Figure 1).

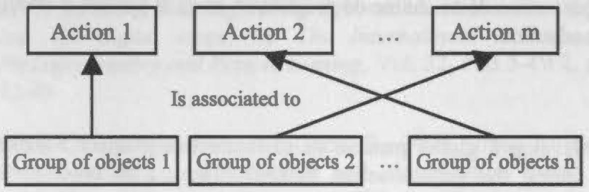


Figure 1: the workflow links groups of objects to actions

A group of objects which must be generalised is defined in the workflow either by a full class of objects or by a condition on the objects of a class (e.g. all the objects of the class *2-D rivers* whose *minimum width* is less than *100 m*; or all the objects of the class *lake* whose attribute value *area* is under *30,000 m2*). An action which is performed is either manual or automated (e.g. select all isolated lakes by means of a query, visually check the data at the junction of two datasheets, apply the algorithm *area to line* with parameters *5 and 0.5*, apply the algorithm *Douglas and Peucker* with the parameter *0.05* [Douglas and Peucker 73]).

It must be noticed that many successive versions of this workflow have been developed and tested by specialists from MRN during several months, before getting an adequate process. Another important point is that the workflow still contains manual operations.

2.3 Purpose of the work

We wondered if the cartographic knowledge included in the workflow could be reused in another generalisation process using other data, with another generalisation software in order to produce another product.

It is important to notice that this workflow is dedicated to a particular GDB creation, from a particular GDB, with a particular tool, in a particular organisational context. Even if the workflow is well documented, it may be over-detailed to be easily reusable when developing a new generalisation process. Anyway, our assumption is that this workflow contains some generic knowledge about which actions should be done in various cases that can be encountered during different generalisation processes.

The purpose of our work is then to:

- *Organise* the knowledge involved in the workflow in order to emphasise the most relevant information potentially useful during the development of a new generalisation process.
- *Extract* the most generic (and then reusable) knowledge from the workflow.
- *Reformulate* the knowledge involved in the workflow in order to make it easily adaptable to a new generalisation process.

3. OUR MODEL FORMALISING THE DATABASE GENERALISATION

3.1 Extraction of the generalisation actions

If the workflow contains many generalisation operations that may be relatively generic, it also takes into account specific problems due to the characteristics of the database and the generalisation software used. The solutions to these problems are specific actions such as changing certain object attributes in the database or simplifying objects before changing their dimension to improve the efficiency of reduction algorithms. In fact, these specific actions can be thought of as a pre-processing which aims at preparing the data to be generalised.

Here is a list of the main reasons justifying the performance of specific actions:

- The initial database may contain some minor errors that could be emphasised during the generalisation process. These errors have to be corrected.
- The initial database model is not adapted. Some modelling choices made in the BDTQ are more adapted to directly draw the data on a map rather than automatically analysing and transforming them.
- Some generalisation algorithms are not efficient in some particular configurations. Actions must then be done to prevent problems in these cases or to to correct them afterwards.
- There exists no algorithm to make certain intended operations. Series of operations are then made to overcome this deficiency.

Hence we must distinguish, first, the actions which one performed to prepare the data and are very specific to the initial data and, second, the actions performed to generalise the data, which means transforming the data in order to respect the specification of the final database. Since our goal is to formalise generic knowledge, our model will only take into account generalisation actions.

3.2 Reformulation of the actions of the workflow in a problem / operation formalism

The actions described in the workflow directly link some object classes of the initial database to algorithms used to transform them (e.g. all objects of the class *River* must be transformed by the algorithm *Simplification* with a parameter *p*). The drawback of this approach is that these object classes are specific to the BDTQ. Hence, in another case of generalisation, the initial database may contain different classes of objects. Since the interest of formalising the knowledge contained in the workflow is to make it reusable, associating a class to an action is not a good option because such an association will be relevant only to MRN's particular generalisation process.

In order to extract generic knowledge from MRN's workflow, we use the notions of *problem* and *operation*. Our assumption is that certain generic problems must be solved with certain generic operations whatever the case of generalisation. The generic part of knowledge relies on the links between problems and operations, while the specific part of knowledge relies in the links between, on the one hand, objects and the problems they have, and on the other hand, operations and the algorithms used to perform them. Thus, as described in Figure 2, our model will consider the generalisation process as a set of *problems* that must be solved by applying certain *operations* [Armstrong 91; McMaster 91; Shea 91].

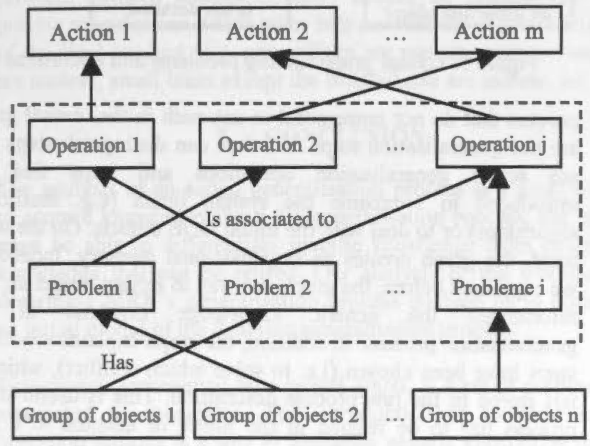

Action 1
Action 2
...
Action m
Operation 1
Operation 2
...
Operation j
Is associated to
Problem 1
Problem 2
...
Probleme i
Has
Group of objects 1
Group of objects 2
...
Group of objects n

Figure 2: Introduction of the notions of *problem* and *operation* in order to extract generic knowledge

A problem results from the fact that a constraint on the final product is not respected [Beard 91; Weibel 96; Ruas 00]. It must be noticed that the problems are not explicitly described in the workflow. The intentions of the cartographers need to be reformulated in a way to describe the problems and the solutions. For instance, if a group of rivers is simplified with *Douglas and Peucker* algorithm, we reformulate this fact to emphasise the reason why this has been done. The proposed reformulation using the notion of problem and operations is: there are in the initial database some objects (the rivers) having a *geometric shape which is too detailed* to be in the final database; and the solution proposed by the cartographers to solve this problem is to *simplify* them (using Douglas and Peucker algorithm). By this way, it is possible to note the problem *geometric shape too detailed* and its solution *simplification*. This kind of knowledge is reusable because in another case of generalisation we suppose that there will still have groups of objects with a geometric shape too detailed and that the solution will still be valid.

In the next sections we will detail the types of problems that we identified, the types of operations selected to solve them, and finally we will associate the problems to the operations.

3.3 Problems and operations

In order to reformulate the workflow into a set of problems linked to a set of operations, one must first identify generic problems and generic operations.

First, the workflow enables us to distinguish two types of problems: problems on one object and problems on one group of objects. More precisely, we decomposed problems on one object into problems related to the object itself and problems related to the relations that this object has with other objects. In particular, we noticed that numerous problems did imply a particular relation: the 'support relation' between one object and another object that must exist to ensure the coherence of the geographic space (e.g. if a house is supported by an island, the house must be removed if the island is removed).

Figure 3 presents a more detailed classification of the problems found in MRN's workflow.

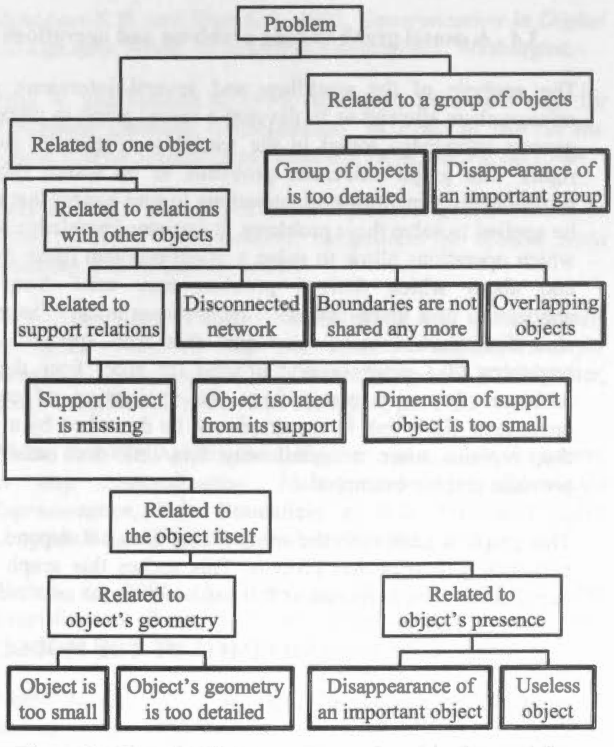

Problem
Related to a group of objects
Related to one object
Group of objects is too detailed
Disappearance of an important group
Related to relations with other objects
Related to support relations
Disconnected network
Boundaries are not shared any more
Overlapping objects
Support object is missing
Object isolated from its support
Dimension of support object is too small
Related to the object itself
Related to object's geometry
Related to object's presence
Object is too small
Object's geometry is too detailed
Disappearance of an important object
Useless object

Figure 3: Classification of problems found in the workflow

Second, we classified the operations made in the workflow. Many classifications of generalisation operations do exist (see for example [McMaster and Shea 92; Peng and Tempfli 96]). For our purpose, we used the classification described in Figure 4. One can notice that the classification follows the same principle than the one used to classify the problems. Operations are classified into two categories: operations on an object and operations on a group of objects. These categories are further divided into subcategories which are associated to typical generalisation operations: simplification, reduction, deletion, etc.

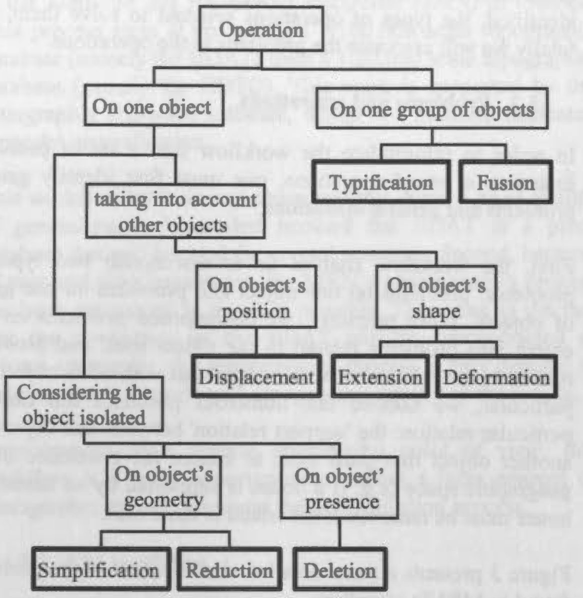

Operation
On one object
On one group of objects
Typification
Fusion
taking into account other objects
On object's position
On object's shape
Displacement
Extension
Deformation
Considering the object isolated
On object's geometry
On object' presence
Simplification
Reduction
Deletion

Figure 4: Classification of operations made in the workflow

3.4 A causal graph linking problems and operations

The analysis of the workflow and several interviews with cartographers allowed us to develop a causal graph to represent generic knowledge found in the workflow (Figure 5, on the right). The graph associates problems to be solved (square boxes) to the generalisation operations (round boxes) that must be applied to solve these problems. It contains knowledge about which operations allow to solve a given problem (dash lines), and about which derived problem may arise from the application of a given operation (continuous lines). The graph also distinguishes initial problems that may appear at the beginning of a generalisation process (in bold) from derived problems that may result from the application of certain operations. Each link in the graph can be described by a form that explains more in detail why this link does exist and provides graphic examples.

This graph is generic in the sense that it does not depend on a particular generalisation process. This makes this graph very useful for several purposes, as it is explained in the next section.

4. UTILITY AND LIMITS OF THE MODEL

4.1 Utility of the causal graph

The first utility of the graph appears when one wants to analyse a special process such as the MRN's workflow. One can do that by associating each step of the process to the path of the graph it corresponds to.

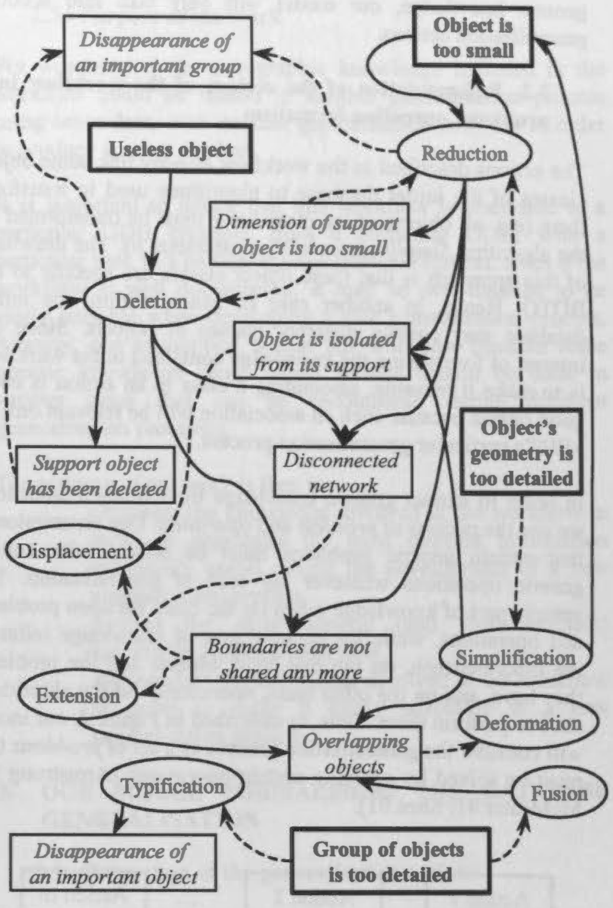

Disappearance of an important group
Object is too small
Useless object
Reduction
Dimension of support object is too small
Deletion
Object is isolated from its support
Object's geometry is too detailed
Support object has been deleted
Disconnected network
Displacement
Boundaries are not shared any more
Simplification
Extension
Deformation
Overlapping objects
Typification
Fusion
Disappearance of an important object
Group of objects is too detailed

Figure 5: Causal graph linking problems and operations

On the one hand, this allows to distinguish the generic and context-related steps of the process. Indeed, the steps of the process that do not correspond to any path in the causal graph are not generalisation steps. We thus can distinguish steps that are actual generalisation operations and steps that are introduced to overcome the system limits (e.g. inefficient algorithms) or to deal with the initial GDB defects. On the other hand, the graph creates an organisational memory. Indeed, as we explained before, the graph is a way to extract, organise, and reformulate the generic knowledge involved in the generalisation process. In addition, the graph explains *why* these steps have been chosen (i.e. to solve which conflict), which is not stored in the raw process description. This is useful if the process has to be refined in the future or adapted to a new product.

The second utility of the graph is to help cartographers to define a new process, when there is a need for a new product. Indeed, thanks to the graph, the workflow designer is able to identify the derived problems that may result from the application of a given operator. Consequently, the number of trials before getting a satisfactory process will certainly be reduced.

Another utility of the graph is to pave the way to a more computer-assisted generalisation process. As proposed in [Weibel 91], one can imagine generalisation platforms that contain both interactive operations, automated algorithms and knowledge bases used to guide the process. Such a platform could use the causal graph as part of its knowledge base in order to propose to the cartographers a set of operations to apply in certain conditions. Of course, this would require a platform containing some tools to help the cartographers to specify their needs (constraints on the final product), some efficient algorithms to automate the generalisation operations, and some efficient measures to automatically detect the problems due to constraints violations.

4.2 Limits of the model

One may argue that it is impossible to formalise all the generalisation knowledge in the form of a set of rules "if Condition then Operation", because there exist too many spatial configurations and too many possible generalisation solutions [Beard 91].

First, one must notice that this does not contradict the fact that the formalism "if Problem then Operation" can be efficient to model the knowledge involved in a particular process, in order to make it reusable.

Second, let us remark that our model only takes into account the generalisation process to produce databases, which is less complex than cartographic generalisation aiming at producing maps, because no legibility constraint appear. The causal graph may then encompass most of the cases encountered during database generalisation.

Finally, we believe that the causal graph presented in this paper is quite generic because of its degree of generality. For example, the problem "useless object" is not decomposed into all the cases in which an object can be useless. If this should be done, the graph may become very complex. The main drawback of this genericity is that this model cannot be implemented directly to automate the generalisation process. For example, the problem named "useless objects" should be refined into a precise specification which takes into account the characteristics of the final product (e.g. power lines are useless, narrow rivers are useless, small lakes except the isolated one are useless, etc.).

5. CONCLUSION

The analysis of an actual generalisation process is a good way to acquire knowledge about the generalisation process, but one must be able to differentiate specific knowledge from generic knowledge that can be reused. Our analysis of the workflow describing MRN's generalisation process enabled us to obtain an initial model of the database generalisation process.

This model, represented by a causal graph, is a set of problems associated to operations. The introduction of problem and operation notions is a way to formalise generic knowledge that can be reused when dealing with others cases of generalisation. Indeed, because our model make explicit the problems underlying the workflow, it can used to anticipate the problems of a new workflow.

The proposed model is a good way to explain which generalisation operations should be done to solve certain problems. In order to go further, some work should be done to explicitly represent knowledge concerning how to sequence the generalisation process: when an object has several problems, which problem must be solved first? When several objects have problems, which object must be processed first?

ACKNOWLEDGEMENT

We would like to gratefully thank MRN's cartographers who have been available to explain us their workflow and to answer our questions. Thanks to the network of centres of excellence, GEOIDE, to National Defence and to Ministère des Ressources Naturelles du Québec for supporting this project.

REFERENCES

Armstrong M.P. 1991. Knowledge classification and organization. In *Map generalization: Making rules for knowledge representation,* B.P. Buttenfield and R.B. McMaster, (Eds), Longman Scientific & Technical, pp. 86-102.

Beard K. 1991. Constraints on rule formation. *Map Generalization : Making Rules for Knowledge Representation*, B.P. Buttenfield et R.B. McMaster (eds), Longman Scientific & Technical, pp.32-58.

Brassel K. et Weibel R. 1988. A review and conceptual framework of automated map generalization. *Int. Journal of Geographical Information Systems,* vol.2, no.3, pp.229-244.

Douglas D.H. et Peucker T.K. 1973. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or it's Caricature. *The Canadian Cartographer journal*, 10(2), pp.112-122.

João E.M. 1998. *Causes and Consequences of Map Generalisation*. Taylor and Francis: London.

McMaster R.B. 1991. Conceptual frameworks for geographical knowledge. In *Map generalization: Making rules for knowledge representation,* B.P. Buttenfield and R.B. McMaster (Eds), Longman Scientific & Technical, pp. 21-39.

McMaster R.B. and Shea K.S. 1992. *Generalization in Digital Cartography*. Assoc. of American Geographers, Washington.

Peng W. and Tempfli K. 1996. An Object-Oriented Design for Automated Database Generalization. In *Proc. of the 7th Int. Symposium on Spatial Data Handling*, Delft, sec.4b, pp.15-29.

Ruas A. 2000. The role of meso objects for generalisation. In *Proc. of the 9th International Symposium on Spatial Data Handling,* Pékin, Chine, sec.3b.

Shea K.S. 1991. Design considerations for an artificially intelligent system. In *Map generalization: Making rules for knowledge representation,* B.P. Buttenfield and R.B. McMaster, (Eds), Longman Scientific and Technical, pp. 3-20.

Weibel R. 1991. Amplified intelligence and rule-based systems. In *Map Generalization, Making rules for Knowledge Representation,* B.P. Buttenfield et R.B. McMaster (eds), Longman Scientific & Technical, pp. 86-102.

Weibel R. 1996. A Typology of Constraints to Line Simplification. *Proc. of 7th Int. Symposium on Spatial Data Handling*, Delft, Vol.2, sec.9a, p.1-14.

Weibel R. and Dutton G. 1999. Generalising spatial data and dealing with multiple representations. In *Geographical Information Systems: Principles and Applications, 2nd edition*, Maguire D.J., Goodchild M.F., Rhind D.W. and Longley P. (eds), Wiley, New York, Vol.1, pp.125-155.

WHAT IS SPATIAL CONTEXT IN CARTOGRAPHIC GENERALISATION?

Sébastien Mustière, Bernard Moulin

Computer Science Department and Geomatics Research Centre, Laval University, G1K 7P4 Ste-Foy, Québec, Canada
{sebastien.mustiere, bernard.moulin}@ift.ulaval.ca

KEY WORDS: Databases, Contextual, Analysis, Modelling, Cartography, GIS, Generalization.

ABSTRACT:

During the cartographic generalisation process, geographic objects cannot just be considered one by one. The way objects are processed clearly depends on their spatial context. In this paper, we first study the nature of spatial context encountered during map generalisation. We differentiate three kinds of relations that an object can have with its environment: being part of a significant group, being in a particular area, being in relation with 'same-level' surrounding objects. We also address the issue of scale-dependency related to spatial context. Then, we study how to represent context-related knowledge in a knowledge-based approach of cartographic generalisation. We discuss several aspects of the representation of objects' spatial context in a geographic data model dedicated to map generalisation: explicit representations of high-level objects, fuzzy representation of objects and relations, multi-scale representation of objects and relations, spatialisation of relations. We also study how context appears in the rules expressed by cartographers to describe the generalisation process: it can be used to express exceptions, to classify typical operations to be done, or to express constraints on the final result.

RÉSUMÉ:

Pendant le processus de généralisation, on ne peut pas considérer les objets géographiques un à un. Le traitement de ces objets dépend clairement de leur contexte spatial. Dans cet article, nous étudions tout d'abord la nature du contexte spatial qui influence la généralisation. Nous différencions trois types principaux de relations qu'un objet peut avoir avec son environnement: faire partie d'un groupe significatif, être dans une zone particulière, ou être en relation avec les objets aspects de même niveau. Nous abordons également un aspect important du contexte spatial: sa dépendance au niveau d'analyse. Nous étudions ensuite comment représenter les connaissances relatives au contexte spatial dans une approche à base de connaissances de la généralisation. Nous présentons diverses considérations sur ce que nécessiterait une représentation du contexte spatial dans un modèle de données adapté à la généralisation: représentation explicite d'objets de haut niveau., représentation floue d'objets et de relations, représentation multi-échelle d'objets et de relations, spatialisation des relations. Nous étudions également comment la notion de contexte apparaît dans les règles de généralisation exprimées par les cartographes: cela peut servir à exprimer des exceptions, à organiser différentes opérations typiques à réaliser dans différentes configurations, ou à exprimer des contraintes sur le résultat final.

1. THE IMPORTANCE OF SPATIAL CONTEXT

Geographic data models usually explicitly represent a set of basic objects, their geometry and their properties. But much of the geographic world's semantics appears in the relations linking objects [Worboys 96; Papadias and Theodoritis 97; Ruas 99; Mark 99]. Nevertheless, most of these relations are not explicitly represented in data models describing geographic databases. Usually, these relations only implicitly appear when one is looking at a *display* of a geographic database. This largely contradicts a principle in knowledge representation that a good representation should make the important things explicit [Winston 84].

This deficiency of geographic data models is a barrier to the analysis and derivation of geographic databases, and this is particularly the case for the cartographic generalisation process [Lagrange and Ruas 94]. Cartographers know that two similar objects located at two different places will not necessarily be generalised in the same way, even for the same target product. This is due to the (implicit) different relations that these objects have with other surrounding objects. "For example, an object can be preserved because it is a representative of a set of objects of the same nature, or because it allows a connection to objects which should be preserved […]" [Lagrange et Ruas 94]. In other words, the generalisation process is dependent on the objects' *spatial context*. Even more, "Context-related rules are probably the most significant of the 'know-how' for generalization" [Kilpeläinen 2000].

In the next section, we study the nature of spatial context encountered during map generalisation: what kinds of spatial context do exist; how does scale influence the notion of context? Then, we make different considerations on what is needed to represent objects' spatial context in a geographic data model dedicated to map generalisation.

2. DIFFERENT TYPES OF SPATIAL CONTEXT

When one reads a map, the geographic space is analysed according to different levels of analysis, from the identification of individual elements to an apprehension of the whole space, through the analysis of groups of objects. This multilevel aspect may be one of the most significant character of the geographic space [Scholl et al. 96].

This consideration is at the basis of our study of what is spatial context or, in other words, how an object is in relation to its environment. We first distinguish relations between objects of the same level of analysis, and relations between objects of different levels of analysis.

2.1 Hierarchical relations

We can find two main types of spatial relations implying geographic objects of different levels of analysis: the relations between a group and its members (e.g. an island is part of an archipelago), and the relations between a part of the space and its elements (e.g. a road is in a sparse area).

2.1.1 Being part of a significant group. A lot of geographic objects take a more precise meaning by being part of groups than on their own. For example, in certain studies an highway interchange may be more significant than the isolated road sections it contains. Certain generalisation frameworks explicitly represent significant groups and manage the interconnected generalisations of the groups and the parts. The identification of groups is seen as a "structure recognition" in the model of [Brassel and Weibel 88], and as a "meso analysis" in the AGENT model [Ruas 99; AGENT 00].

The identification of significant groups is very important during the generalisation process for several reasons [Ruas 00]. First, some generalisation operations cannot be done while looking at objects one by one and must be performed at the group level (such as merging and typifying). Second, the fact of being part of a significant group can influence the way the elements of the group are generalised (for example to insure a certain homogeneity of transformation in the group).

2.1.2 Being inside a particular area. This type of spatial context considers the property of being located inside an area that can be qualified by some global characteristics. For example a road can be inside a urban, rural, or mountainous area; or an island can be either in the middle of the see or in a coastal area.

The property of belonging to a particular area does influence the generalisation operations. It allows to better characterise the importance of objects. Consequently, it is possible to better determine which objects must be represented on the map, and how they must be represented. First, objects that are typical of an area must receive a particular attention during generalisation to efficiently reflect the global character of the area (e.g. the organisation of a street network must be well represented while selecting streets in a urban area). Second, another interest of characterising areas is to detect atypical objects which are considered as important and must be kept (e.g. a river in a desert area, an isolated village in a natural area, etc.).

2.1.3 Distinction between "member/group" and "element/part of space". The frontier between these two kinds of relations seems to be a bit fuzzy. In fact, being an element of a certain part of space can be thought of as being a member of the group made of all the elements located inside this part of the space. There is anyway two main differences between these relations, in terms of the nature of the relation, and in terms of its function.

First the "member/group" relation is a constituting relation (the group is *made of* the members). But the "element/part of space" relation is only a relation of localisation (the element is *inside* the part of the space). In addition, the "element/part of space" relation is usually more fuzzy than the "member/group" relation, because of the fuzziness of the limits of a given area (e.g. what are exactly the limits of a given urban area?).

Second, during the generalisation process, it may be important to know the exact constituents of a group (e.g. to merge individual islands into a unique object archipelago). But it is not important to know the exact limits of a part of the space, it is only important to know that a given element is inside a given part of the space (e.g. rules determining which individual houses to represent on the map may be different in rural and urban areas).

Anyway, let us notice that certain geographic phenomena may some time be viewed as a part of the space and as a group of objects during the generalisation process. For example, a urban area can be seen as the set of its houses when one wants to merge all the houses into a single polygon representing the urban area, or as a part of the space when one deals with the selection of the streets it contains.

The study of the relations between parts and wholes is a field of research in itself (namely mereology [Simons 87]), and we will not discuss it further on. Nevertheless, due to the multilevel nature of the geographic space, more studies should be done to better understand the semantics of the hierarchical relations in the geographic domain.

2.2 Non hierarchical relations

Another type of spatial context that influences the generalisation process corresponds to the local relations that a given object has with its surrounding objects at the same level of analysis (e.g. a house is *near* a road, a house is *aligned* with a road, a road *crosses* a river, a road is *parallel* to a railway, etc.). These relations can influence the generalisation process in different ways.

First, some of these relations must be kept. For example, if a road is *parallel* to a river or if a road is *crossing* a river, it may be important to keep these relations during the generalisation process.

Second, some local relations can be emphasised. For example, some close but distinct houses may be shifted from others to emphasise the *disjoint* relation. This example shows that generalisation can be seen as the process of *representing* an *abstracted* view of the world [Mustière et al. 00]. An abstraction is done when one considers that the disjoint relation is more important than the accurate location of the houses; and shifting houses is a way to represent this disjoint relation while respecting legibility constraints. Other examples where relations between objects are abstracted before being represented may be a house *nearly aligned* to a road that is moved to become *exactly aligned* with the road, or a house *close* to a road that is moved to *touch* it on the map.

It is then important to understand which types of relations may exist between objects in order to understand how they do influence the generalisation process : which relations must be kept, into which relations a given relation can be abstracted, etc.?

In the GIS community, the most studied relations are certainly topology-based relations (e.g. the 9-intersections model of [Egenhofer et al. 89]), followed by distance-based and direction-based relations. It is also often considered that "topology matters, metrics refines" [Egenhofer and Mark 95]. If topologic relation are undoubtedly important, one can moderate the predominance of topologic relations over other relations. First, a lot of relations rely on the notion of shapes and relative positions (e.g. surrounding, being parallel, being aligned, being a landmark) [Mathet 2000]. Quite few works do consider these notions, maybe because of the difficulty of identifying them contrary to topologic relations which are deterministically defined from the geometry of objects. Second, topologic relations such as adjacency must often be refined in order to better reflect significant spatial relations. "Defining connectivity solely on the basis of topological adjacency is inadequate for studying many types of natural and social processes [...].For example, water basins (represented by polygons) are functionally connected only if they are adjacent and water flows from one into another" [Theobald 2001].

It is important to go further the analysis of topology to better understand the semantics of spatial local relations. The study of the way people naturally express these relations may be a good starting point for doing it. Several models have been developed in linguistic studies to represent the cognitive principles lying under the expression of spatial relations in the natural language. Certain authors develop models where topology is the central notion and may be extended with various concepts. For example, [Sablayrolles 95] considers a second "frontier" of the objects as a key notion: the limit of a "proximity area". Other authors go beyond notions related to topology. For example, [Mathet 00] considers that the notions of shape, distance, direction, topology and projection are all important. He also considers that the polymorph aspect of an object is a key issue to understand spatial relations: a given object is either seen as a line, a surface or a point according to the type of relation being considered.

In order to better understand how to manipulate spatial relations during the generalisation process, more attention could be paid to the development of cognitive models of spatial relations, such as the linguistic studies evocated above.

2.3 A scale-dependent notion

Of course, any model of the world depends on its intended use. Thus, the relevant spatial context of an object depends on the purpose of the considered geographic data. This purpose-dependency is by nature an important point for the generalisation process that deals with two different purposes: the purpose of the initial data and the purpose of the final data. The intended scale of analysis of the data is a key aspect of the purpose of the data. The notion of scale is important in cartography because it determines the display capabilities (e.g. minimal distance between objects on the paper map). Even more, the notion of scale is important because it is related to the level of analysis of geographic data (is the analysis done at the level of the town or at the country?). Thus, the influence of the scale on spatial context must be studied.

First, the notion of scale influences which objects are considered relevant in the spatial context of an object. For example, the trails crossing a road are interesting for hikers which have a local view of the geographic space (typically, they may use 1:50,000 scale maps). But these trails are out of interest for drivers who consider space at a larger level of analysis and who may be more interested by the towns accessible by the road (that may be found on 1:500,000 scale maps for example).

Second, the scale influences the definition of the high level objects (the groups and the areas) to be considered to represent spatial context well. For example, the coastal area may stop at a few hundred meters from the coast in a local analysis of marine plants, but it may stop a few miles from the coast in a global analysis of whales' displacement. As another example, let us consider a house which is located in an isolated village, the latter being located in a sparse country. At a large scale the house can be considered as being located in a dense built-up area (the village) comparatively to the houses located in the village's surroundings; but at a lower scale it may be considered as one of the buildings of a sparse built-up area (the countryside).

Finally, the semantics of the relation between different objects may change according to the scale of analysis. For example, a river and a road may be quite parallel from a global point of view but, from a more local point of view, may be not parallel at all because of the sinuous shape of the river. As another example, if a clearing is inside a forest, the meaning of the *inside* relation may vary according to the level of analysis: at a very local scale the clearing can be considered as a frontier of the forest (where the clearing is, there is no tree), and at a larger scale the clearing is considered to be part of the forest (the forest is made of the wooden areas and the clearings).

3. REPRESENTING SPATIAL CONTEXT AND CONTEXT RELATED RULES

3.1 Tool-box orientation vs. explicit representation of spatial context

Spatial context must be taken into account during the generalisation process. Two main approaches can be used to do that.

First, context can be seen as part of the process: it is analysed inside generalisation algorithms. This is the "tool-box metaphor" used in most of the works related to spatial analysis [Egenhofer and Kuhn 99]. In this case, spatial context of objects is not explicitly represented in data models, but it influences the processes used to generalise data. For example, let us consider an algorithm that takes a mechanical paradigm and considers geographic objects as producing forces pushing away other objects with a strength depending of the distance between them: objects will be moved one from the others without explicitly representing any proximity relationship.

Second, context can be seen as part of the data. In this case, the context is explicitly represented in the data model (as the objects are). In this approach, knowledge describing spatial context and knowledge describing how to use it are separated. It must be noticed that representing relations does not necessarily mean that all of them must be pre-computed and stored, it can be better to compute them on-line when needed [Theobald 2001].

Due to the complexity of the geographic world, it may be sometimes a hard work to explicitly represent all the relevant spatial relations useful for a certain purpose. In these cases, it may be better to consider certain parts of the context inside the algorithms transforming the data. But, representing context as an explicit part of the data may be better from a cognitive point of view if we agree that "spatial relations among spatial entities are as important as the entities themselves" [Papadias and Theodoritis 97]. This allows to develop knowledge-based approaches for generalisation, where the knowledge describing how to manipulate the data is separated from the knowledge describing the data (which can thus be easily reused for other applications). The next section explores how the spatial context and context-related rules could be represented in knowledge-based approaches.

3.2 Representing spatial context

In the second section we raised several points about what is spatial context. These considerations have some consequences on how spatial context could be modelled. Proposing a complete model of spatial context representation goes beyond our current work, but we mention in this section some requirements for spatial context modelling.

Representing high-level objects

The generalisation process requires the manipulation of high-level objects such as groups or areas (cf. section 2.1). These objects have relations with basic objects in the database, and are in relation to each other (two archipelagos can be close from each other, an archipelago can be part of a country). If one wants to consider these objects during the generalisation process, they should be explicitly represented [Lagrange and Ruas 94]. There can be several levels of high-level objects (groups, groups of groups, etc.) and these high-level objects dot not necessarily represent a hierarchical partition of the space (there can exist two overlapping groups of aligned houses, rural and urban areas have fuzzy boundaries that can overlap) [Ruas 99; Ruas 00].

Representing fuzzy concepts

One must be able to represent some relations between certain objects and the area they belong to (cf. section 2.1.2). The relationship between a geographic entity (e.g. a house) and an area it belongs to (e.g. a rural area) can be modelled in different ways in an object-oriented paradigm. It can be modelled as a relation between an object (the house) and another object which has a *fuzzy boundary* (the rural area). It can also be modelled as a *fuzzy relation* between an object (the house) and another object which has no explicit localisation in space (the rural area). Then, it can be modelled as an object (the house) with a *fuzzy attribute* (the attribute "is in" with the value "rural area" possibly with a fuzzy characterisation such as "certainly"). Whatever the choice that is made, an adequate representation of the "element/part of space" relation should model some fuzzy concepts: either objects with fuzzy boundaries, fuzzy relations between objects, or fuzzy attributes [Winter 2000].

Dealing with multi-scale representations

Spatial context is a scale dependent notion (cf. section 2.3). As the generalisation process considers at least two levels of analysis (the final and initial scale), it is necessary to be able to have a multi-scale representation of spatial context. For example, if the relation *is parallel to* is characterised by different properties (e.g. degree of parallelism, place where the two objects are parallel), there should be different instances of this relation at different scales.

Characterising the relations

Spatial relations between objects at the same level of analysis have to be represented (cf. section 2.2). Here are some requirements to represent these relations well: we should be able to characterise them by some quantitative or qualitative attributes (the house is *almost* aligned with the road); we should be able to spatialise them (the road is parallel to the river *in this area*) like it is proposed in the MADS [Parent et al. 98] or Perceptory [Brodeur et al. 99] models; we should be able to represent more than topologic relations.

3.3 Representing context-related rules in knowledge-based approaches

Spatial context is taken into consideration in numerous generalisation algorithms. For example, an algorithm dedicated to the selection of roads in a road network may be guided by the analysis of the other spatial objects accessed through this network (e.g. the towns or the houses depending on the level of analysis) [Richardson et Thomson 96 ; Morrisset et Ruas 97]. These algorithms usually implement some cartographic operations while respecting a set of cartographic rules (e.g. keep highways, keep roads necessary to access to main towns, etc.).

Context-related rules can also be explicitly represented in knowledge-based approaches used to combine numerous generalisation operations. Several knowledge-based models do exist (e.g. expert systems, case base reasoning). Identically, cartographers can express the influence of context on map generalisation using different types of cartographic rules, and each of these rules can be more directly represented using a particular model.

Context can be used as a mean to express exceptions (e.g. "remove all houses except the isolated one"). This use of context is particularly adapted to a rule-based representation of cartographic knowledge, where exceptions to the rules contain context-related terms (e.g. [McMaster and Shea 92]).

Context can also be used to classify the different actions to be performed (e.g. "if the island is in the middle of sea, transform it using *operation1*; if it is on a river, transform it using *operation2*; if it is on a lake…"). This use of context is adapted to a rule-based or case-based (e.g. [Keller 94]) representation of cartographic knowledge, where each rule or each case describes a certain context.

Context can also be used to express constraints on the expected result (e.g. "the possible parallelism relation between the road and the river must be kept" or "the river must remain in the valley"). These cartographic rules are more easily expressed in a constraint-based formalism (e.g. [Beard 91; Ruas 00]).

4. CONCLUSION

During the cartographic generalisation process, geographic objects cannot just be considered one by one. The way objects are processed clearly depends on their spatial context. In this paper, we studied the nature of spatial context encountered during map generalisation. We differentiated three types of relations that an object can have with its environment: being part of a significant group, being in a particular area, being in relation with 'same-level' surrounding objects. We also addressed the issue of scale-dependency of spatial context. This allowed us to draw certain requirements necessary to represent the spatial context of objects.

This work is a step towards a better definition of what is spatial context. The interest of better understanding the spatial context of geographic objects is not limited to cartographic generalisation. Most geographic analyses do consider the context of objects (why did these phenomena appeared in certain context? what would happen to this phenomenon if the context is changed? etc.). Objects of the geographic world are highly interrelated. It is thus impossible to explicitly represent in data models all the significant relationships between geographic objects. Anyway, we believe that a better modelling of spatial context could pave the way to the creation of richer and more effective geographic data models.

ACKNOWLEDGEMENT

Thanks to the network of centres of excellence, GEOIDE, to National Defence and to Ministère des Ressources Naturelles du Québec for supporting this project.

REFERENCES

AGENT 00. European research and development project December 1997- December 2000; ESPRIT 24939. Cf. http://agent.ign.fr

Beard K. 1991. Constraints on rule formation. In *Map Generalization : Making Rules for Knowledge Representation*, Buttenfield et McMaster (eds), Longman Scientific & Technical, Harlow, Essex, pp.32-58.

Brassel K. and Weibel R. 1988. A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Systems*. Vol. 2, no. 3, pp.229-244.

Brodeur, J., Bédard, Y. and Proulx, M.-J.., 2000, Modelling Geospatial Application Database using UML-based Repositories Aligned with International Standards in Geomatics, In *Proc. of ACMGIS 2000*, November 10-11, Washington DC, USA

Egenhofer M.J., Frank A. and Jackson J.P. 1989. A Topological Data Model for Spatial Databases. In *1st Int. Symp. on the Design and Implementation of Large Spatial Databases*, LNCS 409, pp.271-286. Springer-Verlag.

Egenhofer M.J. and Mark D.M., 1995. Naive geography. Spatial Information Theory. *Proc. Of COSIT'95, Lecture Notes in Computer Science 98*, Springer Verlag, Berlin, p.1-15.

Egenhofer M.J. and Kuhn W. 1999. Interacting with GIS. In *Geographical Information Systems: Principles and Applications, 2nd edition*, Maguire D.J., Goodchild M.F., Rhind D.W. and Longley P. (eds), Wiley, New York, pp. 401-412.

Keller S. 1994. On the use of case-based reasoning in generalization. *Proc. of 6th International Symposium on Spatial Data Handling*, Edimbourg, Vol. 2, pp. 1118-1132.

Kilpeläinen T. 2000. Knowledge Acquisition for Generalization Rules. *Cartography and Geographic Information Science*, vol.27, n.1, pp.41-50.

Lagrange J.-P. and Ruas A. 1994. Geographic information modelling: GIS and generalisation. In *Proc. of Int. Symposium on Spatial Data Handling*, Edinburgh, pp1099-1117.

Mark, D. M., 1999. Spatial Representation: A Cognitive View. In *Geographical Information Systems: Principles and Applications, 2nd edition*, Maguire D.J., Goodchild M.F., Rhind D.W. and Longley P. (eds), v. 1, pp. 81-89.

Mathet Y., 2000. New paradigms in space and motion: A model and an experiment. In *Proc. of ECAI Workshop on spatio-temporal reasoning*, Berlin, 2000.

McMaster R.B. and Shea K.S. 1992. *Generalization in Digital Cartography*. Assoc. of American Geographers, Washington.

Morrisset B. and Ruas A. 1997. Simulation and Agent Modeling for Road Selection on Generalisation. In *Proc. of Int. Cartographic Conference*, Vol.3, pp.1376-1380.

Mustière S., Zucker J.-D. and Saitta L. 2000. An Abstraction-Based Machine Learning Approach to Cartographic Generalisation. *Proc. of 9th International Symposium on Spatial Data Handling*, Pékin, sec.1a, pp.50-63.

Papadias D. and Theodoridis Y. 1997. Spatial relations, minimum bounding rectangles, and spatial data structures. *International Journal of Geographical Information Science*, vol.11, n.2, pp.111-138

Parent C., Spaccapietra S., Zimanyi E., Donini P., Plazanet C., and Vangenot C.. Modeling Spatial Data in the MADS Conceptual Model. In *Proc. of the 8th Int. Symp. on Spatial Data Handling*, Vancouver, Canada, July 1998.

Richardson D.E. and Thomson R.C. 1996. Integrating Thematic, Geometric and Topologic Information in the Generalization of Road Networks. *Cartographica*, Vol.33, n.1, pp.75-83.

Ruas A. 1999. *Modèle de généralisation de données géographiques à base de contraintes et d'autonomie*. Ph.D. Thesis, university of Marne-la-Vallée, France (in French).

Ruas A. 2000. The role of meso objects for generalisation. *Proc. of the International Symposium on Spatial Data Handling*, Beijing, China, sec.3b.

Sablayrolles P. 1995. *Sémantique formelle de l'expression du mouvement. De la sémantique lexicale au calcul de la structure du discours en français*. PhDhesis, Univeristy Paul Sabatier, Toulouse (in French).

Scholl M., Voisard A., Peloux J.-P., Raynal L., Rigaux R. 1996. *SGBD Géographiques - Spécificités*. International Thomson Publishing, France (in French).

Simons P. 1987. *Parts, A Study in Ontology*, chapter 1-3, pages 5-128. Oxford, Clarendon Press.

Theobald D.M. 2001. Topology revisited: representing spatial relations. *International Journal of Geographical Information Science*, 2001, vol.15, no.8, pp.689-705

Winston P.H., 1984. The Key Role of Representation. In *Artificial Intelligence, second edition.*

Winter S., 2000. Uncertain topological relations between imprecise regions. *International Journal of Geographical Information Science*, vol.14, n.5, pp.411-430

Worboys M.F., 1996. Metrics and topologies for geographic space. In *Advances in GIS research II : proc. of 7th International Symposium on Spatial Data Handling*, Kraak and Molenaar (eds), Taylor and Francis, p.365-375.

SEMANTIC SIMILARITY EVALUATION MODEL IN CATEGORICAL DATABASE GENERALIZATION

Liu Yaolin [1], Martin Molenaar [2], Menno-Jan Kraak [2]

[1] School of Resource and Environmental Science
Wuhan University, Wuhan, 430070, P.R. China
[2] Department of Geo-Informatics
International Institute for Aerospace Survey and Earth Science
7500 AA, Enschede,The Netherlands
Email {yaolin, Molenaar, Kraak}@itc.nl}

Commission IV, WG IV/3

KEY WORDS: Hierarchical structure, Similarity Model, Categorical Database, Classification System,

ABSTRACT

Database generalization process will be used to derive a new database with less detail for some application purposes from a single detailed database. In a database generalization process, semantic similarity measures among objects and among object types play a key role in object aggregation process. In this paper a generalization procedure will be developed based on object aggregation. The decision to aggregation objects will be based on the geometric properties of the objects, on their spatial relationships and on their thematic similarity. Normally, this similarity among objects acts as decisive rule that controls the generalization operations. This paper focuses on semantic similarity evaluation model for categorical database generalization. It presents a semantic evaluation model after reviewing current similarity models. The models are based on classification hierarchies, set theory, and attribute structure of classes. It is a computation model and can not only be used to compute the similarity between objects at same level, or at different levels, but also the similarity between object types at the same level and at different level.

1. INTRODUCTION

The similarity measures among objects and among object types play a key role in object aggregation in database generalization. The aggregating two objects not only depends on the geometric properties of the two objects, but also the thematic properties of the two objects. Using Set theory, Tversky (1977) defines a similarity measure in terms of matching process. This measure produces a similarity value that is not only the result of the common, but also the result of the different characteristics between objects, which is in agreement to an information theory definition of similarity (Lin,D 1998, Rodriquez and Egenhofer 1999, Bishr, 1997, Chakroun et al,2000, Rodríguez and M. Egenhofer, 1999, Rodríguez, M. Egenhofer and Rugg, 1999). Although many models of similarity are defined in the literature, the similarity measure method based on set theory, hierarchy structure and attribute structure of class is still lack. This paper mainly discusses the similarity measure method in the database generalization used set theory, hierarchy structure and attribute structure of class. The paper is organized as following. First, the brief review is given; secondly, the basic concepts for similarity measure are given and followed by the computing similarity model is proposed; finally an example is used to test the model.

2. BASIC CONCEPTS

Some basic concepts must be clear before discussing similarity computing model.

2.1 Class, Object Type and Object

In this paper, the class and object type have the same meaning. A class or object type determines a set of attributes to form its attribute structure. Each class c_j or object type c_j has its own attribute structure *List* (c_j) as following:

$$List(c_j)=\{A_1,\ldots A_i,\ldots A_n\}$$

A_i denotes one of the attributes of class c_j. Each attribute will have a name, a domain that will be specified by defining the range of the attribute values and scale type of the domain which indicates whether these values are from a nominal, an ordinal, an interval or ratio scale. An attribute A_i can be specified by a three tuple (Molenaar 1998):

$$A_i = \{NAME(A_i), SCALETYPE(A_i), DOMAIN(A_i)\}$$

For each object which belong to class c_j, the attribute structure defined by a class specifies the description structure of the object. A value is assigned to every attribute in the attribute structure of object. For any object, its direct class is unique and lowest in a classification hierarchy since different classes have different attribute structures. These values must fall within the range of the attribute domain, which must be defined prior to the actual assignment of attribute values.

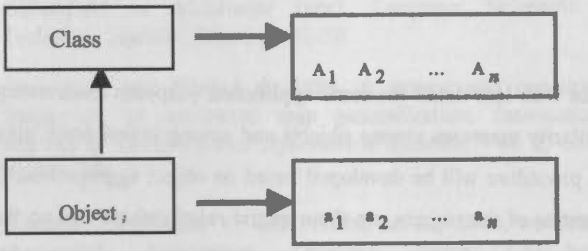


Figure 1 Diagram representing the relations between objects, classes and attributes (after Molenaar 1998)

Each object has an attribute structure list containing one value a_j for every attribute of its class. The thematic description of an object can now be specified by its class (which specifies the attributes of the object) together with the list of attribute values.

2.2 Intension and Extension of A Class

The intension of a class c_j can be expressed as a condition for the value of a combination of attributes. The condition specifies a subset of the set containing all the possible combinations of the values of the attributes, denoted as $Int(c_j)$, while the set of all the objects that belong to c_j with the same attribute structure is commonly identified as the extension of the class c_j, denoted as $Ext(c_j)$.

2.3 Formalizing Classification Hierarchy and Aggregation Hierarchy

The class hierarchy has been studied for many years in databases and knowledge bases, especially in relation to data abstraction and generalization (Smith and Smith 1977, Yee,Leung, Kwong,S.L. nad He J.Z 1999, Molenaar 1996). However there are still lack of formalization of classification hierarchy and aggregation hierarchy based on Set theory and properties of class. In the following part, the formalizations of classification hierarchy and aggregation hierarchy are discussed.

The object types in a geo-spatial model are normally determined by classification hierarchy and aggregation hierarchy. Changing classification hierarchy and aggregation hierarchy will results in changing geo-spatial model associated with database, and in turn changing the contents of the database. Changing the attribute structure and extension of classes at different level in classification hierarchy and aggregation hierarchy associated with a database will induce a new classification hierarchy and aggregation hierarchy and define a new data model of database and rebuild the corresponding contents of database.

Let S be the set of objects $\{o_1, o_2, o_3, \ldots, o_i\}$ in space, U be the set of classes $\{c_1, c_2, c_3, \ldots, c_j\}$ for the database space. Before formalizing classification and aggregation hierarchy, the relations between classes must be identified.

Let $c_i, c_j \in U$ be two arbitrary spatial classes, there should be no objects that belong to the extensions of the two different classes of U. Based on the definition of classification hierarchy last chapter and relations among classes, we can formalize classification hierarchy with intension and extension of the classes.

- $c_i \; \psi \; c_j$, only if $Ext(c_i) \cap Ext(c_j) \neq \phi$;(Relations of consistent among classes symbolized by ψ), (c_i is consistent with c_j);
- $c_i \equiv c_j$, only if $Ext(c_i)$=$Ext(c_j)$, denoted as $c_i = c_j$; (Relations of equivalence among classes symbolized by $\equiv$),($Ext(c_i)$ of c_i is equal to $Ext(c_j)$ of c_j and c_i is identical to c_j);
- $c_i \subseteq c_j$, only if $Ext(c_i) \subseteq Ext(c_j)$, denoted as $c_i \leq c \; c_j$;(Relations of inclusion among classes symbolized by $\subseteq$), ($Ext(c_i)$ of c_i include $Ext(c_j)$ of c_j and c_i belongs to c_j) or

- $c_i \subset c_j$, if $\mathrm{Ext}(c_i) \subset \mathrm{Ext}(c_j)$, denoted as $c_i < c\ c_j$. We also call c_i a sub-class of c_j and c_j a super-class of c_i. (Relations of complete inclusion among classes symbolized by $\subset$), ($\mathrm{Ext}(c_i)$ of c_i completely include $\mathrm{Ext}(c_j)$ of c_j and c_i completely belongs to c_j).
- $c_i \neq c_j$, only if $\mathrm{Ext}(c_i) \cap \mathrm{Ext}(c_j) = \phi$ (Relations of disjunction among classes symbolized by $\neq$), (there is no common object between c_i and c_j, and c_i is different from c_j).

A class c_i is included by a class c_j if and only if the extension of c_i is subsumed by the extension of c_j. We call c_i 'IS-A' c_j.

Obviously, $\leq c$ is a partial binary relation on U, called the 'Belong to' relation, and (U, $\leq c$) is a partially ordered set that we call a classification hierarchy. The process of formalization of classification hierarchy depicts how the object types (classes) and super object types can be formed into a hierarchical structure. For creation of a new super object type in the classification hierarchy there will be:

- If any A, B$\in$ H and no D$\in$ H satisfying Ext(D) =Ext(A) $\cup$ Ext(B),
  then generate such a class D, and let
  Ext(D) =Ext(A) $\cup$ Ext(B), LIST(D) =LIST(A) $\cap$ LIST(B) and D$\in$ H,
  noted as IS-A links;

The upward connections from objects to classes and classes to super classes are is-a links, which express that an object is an instantiation of a class and that a class is a special case of more general super-class. At each level, the classes inherit the attribute structure of their super-classes at the next higher level and propagate it normally with an extension to the next lower level. At lowest level in the hierarchy are elementary objects (Molenaar 1998).

2.4 Hierarchic Semantic Similarity Matrix

For a hierarchical structure as shown in Figure 2. a semantic similarity matrix as shown in Table 1 can be defined based on the properties of the hierarchical structure. A, B and C in Figure 2 represent the different branches in the hierarchical structure. For later use, they are called sub tree. T in the same Figure is called the top of the structure and c_i are object or object type.

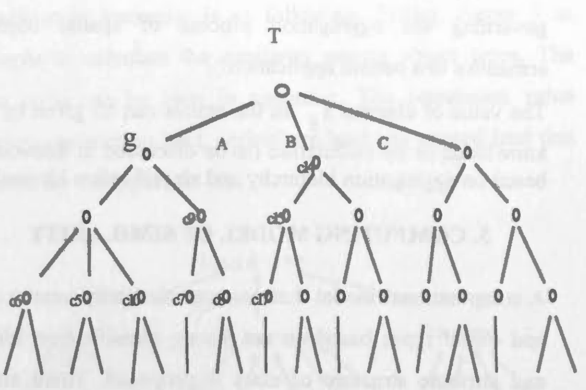


Figure 2 Example of a hierarchical structure

The semantic similarity matrix represents the similarity between object types.

Table 1 Example of semantic similarity matrix

SIMILARITY	Sub-type1	...	type1	type2	...	Sup-type1	..
Sub-type1	s_{11}	...	s_{14}	s_{15}	...	s_{17}	..
Sub-type2		...	s_{24}	s_{25}	...	s_{27}	..
...		...	...	...	...	...	..
type1			s_{44}	s_{45}	...	s_{47}	..
type2				s_{55}	...	s_{57}	..
...					...	...	..
Sup-type1						s_{77}	..
...							..

Where:

sub-type1 etc denote different elementary object types;

type1,type2 denote different object types;

sup-type1etc denote (super) composite object type;

s_{ij} denotes similarity value among object types.

The larger the value of an element in the matrix is, the more the similarity between two object types that the element links is. The matrix is symmetric and reflexive one, and has the property that s_{ij} is equal to s_{ji} ($s_{ij} = s_{ji}$) and s_{ii} is equal to

s_{jj} ($s_{ii} = s_{jj} = 1$) in the matrix. s_{ji} is a value between o and 1.

This matrix shows the similarity among different levels of object types. It will provide potential possibility to chose objects of different types to be merged or aggregated. The similarity matrix will be used as a look-up table for guiding or governing the aggregation process of spatial objects in semantics to a certain application.

The value of element s_{ij} in the matrix can be given by expert knowledge or by calculation (to be discussed in Section 6.3.2) based on aggregation hierarchy and classification hierarchy.

3. COMPUTING MODEL OF SIMILARITY

A computational model that assesses similarity among objects and object types based on set theory, classification hierarchy and attribute structure of class is proposed. There are three distances. One is the distance (number of the link edges) from immediate super object type that subsumes c_i and c_j to the top of a hierarchy such as object type g to top of the tree T in Figure 2 which represents common part of attribute structure between two object types c_i and c_j. Another distance (number of the link edges) from immediate super object type that subsumes c_i and c_j to c_i such as object type g to c_1 in Figure 2 which represents the different part of attribute structure between object type c_i and c_j ($|c_i - c_j|$). And the third is the distance (number of the link edges) from immediate super object type that subsumes c_i and c_j to c_j such as object type g to c_5 in Figure 2 which represents the different part of attribute structures between object type c_i and c_j ($|c_j - c_i|$), The proposed model is shown in Equation 2. It suits for two cases. One is for two given objects or object types belonging to the same sub tree such as sub tree A in Figure 2 and the other is for two given object types belonging to two different sub tree such as A and B as shown in Figure 2. For the first case, the model uses two types of distances to define the common and difference properties between the given object types. One is the distance between given objects or object types and immediate super object types that subsumes them which reflects difference properties between two given object types and the other is the distance between immediate super object types that subsumes two given object types and the top of hierarchical structure which reflects the common properties of two given object types. For the second case, the distance between immediate super object types that subsumes two given object types and the top of hierarchical structure will be zero since the two given object types belong to different sub tree such as c_1 and c_{15} in Figure 2. So this distance will be replaced by the correlation value between two sub trees in the Equation 2.

$$s_{ij}(c_i, c_j) = \begin{cases} \dfrac{l}{l + \alpha(c_i, c_j) * d_{ci} + (1 - \alpha(c_i, c_j)) * d_{cj}} \\ \\ \dfrac{\beta}{\beta + \alpha(c_i, c_j) * d_{ci} + (1 - \alpha(c_i, c_j)) * d_{cj}} \end{cases} \tag{1}$$

Where:

l: the shortest distance (number of the link edge) from immediate super object type that subsumes c_i and c_j to the top of a hierarchy;

d_{ci}: the shortest distance (number of the link edges) from immediate super object type that subsumes c_i and c_j to c_i ;

d_{cj}: the shortest distance (number of the link edges) from immediate super object type that subsumes c_i and c_j to c_j ;

α: a function of the distance (number of the link edge) between immediate super object type that subsumes c_i and c_j to the class c_i and c_j .

β: correlation degree among different sub-trees, such as similarity among agriculture land use, forest land use and building up land use, and its value can be given by experts based on application requirement.

A natural approach to comparing the degree of generalization between object types is to determine the distances from these object types to the immediate super object types that subsumes them in a classification hierarchy as shown in Figure 1, that is, their least upper bound in a partially ordered set. In a sense, the difference in the distances from these object types to the immediate super object types that subsumes them in a classification hierarchy reflects the difference in attribute structure between two object types. The $\alpha(c_i, c_j)$ can be expressed as a function of the distance d_{ci} and d_{cj}. In order to get final values of α, the function (Equation (2)) is defined as following:

$$\alpha(c_i, c_j) = \begin{cases} \dfrac{d_{ci}}{d_{ci} + d_{cj}} \\ 1 - \dfrac{d_{ci}}{d_{ci} + d_{cj}} \end{cases}$$

where:

d_{ci}: the shortest distance (number of the link edges) from immediate super object type that subsumes c_i and c_j to c_i ;

d_{cj}: the shortest distance (number of the link edges) from immediate super object type that subsumes c_i and c_j to c_j.

This similarity function yields values between 0and 1. The extreme value 1 represents the case that the two entity classes are completely the same, whereas the value 0 occurs when the two entity classes are completely different.

4. EXAMPLE

An example for computing element of similarity matrix from classification hierarchy is as following: Taking Figure 3 as example to calculate the similarity among object types. The class code can be seen in appendix. The correlation value among construction land, agriculture land and unused land that is given by the experts is 0.5.

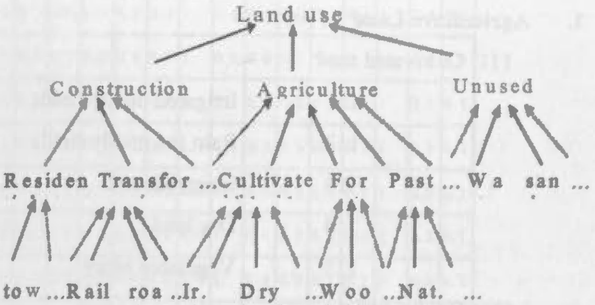


Figure 3 Land use classification hierarchy before generalization

The computing result of similarity among object types based on the model (see in Table 2.)

5. CONCLUSIONS

The computing model that has been proposed has taken set theory, classification hierarchy and attribute structure into account. It can be used to measure the semantic similarity among object types in classification hierarchy. The example shows that the computing result of the model are reasonable and efficient. How to decide the correlation parameter β will still need to research in the future work.

The authors would thanks the financial support from ITC, Ministry of Education,P.R. China and State Bureau of Surveying and Mapping.

6. REFERENCES

Bishr, Y., 1997, Semantic aspects of interoperable GIS. ITC, Publication No. 56, Enschede: International Institute for Areospace Survey and Earth Sciences (ITC).

Chakroun,H., Benie,G.B., Oneill,N.T., and Desilets,J. 2000, Spatial analysis weighting algorithm using Voronoi diagrams. International Journal of Information Science, Vol.14,No.4 pp.319-336.

Lin, D. 1998,An Information theoretic definition of similarity (eds.) International conference on amchine learning, ICML'98.

Molenaar, M., 1998,An Introduction to the Theory of Spatial Object Modelling for GIS, Taylor & Francis Ltd. London.

Molenaar, M.,1996, The Role of Topologic and Hierarchical Spatial Object Models in Database Generalization, In Molenaar,M.(Eds.) Methods for the Generalization of Geo-databases, Delft: Netherlands Geodetic Commission, New Series, Nr 43,pp.13-36.

Rodríguez and M. Egenhofer, 1999, Putting similarity assessment into context: matching functions with the user's intended operations. *Modeling and Using Context*, CONTEXT'99, Trento, Italy. In: P. Bouquet, L. Serafini, Patrick Brézillon, Massimo Benerecetti, and Francesca Castellani (eds.), *Lecture Notes in Computer Science*, Vol. 1688, Spiringer-Verlag, pp. 310-323,.

Rodríguez, M. Egenhofer, and R. Rugg , 1999, Assessing semantic similarity among geospatial feature class definition Interoperating Geographic Information Systems, Second International Conference, INTEROP'99, Zurich, Zwitzerland. In: . Vckovski, K. Brassel, and H.-J. Schek (eds.), *Lecture Notes in Computer Science*, Vol. 1580, Springer-Verlag, pp. 189-202,.

Smith, J.M. and Smith, D.C.P, 1977, Database abstractions: aggregation and generalization, ACM Transactions on Database System, 2, pp.105-133

Tversky,A. 1977, Features of similarity. Psychological review, 84(40:PP.327-352).

Appendix

Land use classification (code)

1. **Agriculture Land** (100)

11□ Cultivated land

111	Irrigated paddy fields
112	Rain fed paddy fields
113	Irrigated land
114	Dry land
115	Vegetable plots

12□ Garden Land

121	Orchards
122	Mulberry fields
123	Tea fields
124	Rubber plantation
125	Other

13□ Forest

131	Wood land
132	Shrubbery land
134	Young forestation land
135	Slashes

15□ Water area

151	Rivers
153	Reservoir
154	Pond
156	Beaches and flats
158	Hydraulic building

2. **Construction Land (200)**

21 Residential quarters and industrial and mining land

211	Area of cities and towns
212	Residential quarters in rural areas
213	Isolated industrial and mining land

3. **Unused lands (300)**

310	Waste lands
380	Others

Table 2 Similarity Table

S	111	112	113	114	115	121	122	123	124	125	131	132	134	135	151	153	154	156	158	211	212	213	110	120	130	150	210	310	380	100	200	300
111	1	0.7	0.7	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	1	0.4	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
112	0.7	1	0.7	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	1	0.4	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
113	0.7	0.7	1	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	1	0.4	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
114	0.7	0.7	0.7	1	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	1	0.4	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
115	0.7	0.7	0.7	0.7	1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	1	0.4	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
121	0.3	0.3	0.3	0.3	0.3	1	0.7	0.7	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	1	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
122	0.3	0.3	0.3	0.3	0.3	0.7	1	0.7	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	1	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
123	0.3	0.3	0.3	0.3	0.3	0.7	0.7	1	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	1	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
124	0.3	0.3	0.3	0.3	0.3	0.7	0.7	0.7	1	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	1	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
125	0.3	0.3	0.3	0.3	0.3	0.7	0.7	0.7	0.7	1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	1	0.4	0.4	0.2	0.1	0.1	1	0.2	0.1
131	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	1	0.7	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	0.4	1	0.4	0.2	0.1	0.1	1	0.2	0.1
132	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.7	1	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	0.4	1	0.4	0.2	0.1	0.1	1	0.2	0.1
134	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.7	0.7	1	0.7	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	0.4	1	0.4	0.2	0.1	0.1	1	0.2	0.1
135	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.7	0.7	0.7	1	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.4	0.4	1	0.4	0.2	0.1	0.1	1	0.2	0.1
151	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	1	0.7	0.7	0.7	0.7	0.2	0.2	0.2	0.4	0.4	0.4	1	0.2	0.1	0.1	1	0.2	0.1
153	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.7	1	0.7	0.7	0.7	0.2	0.2	0.2	0.4	0.4	0.4	1	0.2	0.1	0.1	1	0.2	0.1
154	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.7	0.7	1	0.7	0.7	0.2	0.2	0.2	0.4	0.4	0.4	1	0.2	0.1	0.1	1	0.2	0.1
156	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.7	0.7	0.7	1	0.7	0.2	0.2	0.2	0.4	0.4	0.4	1	0.2	0.1	0.1	1	0.2	0.1
158	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.7	0.7	0.7	0.7	1	0.2	0.2	0.2	0.4	0.4	0.4	1	0.2	0.1	0.1	1	0.2	0.1
211	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1	0.7	0.7	0.2	0.2	0.2	0.2	1	0.1	0.1	0.3	1	0.2
212	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.7	1	0.7	0.2	0.2	0.2	0.2	1	0.1	0.1	0.3	1	0.2
213	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.7	0.7	1	0.2	0.2	0.2	0.2	1	0.1	0.1	0.3	1	0.2
110	1	1	1	1	1	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.2	0.2	0.2	1	0.5	0.5	0.5	0.2	0.2	0.2	1	0.4	0.1
120	0.4	0.4	0.4	0.4	0.4	1	1	1	1	1	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.2	0.2	0.2	0.5	1	0.5	0.5	0.2	0.2	0.2	1	0.4	0.1
130	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	1	1	1	1	0.4	0.4	0.4	0.4	0.4	0.2	0.2	0.2	0.5	0.5	1	0.5	0.2	0.2	0.2	1	0.4	0.1
150	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	1	1	1	1	1	0.2	0.2	0.2	0.5	0.5	0.5	1	0.2	0.2	0.2	1	0.4	0.1
210	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1	1	1	0.2	0.2	0.2	0.2	1	0.2	0.2	0.3	1	0.2
310	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	1	0.5	0.2	0.2	1
380	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.5	1	0.2	0.2	1
100	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.3	0.3	0.3	1	1	1	1	0.3	0.2	0.2	1	0.4	0.3
200	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1	1	1	0.4	0.4	0.4	0.4	1	0.2	0.2	0.4	1	0.3
300	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.2	1	1	0.3	0.3	1

GENERALIZATION OF 3D BUILDING DATA

Frank Thiemann

Institute for Cartography and Geoinformatics, University of Hanover, Appelstraße 9a, 30167 Hanover, Germany
Frank.Thiemann@ikg.uni-hannover.de

Commission IV, WG IV/3

KEYWORDS: Three-Dimensional, Generalization, Multiresolution, Visualization, Modelling

ABSTRACT

Today, more and more (semi-) automatic tools and techniques are getting available, that allow for an efficient capture of more and more detailed 3D city models. With the availability of such data, the question rises, how these typically huge data sets can efficiently be handled, manipulated, analysed and visualized. Concerning visualization, the Level of Detail (LOD) concept from Computer Graphic provides a mechanism to only present that information, that is currently visible in an adequate resolution level. The problem, however, is how such a multi-scale representation of buildings can be automatically gained. This generation is a generalization problem. In the paper a concept is sketched, that relies on the representation of buildings in terms of Constructive Solid Geometry (CSG). Firstly, the complex building shape has to be transformed into CSG, which involves a segmentation of the building into its relevant parts. Secondly, generalization mechanisms can operate on the CSG-tree. Finally, the transition between neighbouring generalization levels is achieved using a differential scheme, where only the differences between adjacent generalization levels are stored, instead of the whole models.

1. INTRODUCTION

1.1 Motivation

Three-dimensional models of cities and landscapes are getting increasingly popular and widely used. The most important applications are models for propagation of noise, electromagnetic waves and pollutants in the air. The main driving forces for the capture of 3D city models have been telecommunication companies that need such models for the planning of the distribution of their antennas.

In addition to these calculation models, 3D city models are also used for visualizations. It is for instance possible to visualize whole towns, which is especially important for new architectural projects, that can thus be visualized in the context of the existing situation. A rather new field are 3D city information systems, where tourists can either visit a city virtually before a trip, or use the 3D-system as a guide during their trip. In addition, in car navigation, that currently relies on the visualization of 2D maps, 3D visualization are being introduced – especially in Japan: On complex road junctions, a 2D-map alone does not give an unambiguous presentation of the situation. Here, sophisticated and rich 3D-presentations of the spatial situation are helpful in order to give a more realistic impression and thus lead to a more save guidance.

There are different requirements concerning the level of detail of the 3D models: whereas 3D information systems and architecture needs detailed, photo realistic visualizations, they are lower for tourist information purposes, and still lower for simulations [Brenner 1999]. Thus, such models have to be available at different levels of detail on order to satisfy various needs. Furthermore, on-line visualization of detailed 3D datasets require a lot of rendering time. Thus typically only those levels of details are transmitted and rendered on the screen, that are currently needed for the visualization. The generation of the different levels of detail is a generalization problem.

1.2 Overview of paper

The paper reports on on-going research concerning the generalization of highly detailed 3D buildings. It presents a step-by-step procedure consisting of a segmentation of the building parts into its constituting details, that are then stored in terms of a CSG representation. These segments are subsequently generalized. The possibility of storing the different generalization levels in a multi-scale structure is described in detail.

2. BASICS AND STATE OF THE ART

2.1 Why is simplification and generalization necessary?

The aim of computer graphics is to create photo-realistic visualizations. Simplifications are needed to reduce the amount of data, so that is possible to calculate the visualization in real-time.

For surface simplification, vertex-, edge, and face-reductions are used. The user should not see the difference between the original model and the simplified one in the optical presentation. Therefore, only non-recognizable details can be omitted. By increasing the distance, an object will contract to a point before it will be invisible.

The disadvantage of photo-realistic presentations is the large number of presented details. The user is distracted from the essential parts. The possibility to recognize the parts of interest is deteriorated by the uninteresting details.

In cartography, there are additional aims. Single features must either be clearly distinguishable and visible or not visible at all.

Thus there are discrete gaps, where the representation changes from one level of detail to the next. The main characteristics of an object will be saved until it cannot be visualized anymore. There are different generalization methods: First of all, there is the possibility to select objects based on their thematic contents. Then, geometric generalization procedures are used: besides simplification there are also emphasizing, selection, displacement, aggregation and classification (e.g. [Shea & McMaster, 1989], or [Hake, Grünreich & Meng, 2002]). The main task that is solved in a cartographic presentation is the fact that important objects are clearly visualized, whereas the unimportant details are suppressed. Thus, it is possible to understand and comprehend a spatial situation at a glance, without being distracted by too many details.

2.2 Why is it necessary to generalise interactively scaleable and thematically selectable representations at all?

The question is, why to generalize interactive representations at all, as the user can influence them so easily: There is the possibility to flexibly zoom to objects as well as to select and unselect information, to inspect the interesting details.

In order to perceive an object, it must be identified and localised. Therefore, one typically gets an overview of the area in a small scale. However, without generalization the interesting objects could be already unrecognisable, since they are too small at that scale. With emphasizing and classification, this problem can be solved.

Furthermore, the user should get the ability to generate static visualizations (for instance printed maps). In this case, obviously manipulations (selection or scaling) are no longer possible – still the important information has to be visible. For the generation of such static representations and visualizations, cartographic generalization has proven to be suitable.

2.3 Multi-resolution-models

Scale dependent data has to be visualized and analysed with different scales. For each scale another degree of simplification or generalization is needed. This derivation of simplified and generalised models is typically time consuming, as in most cases it involves complex geometric operations. Therefore, the generation of the simplification is much more cost intensive than calculating the visualization. Thus in most cases, simplification and generalization cannot be done at real-time. There are approaches that however mainly aim at the transition of small scale differences [Lehto, & Kilpeläinen, 2000].

The Level of Detail concept (LOD) from computer graphics is designed to calculate and store several models of increasing simplification for every object. Depending on the scale for the visualization (or the distance of the observer from the objects) the appropriate model is selected and used and so calculation time is saved.

In the domain of visualization of 3D surfaces, mesh simplification or triangulation methods are used ([Hoppe 1997], [De Floriani & Puppo 1995], [Schmalstieg 1997], Schröder 1992], or [Heckbert & Garland, 1997] for a summary). These different techniques mainly depend on geometric properties. However, a mere geometric consideration does not respect the relative importance of objects. Similar to generalization approaches in cartography, objects have to be generalized according to their importance, not (only) with respect to their geometric size. In order to integrate semantic and geometric generalization, object-based methods are being applied, which are usually realized by knowledge-based systems [Mackaness et al. 97].

An approach that uses rules similar to cartographic generalization is presented in [Sester & Klein 1999]. Mayer [2000] uses a method from image processing – scale space theory – for the generalization of buildings. Typical approaches in building generalization design three levels of detail: the highly detailed object as first level, a bounding cuboid as a second level, followed by the centroid of the cuboid as third, most coarse level (e.g. [Coors 2001, Könniger & Bartels 1998, Kofler 1998]).

3. APPROACH FOR 3D-BUILDING SIMPLIFICATION

3.1 Scenario – working hypothesis

The approach presented in this paper is based on the following assumption: the building is described in a highly detailed form. The data are taken from construction plans or from a fine surveying. The aim is to use the data for a virtual tour through a town. The visualization is not intended to be photo-realistic, but has to comply with cartographic standards. Besides simplification by smoothing and omitting other generalization operations such as emphasizing, aggregation, classification and displacement are used. However, these operations require information about the importance of the objects. In the ideal case, this information can be taken from attributes. In most cases, however, importance has to be extracted from geometry (size, form) and topology (singularity).

Generalization of one building cannot be done without respecting its surrounding. A single building will typically be generalised in another way than buildings in low- and high-density-areas. The main constraint for all generalization approaches is to observe minimum measures and preserve important characteristics. The methods proposed in this paper are useful for the generalization of single buildings. Such generalizations are needed when going from the highest detail levels to a mid detail. For further generalizations, generalization of groups of objects – mainly aggregation – has to be taken into account, additionally.

3.2 Input data formats

There are different kinds of possibilities for the 3D modelling of buildings, e.g. boundary representation, parametric description, spatial enumeration or constructive solid geometry (CSG). The simplest way to describe a building geometry is to use boundary representation. Such a representation can be transformed from all the other modelling schemes.

The objects' boundary can be parameterised by a triangulated network. Usually, coordinated points will be connected by straight lines and form planar triangles. There is also the option of using NURBS for surface parametrization, which, however, in this case is not necessary, as most buildings are bounded by planar faces. In this way, the planar face that the building borders primarily consist of will be fractionised into several triangles, leading to many small faces.

Another way to model a building is to directly parameterise the boundary. This allows that the main characteristics of building faces (horizontal, vertical, orthogonal and coplanar) can directly be introduced in the data model [Thiemann 2001].

3.3 Procedure

The proposed procedure consists of the following steps: The building has to be segmented into its elementary parts. These parts will be generalized by distinguishing between important and unimportant structures. The results of the generalization steps are stored in a multi-scale data structure. In the following sections, these steps are described in more detail.

3.3.1 Feature detection

Chimneys, balconies, windows and doors are features that stick out of the facades, or form holes in the boundary. These features have to be extracted and separated from the main body of the building. In addition, features can also be split in a main part and feature-features. There are two different methods:

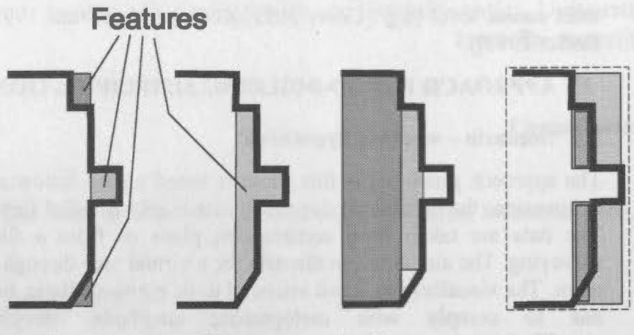


Figure 1: Two possibilities for feature detection: a: two different results of intersection with boundary planes b: Circumscription of elementary primitives: enclosing (b.1), wrapping (b.2).

a) The building is intersected with one or more planes of its boundary. This has effects for all concave and convex parts. All bumps, holes and notches will be detected. The features will be ranked in quality order [Heckbert et al. 2001] The ranking process can also take the special characteristic of buildings into account. Figure 1a.1) visualizes the effect of detecting convex and concave building structures. In Figure 1a.2) larger features are being detected. In the ranking process, the size of the added or removed parts is taken into account

b) The building is first circumscribed with enclosing or wrapping primitive solids like cylinder, prism and pyramid. Unmask building areas and mask no-building areas will be examined in next iterations. The prominent property of buildings, namely being composed of vertical walls, can easily be modelled with standing prisms. Figure 1b) demonstrates circumscription with 1b.1) only enclosing, and 1b.2) only warping primitives.

3.3.2 Segmentation into main body and features – derivation of a CSG-representation

The building will be modelled as a combination of the main body and the features. These parts are combined using Boolean operators especially union and difference and in some cases intersection. The sequence of operations is stored in a (CSG-) tree.

It is allowed that the parts (features) overlap. In this case, the Boolean operations are not reversible. Without overlapping, union and difference are reverse operations.

If $A \cup B = C$ and $A \cap B = \varnothing$,
then $C \setminus A = B$ and $C \setminus B = A$, respectively.

3.3.3 Generalization

A local and global analysis decides – depending on the scale – which parts of the building must be emphasized, merged or removed. These operations have to consider that exaggeration, aggregation and classification can lead to the necessity of displacement. The analysis will be based on given constraints concerning minimal object properties: the size of a building plane, the volume of an extrusion or notch, as well as the local neighbourhood of such phenomena influence the selection of appropriate generalization methods.

These generalizations can be executed with CSG-operations based on the following methods:

Generalization Operation	Realization with CSG
Omitting	Remove features
Displacement	Move anchor point, rotate orientation
Emphasis	Scale parameters
Aggregation:	Replacement of n features by one feature
Typification	Replacement of m features by n features
Symbolising	Is not used in this abstraction level of building generalization

Different levels of generalization can be achieved by selecting different values for the control parameters determining minimal constraints.

3.3.4 Saving the results of generalization

The result of generalization is a sequence of models for different scales. To limit the number of generated models, for larger scale ranges the same highly detailed model can be used. If only a few generalization levels are used, typically there will be many differences between these successive generalization levels of one object. While zooming in and out, details pop up or disappear abruptly when the model changes. This is known as "popping effect". Such abrupt changes attract the attention of the observer, what is considered as distracting and uncomfortable.

To avoid the popping effect, there are some strategies:

a) Many different models will be generated. Thus only few small differences between neighboured models occur.

b) With slow deformations, scalings, translations and rotations the model changes appear being continuous. Details, that will be removed, can be contracted to a geometrically lower dimension and vice versa: An edge can be contracted to a vertex, an area to an edge and a volume to an area. This is known as "geomorphs". For computer graphics; Hugues Hoppe's progressive meshes are based on this strategy [Hoppe 1996].

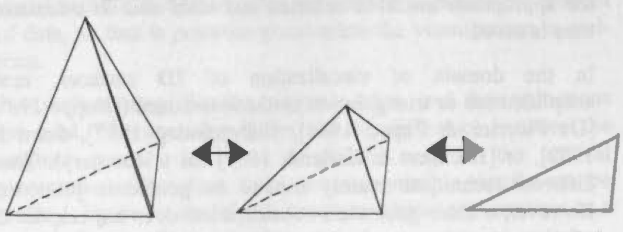


Figure 2: Geomorphing: Contracting a line, two faces and the whole tetrahedron to a triangle.

c) Typification denotes the process of replacing a number of objects by a smaller number of objects. As this is a discrete process, it is not defined which object in one representation is related to which object in the next generalization step. A seemingly continues change can be generated by cross fading using decreasing/increasing transparency of the two generalization levels involved.

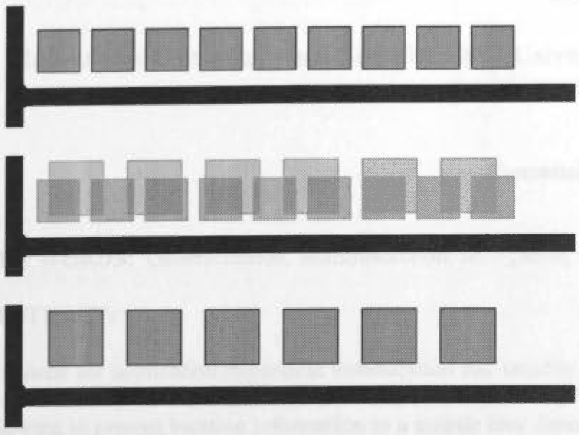


Figure 3: Fading; taken from [van Kreveld, 2001]: the nine objects from the first row are replaced by the six objects in the lower row.

Consequently, it is better to generate as much models as possible in order to achieve visually continuous changes in representation. This however leads to a huge amount of data, when each intermediate model is stored. To reduce the amount of data, instead of the whole models only the differences between them are stored. The models can be calculated by sequentially adding and deleting details, respectively. Figure 4 visualizes this concept: between generalization levels (LOD's), only the difference operations are stored (e.g. adding or deleting objects or object parts). It is assumed, that storing (small) differences between adjacent LOD's involves smaller amounts of data, and can thus be performed in real time.

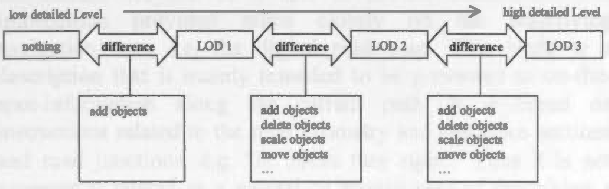


Figure 4: Differential multi-scale data structure: storing only differences between adjacent levels of detail.

If this data structure is designed from low to the high detailed level, incremental loading and transferring of the data will be possible. Data have only to be read until the requested level of detail is reached. Such techniques are well known in image processing, e.g. the progressive updating of GIF image in the Internet. They are being proposed also for vector data, e.g. by [Bertolucci & Egenhofer, 2000]. This leads to a reduced number of data to be transferred, which is especially important for mobile applications.

For visualization, a data model is needed, which enables a fast change between the neighboured models. When approaching an object, models that are more detailed are needed. When moving away, in turn, the less detailed models can be used again. The visualization of these models has to be possible in real-time. Therefore, only a boundary representation is suited, as today's rendering engines are optimised for drawing triangulated surfaces.

For the analysis, interpretation and manipulation of the 3D objects, however, a CSG representation is more suitable, especially when it comes to generalization operations. However, the conversion from CSG to boundary representation is always possible (the reverse conversion is not trivial). Thus, in this approach, the generation of the generalization space is performed off-line and based on CSG, whereas the rendering is done with a boundary representation of the data.

An important prerequisite is that the Geomorphs run exactly synchronized, so that the main characteristics of buildings (horizontality, verticality, etc.) can be preserved. If the data model includes these characteristics implicitly, it is straightforward to ensure it. Such a data model is described in [Thiemann 2001].

In this work, a continuous scale representation based on a differential CSG model will be developed. For visualization, it can be converted into several boundary representations. If required, a multi-scale boundary representation will be created.

4. DISCUSSION AND OUTLOOK

In the next steps of this research, these concepts will be implemented. First, feature splitting and generalization will be realized. Then, the automatic generalization based on minimal constraints is implemented. Based on such a mechanism, investigations concerning the necessary number of generalization levels for different application scenarios will be conducted.

Finally, the generalization of adjacent buildings, building blocks and bigger housing areas will be studied, where special problems arise, like the definition of a simplified roof of two adjacent buildings.

5. REFERENCES

Berlotto, M. & Egenhofer M. [1999], Progressive Vector Transmission, Transactions of the ACMGIS99, Kansas City, MO, pp. 152-157.

Brenner, C. [1999], Interactive modeling tools for 3D building reconstruction, Photogrammetric Week 99, Fritsch/Spiller (eds.), Wichmann, pp. 23-34.

Coors, V. 2001: Feature-preserving Simplification in Web-based 3D-GIS. 1st International Symposium on Smart Graphics, http://www.dfki.de/~krueger/sg2001.

De Floriani, L. & E. Puppo [1995]: Hierarchical Triangulation for Multiresolution Surface Description. ACM Transactions on Graphics 14(4), pp. 363-411.

Hake, G., D. Grünreich & L. Meng [2001]. Kartographie. De Gruyter Verlag.

Heckbert, P., Ribelles, J., Garland, M., Stahovich, T., Srivastava, V. [2001]: Finding and Removing Features from Polyhedra. Proceedings of DETC'01.

Heckbert, P. & M. Garland [1997]: Survey of Polygonal Surface Simplification Algorithms, Technical report, School of Computer Science, Carnegie Mellon University, Pittsburgh, Multi-resolution Surface Modeling Course Siggraph, 97, http://www.cs.cmu.edu/~ph.

Hoppe, H., [1997], View-dependent refinement of progressive meshes, Computer Graphics (SIGGRAPH '97 Proceedings), 1997, pages 189-198.

Kofler, M. [1998]: R-Trees for Visualizing and Organizing large 3D Databases. Doctoral Thesis, University of Graz.

Könninger, A. & S. Bartel [1998], 3D-GIS for Urban Purposes, GeoInformatica, Vol., Nr. 1, pp. 79-103.

Lehto, L. & T. Kilpeläinen, 2000. Real-time Generalization of Geodata in the WEB, in: 'International Archives of Photogrammetry and Remote Sensing', Vol. 33, Part B4, ISPRS, Amsterdam, 2000, pp. 559–566.

Mayer, H. [2000], Three Dimensional Generalization of Buildings Based on Scale-Spaces, *in:* `International Archives of Photogrammetry and Remote Sensing', Vol. 33, Part B4, ISPRS Amsterdam, 2000.

Schmalstieg, D.[1997]: Lodestar - An Octree-Based Level of Detail Generator for VRML. in: Proceedings of SIGGRAPH Symposium on VRML.

Schröder, W. J., Zarge, J. A., Lorensen W. E., [1992], Decimation of Triangle Meshes, Computer Graphics, 26, 2, July 1992, pages 65-69.

Sester, M. & A. Klein [1999]: Rule Based Generalization of Buildings for 3D-Visualization, Proceedings of the International Cartographic Conference (ICA), Ottawa.

Shea, K. & McMaster, R. [1989], Cartographic generalization in a digital environment: when and how to generalize, in: `Ninth International Symposium on Computer-Assisted Cartography', Baltimore, Maryland, pp. 56-67.

Thiemann, F. [2001]: Parametrisierung, Modellierung und Ausgleichung von 3D Gebäudedaten für das Facility-Managment. Diploma Thesis, Institute for Geodesy and Geo-Information, Technical University of Berlin.

van Kreveld, M. [2001], Smooth Generalization for Continuous Zooming, Proceedings of the ICC, Beijing, China, 2001.

APPLICATION DEPENDENT GENERALIZATION – THE CASE OF PEDESTRIAN NAVIGATION

Monika Sester

Institute for Cartography and Geoinformatics, University of Hanover, Appelstraße 9a, 30167 Hanover, Germany
monika.sester@ikg.uni-hannover.de

Commission IV, WG IV/3

KEY WORDS: Generalization, Multiresolution, Navigation, Visualization, Automation

ABSTRACT:

The need for application dependent visualization has steadily been growing with the advent of the internet and is even enforced by the availability of mobile computers. One of the assumed "killer applications" for mobile phones are location based services, allowing to present location information to a mobile user depending on his/her current position. Typical examples for such a location based service is the provision of information about restaurants in the vicinity of the user, or the shortest path from the current position to the next bus stop, etc. When displaying spatial information on a small display, the information presented has to be clear and distinct in order to be visible, and understandable by the user. This is even more true, as the mobile user is typically subjected to a high cognitive load while navigating and has to rely on extremely reduced and adequate information. An approach is presented to generate a compact description and a visualization of a route depending on the starting point and destination of the user. For a save and unambiguous guidance, it must be guaranteed that the relevant information can be grasped at a glance. This involves that important information is highlighted, whereas irrelevant information is reduced. Obviously, this is a generalization problem. Different cartographic operations will be tested and evaluated with respect to their applicability. Important objects like starting and endpoint will have to be enhanced. An important issue is the fact that different generalization levels will be applied in one description, thus there will be several "scales" within one presentation.

1. INTRODUCTION

Navigation systems are gaining an increasing popularity. Today, many middle and upper class cars are equipped with navigation systems that provide facilities for efficient road guidance from the current location of a user to his/her destination. The information provided relies closely on the underlying navigation data, i.e. the digital road map. This leads to a description that is mainly intended to be presented as on-the-spot-information along the current path. It is based on instructions related to the road geometry and describes sections and road junctions, e.g. "in 300m turn right". Thus it is not necessarily related to a natural or human way of describing a path.

Whereas mainly geometry related instructions are appropriate for car navigation systems, things are different when it comes to mobile navigation of a pedestrian. Here, human centred instructions seem to be more appropriate. There are two main reasons for that: Firstly, in car navigation, a combination of different sensors allows for the exact determination of the current position of the driver, therefore, exact position dependent instructions can be provided and are adequate. In pedestrian's navigation, however, the positioning accuracy is currently limited to a few hundred meters, when the user is located by mobile network cells. Even when assuming that in the future mobile devices will be equipped with more precise positioning systems like GPS, there are still inaccuracies e.g. due to multipath effects in dense city areas, that lead to the fact that the positioning accuracy is limited. Therefore, as the mobile user does not have the possibility of an exact positioning, the orientation and guidance information has to rely on intuitively understandable concepts that can immediately be recognized and matched to reality of the environment.

Secondly, today's and also probably future display devices will be of a limited size. This naturally prevents that – like on a map – all the details of the environment can be presented. Guiding instructions have to be abstracted and reduced to the elementary important issues. Besides reducing the amount of objects to be presented, this abstraction and concentration also helps to discern between relevant and irrelevant information: only those objects have to be presented, that are important for the current task – irrelevant information can be suppressed. Importance and relevance, however, depends on several factors, like the application, the current route, and the user – therefore, this generalization has to be performed on demand and on-the-fly.

Thus, generalization is a major issue for this visualization. Furthermore, if routes are to be presented, it is of enormous importance, that the mobile, moving user is able to grasp the spatial situation at a glance. This leads to presentations, that include only the major route elements, together with important landmarks that are placed at positions, where the user has to take decisions. These presentation naturally adopt different scales in one presentation: in the vicinity of start and endpoint of a route there is detailed information presented at a large scale, whereas the information in between can be reduced to the major roads and landmarks.

In the paper an approach is presented, that generates several options for an application dependent generalization of route maps. The assumptions are that all the information is given (i.e. the route, as well as the important landmarks) – the problem to be solved is the adequate generalization and presentation of the spatial situation in order to be understandable in short time and at a glance. The approach uses the following geometric generalization functions: simplification, enhancement and displacement, that are integrated in one framework. Several examples are given that visualize the effect of the different choices of operations and parameters. A conclusion summarizes the paper and gives a short sketch of future work.

2. RELATED WORK

The automation of generalization is an issue that has been tackled by cartographers for several decades. There are well known classifications of generalization operations (e.g. Shea, K. & McMaster, R. [1989], or [Hake, Grünreich & Meng, 2002]), as well as several approaches for the implementation of the generalization methods. Current generalization research concentrates on the integration of different modules in order to achieve a holistic solution. Here, techniques from Artificial Intelligence (knowledge based systems, agents [Lamy et al. 1999]) as well as optimisation approaches [Højholt 1998, Ware & Jones 1998, Sester 2000] are used. This research mainly concentrated on the automatic derivation of traditional map products, e.g. map series of different scales. Glover & Mackaness [1999] discuss the possibility of producing application dependent presentations in different scales: Based on users needs either a tourist map, or a topographic map can be derived from one data set.

New challenges are posed with the advent of small mobile devices, that demand for quick, adequate and readable information visualization [Gartner & Uhlirz 2001]. As the small displays need a high level of abstraction, methods for on-line zooming have to be available in order to inspect details and see the overview. Thus algorithms are needed, that are able to present the information on-demand and on-line [Letho & Kilpälainen, 1999]. Due to the flexible possibility of zooming the requirements concerning the cartographic quality can be relaxed to some extent.

A popular application using spatial data in the internet is route calculation (e.g. mapblast.com, teleinfo.de). Based on shortest path algorithms, the optimal route is selected and highlighted on the corresponding map. Whereas such presentations and descriptions are very well suited for car navigation systems, their usefulness for human navigation is somehow limited: as the whole map is presented in one scale, the typically more complex information in the starting and endpoints of the route cannot be represented adequately in all details, since most of the presentation is occupied by the large part of the connection between the two points. Findings from cognitive psychology show that humans use a different way of describing ways and paths (which also can easily be supported by our personal experience when sketching a route for a friend): only the orientation and navigation relevant information is presented, unnecessary details are dropped; orientation relevant information is needed, when the user has to take decisions like turning, or changing roads; only relative lengths of the route sections are needed, not the exact distances (as people are in general not good at estimating distances). This leads to one presentation, where the information is given at different scales, namely high details (large scales) in the vicinity of start and endpoint, and small scales in between. LineGraph is an approach, that automatically generates a route description based on these principles [Agrawala & Stolte, 2001].

[Harrie et al., 2002] present an approach to derive a vario-scale presentation for different user locations, similar to mono- or polyfocal maps, often used for city maps.

3. APPLICATION DEPENDENT VISUALIZATION

There are different possibilities to graphically highlight important information in order to lead to an immediate recognition of the spatial situation by the user. One obvious way is to use different colours or textures for the objects, i.e. design a presentation based on different graphical variables [Bertin, 1983]. When dealing with small displays, the possibility to use elaborated graphical means is limited, even colour is not yet a standard. Besides the graphical variables, there is the possibility to use generalization operations leading to changes in geometry in order to emphasize relevant information.

In the following, a spatial situation is given (see Figure 1), that will be visualized with different possibilities. The idea is that a visitor looks for a specific building, here the marked target building. An adequate presentation should help him/her to immediately identify the building.

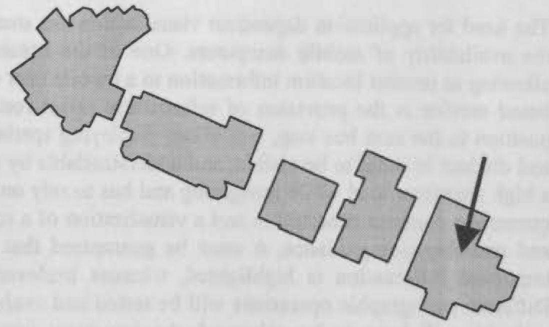


Figure 1: Original Situation and target object marked with arrow.

3.1 Graphical Variables

In the row of different buildings, the one that is relevant can be highlighted by a different colour or in different shades of grey. Also texture can be used to differ between important and unimportant information. Furthermore, the objects can be presented in different size – this aspect will be treated in the next section, however, when dealing with geometry.

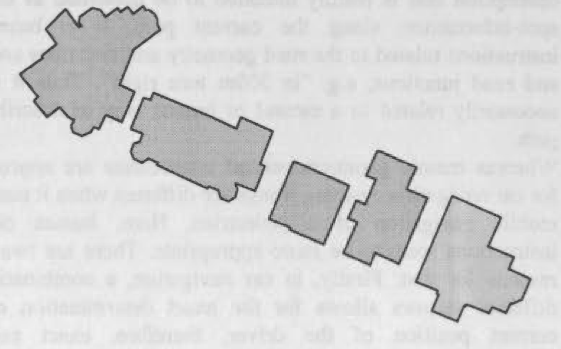


Figure 2: Graphical variable colour to emphasize the target building.

3.2 Generalization

The main idea is to present the information in different levels of detail or scales. The decision on the actual local scale depends on the application. In the following, the assumption is made, that already all relevant information is selected, i.e. generalization methods like selection and deletion have already been applied. Then the following options for generalization are possible.

3.2.1 Enhancement

Consider the row of buildings in the above example: the target building can be enhanced by enlarging it with respect to its neighbours. At the same time, the neighbours can be reduced in size in order to make the target more distinct. Another option is to decrease the enlargement factor smoothly from the target buildings down to the farthest neighbours. This leads to a more continuous transition between the representations.

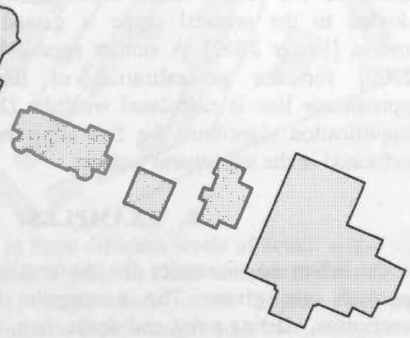


Figure 3: Enhancement: target object is enlarged with factor 1.4, whereas "background objects" are reduced by factor 0.6.

3.2.2 Simplification

The idea is to present the target building in all its details, and simplify the neighbour's shapes. Simplification of ground plans can be achieved by specifying the minimum visible façade width of a building.

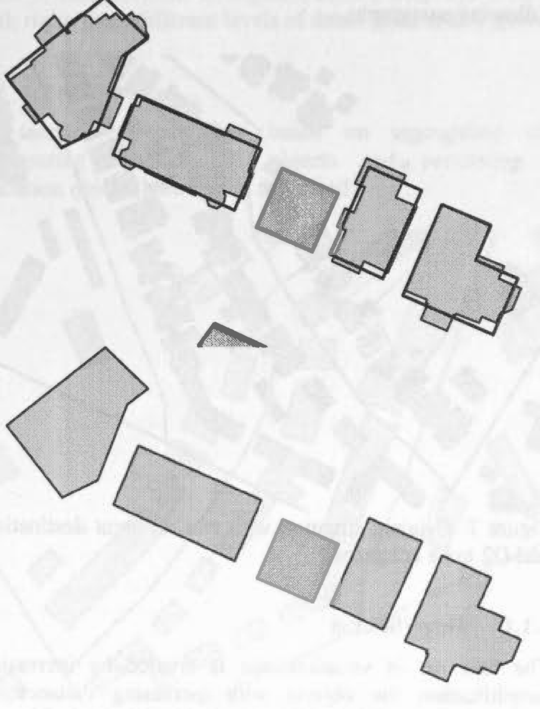


Figure 4: Simplification of object shape: a) overlay of simplified and original shapes, b) result of partial simplification: target shape without simplification.

3.2.3 Aggregation

A further decrease in the level of detail is achieved by aggregating objects that are of minor importance. Thus, several buildings in a row can be combined to form one large object; among those, the target object is separated and presented individually.

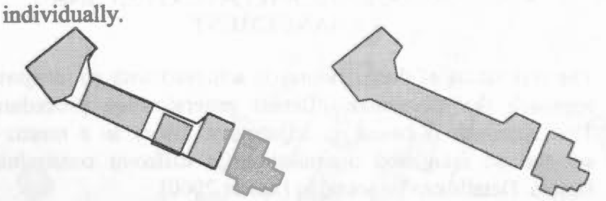


Figure 5: Aggregation of "irrelevant objects" – separation and visualization in full detail of destination object: a) overlay (outline), b) Result of aggregation.

3.2.4 Displacement

This operation is needed in order to compensate for the lack of space that was imposed by the preceding operations: when objects are enlarged, they occupy the space of neighbouring objects. In order to maintain the relative arrangements of the objects, as well as to enforce graphical constraints like minimal distances or minimal sizes, the displacement operation is needed. It has the effect of rearranging the objects while preserving their spatial distribution.

3.2.5 Presentation in 3D

The traditional way of presenting spatial information is in on a 2D map sheet. An interactive systems also allows for a 3D-presentation of a spatial situation, thus, the before mentioned operations can also be visualized in 3D. However, the provision of these operations is still in a research phase, especially the generalization of the shape of 3D buildings (cf. [Mayer 2000, Thiemann 2002]). Figure 6 gives a 3D presentation by simple extrusions of 2D-situations generalized with the before mentioned means.

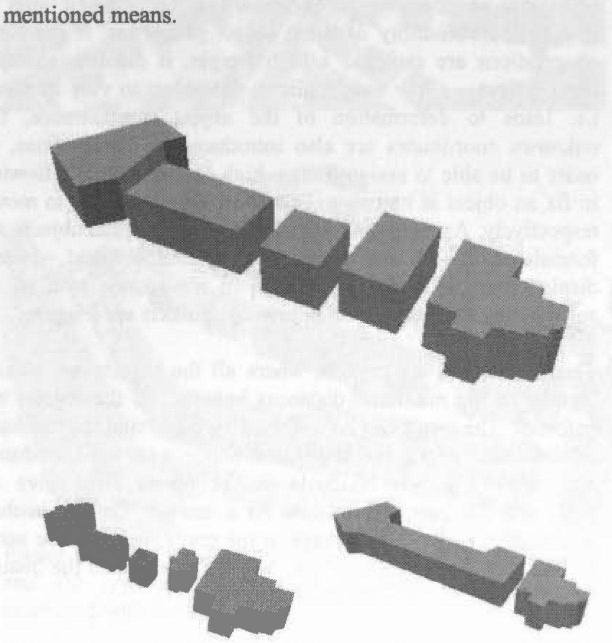


Figure 6: Visualization of different enhancement results in 3D: simplification of "background" objects (up), enlargement (left), aggregation (right).

4. REALIZATION – INTEGRATED APPROACH OF DISPLACEMENT, SIMPLIFICATION AND ENHANCEMENT

The realization of these features is achieved with an integrated approach that combines different generalization procedures. This approach is based on adjustment theory as a means to achieve an integrated optimisation of different constraining factors. Details can be found in [Sester 2000].

Adjustment theory is a means to determine a set of unknowns based on given observations. The observations are described as functions of unknowns in the so-called *functional model*; in addition, the accuracy of the observations can be described in the *stochastic model*. If the functional dependencies are not linear, they have to be linearized, leading to the fact, that approximate values for the unknowns have to be given. The unknowns are determined by the following equation:

$$\hat{x} = (A^T PA)^{-1} A^T P(l - f(x_0)),$$

where A is the Jacobean Matrix of the derivations of the functions according to the unknowns x, P is the weight matrix of the observations, l are the observations and $f(x_0)$ is the value of the function calculated at the approximate values x_0.

Obviously, the use of this scheme is straightforward, as soon as the observations and the unknowns for a given problem are identified. In the case of using adjustment theory for displacement, the unknowns are the coordinates of the points of the objects involved. The observations are distances between the objects, that have to be enforced and set to be at least as large as the minimum legible distance between the objects. Additional object specific observations can be introduced in order to define the form and orientation of the objects. These additional observations are necessary in order to be able to specify the variability of these object properties: if the form observations are assigned a high weight, it enforces that the form is kept – a low weight allows the object to vary its form, i.e. leads to deformation of the object. Furthermore, the unknown coordinates are also introduced as observations, in order to be able to assign them a high or low weight, allowing to fix an object at its original position, or allowing it to move, respectively. As distance constraints between all the objects are formulated, it is ensured that a global solution is found, where a displacement of one object occurs in accordance with all its surrounding objects, and no follow-up conflicts are triggered.

Thus it leads to a situation, where all the objects are clearly legible, as the minimum distances between all the objects are enforced. The result can be analysed by inspecting the residuals of the observations and their accordance with the introduced accuracies. E.g. the residuals in the object sides give an indication for their deformation. As a measure for the absolute positional accuracy, the change in the coordinates can be used. In this way, these measures can be used to evaluate the quality of the result and allow for self-inspection.

The integration of two other generalization operations is possible: Enhancement of an object (enlarge or reduce) is easily integrated, as this can be controlled by the form parameters (object edges) of the object. Aggregation of objects can be achieved by setting the distance between two objects to zero, instead of to the minimum distance. This leads to the fact, that the objects will be adjacent after the adjustment; merging them to form one object has to be performed in a separate step.

Simplification involves a change in the object structure itself, typically a reduction of the number of points. As such discrete changes cannot directly be achieved by the approach, these changes have to be calculated a priori and introduced into the system for a fine adjustment. This is used for building simplification in the following way: based on a set of rules a simplification of the building is achieved. The control parameter for this simplification is the minimum façade width that is just visible at the target scale. Subsequently, the exact shape adopted to the original shape is gained in the adjustment process [Sester 2000]. A similar approach is used by Harrie [2000] for the generalization of linear elements: the approximate line is calculated with the Douglas-Peucker line simplification algorithm; the fine adjustment of the line was performed in the adjustment process.

5. EXAMPLES

In the following, examples for the results achieved with the approach are given. The assumption is that the route description, starting point and destination, as well as important landmarks along the route are given, e.g. derived by approaches like [Elias 2002]. The task is to automatically visualize this situation in a way, that the important objects are enhanced and thus immediately distinguishable.

The spatial situation is given in Figure 7. In the examples only the vicinity of the destination is given for clarity. Two destination objects D1 and D2 are given to compare the different results. The results applying the generalization and visualization methods described above are presented in the following paragraphs.

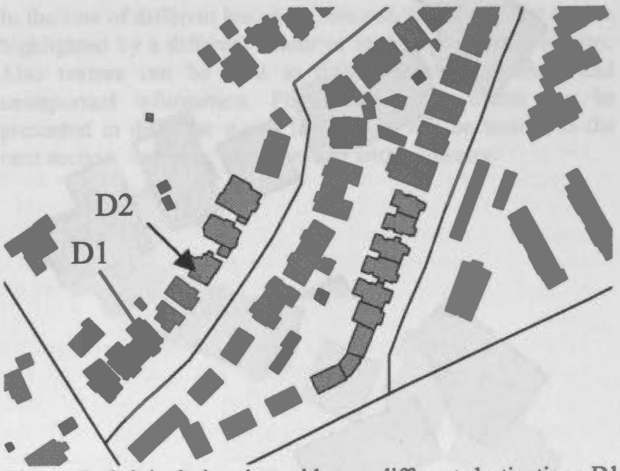


Figure 7: Original situation with two different destinations D1 and D2 to be enhanced.

5.1.1 Simplification

The first set of visualizations is created by increasing the simplification the objects with increasing distance to the destination object. These Levels of Detail (LOD) are realized with the control parameter minimal façade length of a building. In Figure 8, three different LOD's are realized to enhance destination object D1: D1 is given with all its detail in the first LOD, its neighbours are presented with small simplification level (here 3m), whereas all the other objects are highly simplified with a value of 6m.

In Figure 9 the result for target object D2 is given.

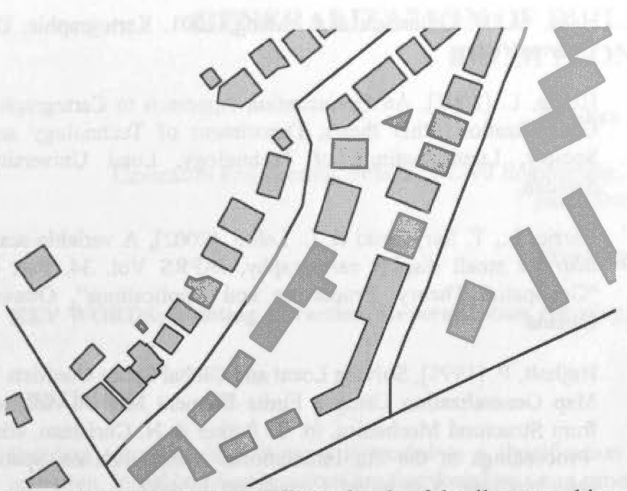


Figure 8: Objects in three different levels of detail: target object D1 in original detail, objects in close vicinity in detail level 2 (minimum façade length 3m), other objects in coarse representation (minimum façade length 6m).

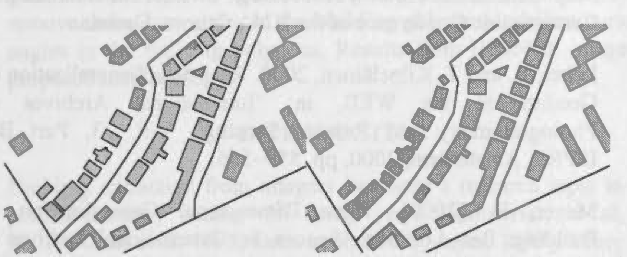


Figure 9: Enhancement of object D2: left: only two levels of detail; right: three different levels of detail (similar to Figure 8).

5.1.2 Aggregation

The second realization is based on aggregating close neighbouring "background" objects and presenting the destination object (here D2) in full detail.

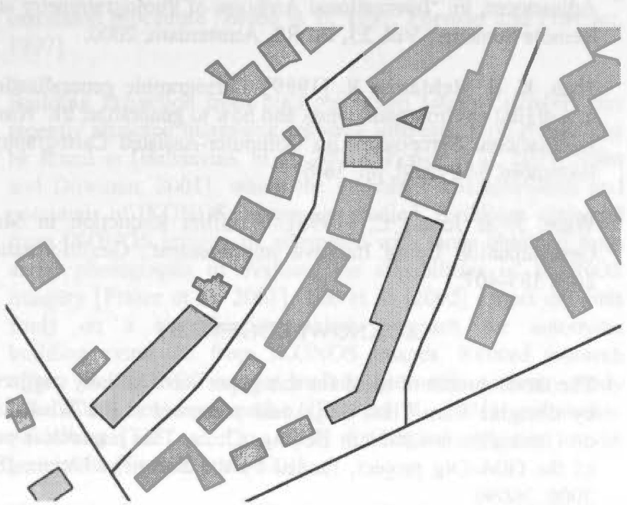


Figure 10: Aggregation of "background" objects (target object D2).

5.1.3 Enlargement

The third possibility is the enlargement of the destination object while the other objects are reduced in size. This leads to the visualization in Figure 11.

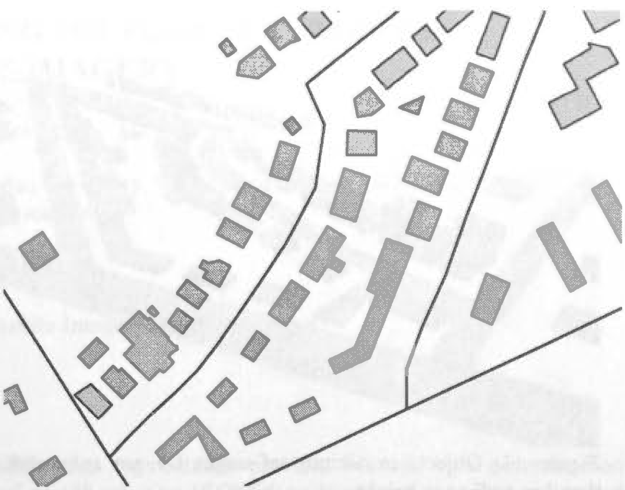


Figure 11: Enlargement of destination object and reducing of "background" objects (target object D1).

5.1.4 Visualization in 3D

Finally, these visualization can also be given in 3D. In Figure 12 selective simplification and aggregation is used.

Another degree of freedom or control parameter can be applied in 3D: the height of the objects can be set either equal for all the objects (see Figure 12), or dependent on the importance of the objects: the more important the objects are, the higher they are (Figure 13).

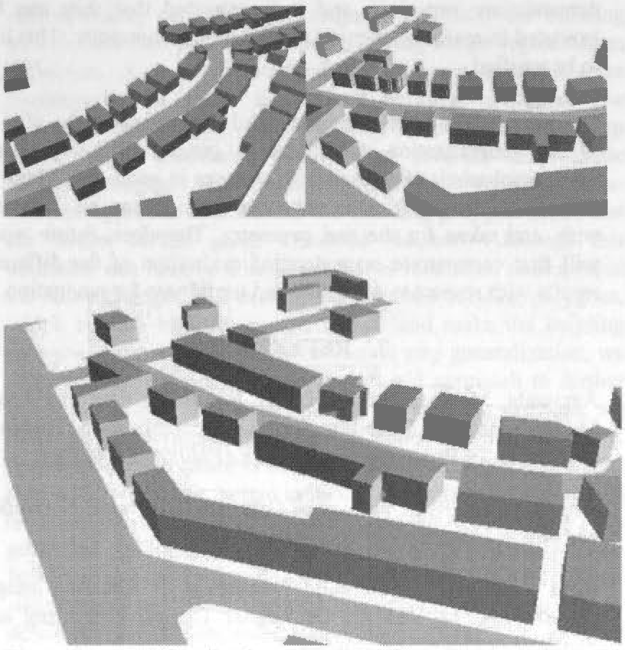


Figure 12: 3D-Visualization: simplification of target objects D1 and D2 (up) and aggregation of target object D2 (low) combined with 3D.

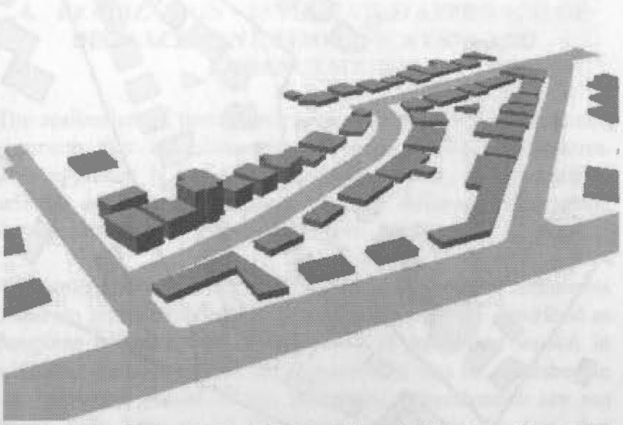


Figure 13: Objects in vicinity of target D1 are enhanced in detail as well as in height.

6. DISCUSSION AND CONCLUSION

The paper presented a toolbox of possibilities to automatically derive different visualizations depending on a given application. A target destination is enhanced with respect to its environment. The different possibilities of rendering have been automatically calculated based on an optimisation procedure, relying on least squares adjustment theory. The complexity of the algorithm depends on the number and complexity of objects involved, as the normal equations are of the dimension of the number of unknowns, i.e. the number of points. As for this application only a few objects are involved, the computational time demands are very low, and it is expected that they can be executed in real time even on a small mobile computer. This has to be verified.

Different results have been generated depending on the choice of the generalization operations. A general difficulty arises when emphasis is expressed with changes in geometry: There is always the danger that the enhanced version can be confused with. and taken for the real geometry. Therefore, future work will first concentrate on a detailed evaluation of the different results with respect to adequacy and usefulness for navigation.

7. REFERENCES

Agrawala, M. & C. Stolte, [2001], Rendering Effective Route Maps: Improving Usability Through Generalization, Stanford University, In *Proceedings of SIGGRAPH01*, pp. 241-249.

Bertin, J. [1983], Semiology of Graphics: Diagrams, Networks, and Maps, University of Wisconsin Press.

Elias, B. [2002], Automatic Derivation of Location Maps, IAPRS Vol. 34, Part 4, "GeoSpatial Theory, Processing and Applications", Ottawa, Canada.

Gartner, G. & S. Uhlirz [2001], Cartographic Concepts for Realizing a location-based UMTS service: Vienna City Guide Lol@, *in:* 'Proceedings of the 20th International Cartographic Conference of the ICA', Beijing, China, Vol III, pp.3229-3239.

Glover, E. & Mackaness, W. [1999], Dynamic Generalization from Single Detailed Database to Support Web Based Interaction, in: 'Proceedings of the 19th International Cartographic Conference of the ICA', Ottawa, Canada.

Hake, G., D. Grünreich & L. Meng, 2001. Kartographie. De Gruyter Verlag.

Harrie, L. [2001], An Optimization Approach to Cartographic Generalization, PhD thesis, Department of Technology and Society, Lund Institute of Technology, Lund University, Sweden.

Harrie, L., T. Sarjakoski & L. Lehto, [2002], A variable-scale map for small display cartography, IAPRS Vol. 34, Part 4, "GeoSpatial Theory, Processing and Applications", Ottawa, Canada.

Højholt, P. [1998], Solving Local and Global Space Conflicts in Map Generalization Using a Finite Element Method Adapted from Structural Mechanics, *in:* T. Poiker & N. Chrisman, eds., 'Proceedings of the 8th International Symposium on Spatial Data handling', Vancouver, Canada, pp. 67-689.

Lamy, S., Ruas, A., Demazeau, Y., Jackson, M., Mackaness, W. & Weibel, R. [1999], The Application of Agents in Automated Map Generalization, *in:* 'Proceedings of the 19th International Cartographic Conference of the ICA', Ottawa, Canada.

Lehto, L. and T. Kilpeläinen, 2000. Real-time Generalization of Geodata in the WEB, in: 'International Archives of Photogrammetry and Remote Sensing', Vol. 33, Part B4, ISPRS, Amsterdam, 2000, pp. 559–566.

Mayer, H. [2000], Three Dimensional Generalization of Buildings Based on Scale-Spaces, *in:* 'International Archives of Photogrammetry and Remote Sensing', Vol. 33, Part B4, ISPRS Amsterdam, 2000.

Thiemann, F. [2002], Generalization of 3D building data, IAPRS Vol. 34, Part 4, "GeoSpatial Theory, Processing and Applications", Ottawa, Canada.

Sester [2000]: Generalization based on Least Squares Adjustment, in: 'International Archives of Photogrammetry and Remote Sensing', Vol. 33, ISPRS, Amsterdam, 2000.

Shea, K. & McMaster, R. [1989], Cartographic generalization in a digital environment: when and how to generalize, in: 'Ninth International Symposium on Computer-Assisted Cartography', Baltimore, Maryland, pp. 56-67.

Ware, J. & Jones, C. [1998], 'Conflict Reduction in Map Generalization Using Iterative Improvement', GeoInformatica 2(4), 383-407.

ACKNOWLEDGEMENT

The development of ideas for this paper were initially triggered by thoughts from Tiina Sarjakoski presented at the Workshop on Generalization 2001 in Beijing, China. This research is part of the GiMoDig project, funded by the European Union, IST 2000, 30090.

GENERALIZATION OF BUILDING POLYGONS EXTRACTED FROM IKONOS IMAGERY

Jie Shan D. Scott Lee

Geomatics Engineering, School of Civil Engineering, Purdue University, West Lafayette, IN 47907-1284, USA
jshan@ecn.purdue.edu

Working Group IV/3

KEY WORDS: Building extraction, Generalization, Shaping, Satellite Imagery, High resolution

ABSTRACT

Building extraction in this paper is treated as a classification guided image segmentation problem followed by a generalization operation. Initial building locations are first found by using supervised classification on IKONOS multispectral images. A supervised classification/cluster based approach is then used to segment the building image from its background and surroundings in the panchromatic image. A series of polygon operations are implemented to determine the best delineation of buildings based on the segmentation results. The delineated building polygons are subsequently refined by a two-step generalization process, distance-based generalization and curvature-based generalization. The distance-based generalization uses the classical Douglas-Peucker approach to remove redundant nodes in the building polygons, while a curvature-based approach is proposed to eliminate sharp and unrealistic angles in the building polygons. Results from IKONOS imagery are presented to show the implementation and efficiency of the proposed methodology.

1. INTRODUCTION

Building extraction from imagery has been a research topic in photogrammetry society for many years [Baltsavias, et al, 2001b; Grün et al, 1995; Grün et al, 1997; Mayer, 1999]. Many efforts use black and white images from air and space as the single data source. However, recent achievements suggest the combination of multiple data sources and the integration of photogrammetric tools with image processing and geographic information systems [Föstner, 1999; Haala, 1999; Huertas, et al, 1999; Huertas, et al, 2000; McKeown, et al, 1999]. As another experience gained in the past years, various building models have been introduced to facilitate and automate the building extraction procedure [Braun et al, 1995; Försnter and Pluemer, 1997].

Building extraction from high resolution satellite imagery has recently attracted increasing research interests. First results can be found in [Baltsavias, et al, 2001a; Fraser et al, 2001; Sohn and Dowman, 2001], where the geometric characteristics and potentials of IKONOS images are studied. Buildings extracted from IKONOS images are compared with those obtained from aerial photographs to evaluate the capabilities of IKONOS imagery [Fraser et al, 2001]. Lee et al [2002] report on their study on a classification-guided approach for automatic building extraction from IKONOS images. Related research efforts on IKONOS images include geometric accuracy study [Dial,2000], road extraction [Dial et al, 2001] and sensor modelling [Tao and Hu, 2001] for digital elevation model generation [Toutin, et al, 2001].

This paper will report on our recent experience gained in automatic building extraction from IKONOS imagery. Although the entire extraction process will be presented, the paper will be focused on the automatic delineation and generalization of building polygons obtained from classification results of multispectral IKONOS images. The objective is to refine the classification results to obtain accurate and realistic building boundaries. Our approach consists of following steps. First, the IKONOS multispectral images are classified into several classes, including water, road, roof, tree, marsh, grass and sand based on the selected training sites. This step is carried out by using a Maximum Likelihood classifier. The resultant class maps are then vectorized to feature classes. Each vectorized building object will be used to define the searching area on the corresponding panchromatic image to delineate the building boundaries. This is accomplished by using image segmentation technique. A series of vector-based polygon operations are implemented to select the refined building polygons. The delineated building polygons will then be subject to a two-step generalization processing to obtain an accurate and realistic building boundary. For generalization, we first use Douglas-Peucker approach to eliminate the building polygon nodes that are within certain given threshold distance. Although this operation can remove a large amount of redundant nodes, rapid curvature changes still exist in the resultant building polygons, which contain high frequency zigzags and make the building polygon shape untrue. As for the second step generalization, we propose and implement a curvature-based approach to further refine the results. For each node in the building polygon, we calculate its curvature and label it as either convex or concave. Nodes whose curvature is beyond a given threshold and whose sign is different with its two immediate neighbouring nodes will be eliminated from the building polygon. In this way, the generated building polygons are closer in shape to real buildings and coincide with the building image well. This paper will present technical descriptions of our approach, shows details of the implementation and illustrate the principles by providing step by step intermediate results.

The rest of the paper is organized as follows. Section 2 briefs the data set used in the test and its classification. Section 3 describes the building delineation process on the panchromatic image guided by the classification results. Section 4 presents the technical details for the two-step generalization process. Sample results for each processing step are shown in every section. The paper concludes at Section 5 by summarizing the gained experience and prospecting the future work.

2. BUILDING CLASS GENERATION

The study area is located at Camp Lejeune and its surroundings (North Carolina, USA). The images used in this study are a single mosaicked IKONOS four band multispectral image and the corresponding IKONOS panchromatic image. The images are pre-rectified by the image vendor Space Imaging, Inc. The properties of the IKONOS images are listed below:

- Spectral resolution: 4 bands (Near IR/R/G/B), 11 bits/pixel
- Spatial resolution: 4 meters/pixel (multispectral) or 1 meter/pixel (panchromatic)
- Region: about 2.3*1.7 square kilometres
- Preprocessing from Space Imaging, Inc.: Standard Geometrically Corrected, Mosaicked
- Map projection: UTM Zone 18, WGS-84

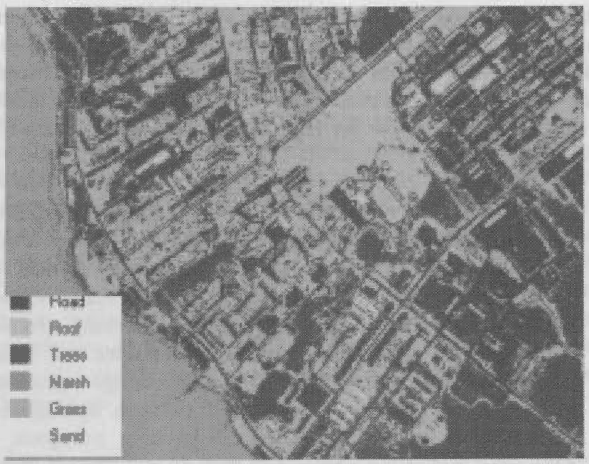


Figure 1. Class map of supervised classification

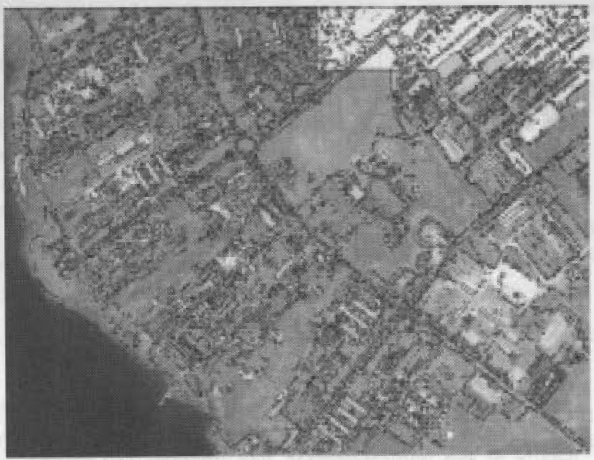


Figure 2. Vectorized building class overlaid on the panchromatic image

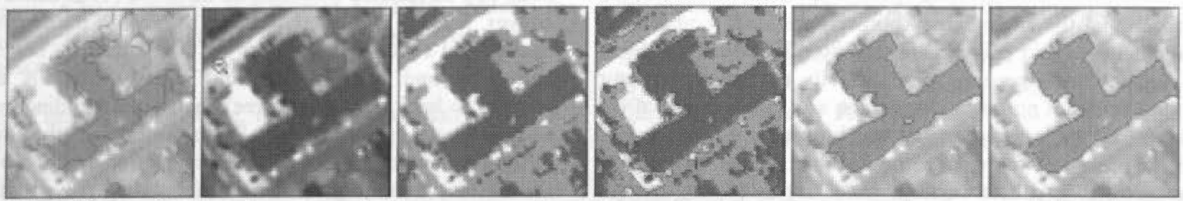
As the first step of the study, the multispectral image is classified to several thematic classes, from which the building class is separated for further processing. In order to determine what classes should be used for the study, an unsupervised classification (clustering) based on the ISODATA (Iterative Self-Organizing Data Analysis Techniques) algorithm [Mather, 1999] is first carried out. A maximum of 12 clusters is specified for the clustering operation. A visual inspection of the clustering results leads to seven (7) primary classes: water, road, roof, tree, marsh, grass and sand. The class Road is used for roads comprised of grayish and white or bluish-white pixels. The class Roof is used for gray, white and light blue rooftop pixels. The class Marsh is used for the reddish-brown areas adjacent to waterways. After the class category is determined, an ordinary maximum likelihood classifier is used to carry out the supervised classification. The classification is conducted with two packages, a campus-developed software Multispec [Biehl and Landgrebe, 1996] and the commercial image processing package ENVI, both yield essentially the same results. Figure 1 shows the class map resulting from the supervised classification as described above, whereas Figure 2 presents the vectorized building class overlaid over the corresponding panchromatic IKONOS image.

3. BUILDING DELINEATION

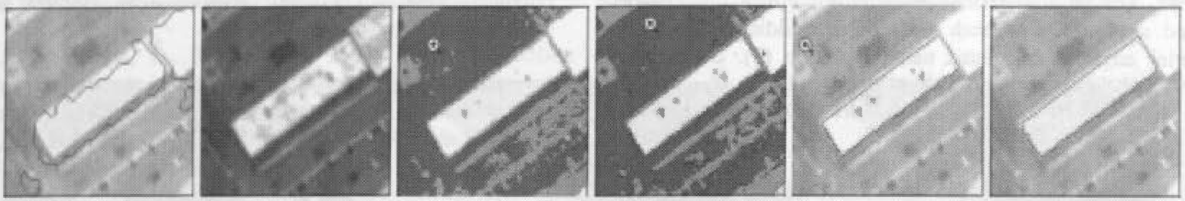
As shown in Figure 2, the original vectorisation of building classes contains many artefacts and non-building objects. The objective of building delineation is to obtain accurate building boundaries from the vectorisation results. This task consists of several steps.

Artefacts from the classification process need to be removed before further processing. After the building class is converted to vector form, each potential building will then be represented as a polygon object. Its primary geometric properties, such as length and area, are then calculated for each building polygon. Building polygons below given thresholds of areas and lengths are then regarded as artefacts caused by misclassification and eliminated from further processing. Our experience shows that a large percentage of classification artefacts can be eliminated through this process.

The next step is to precisely determine the building boundary on the panchromatic image guided by the classification results. The extension of the building polygon obtained from the classification is first used to define the working window for segmentation. Instead of using a conventional binary image segmentation approach [Otsu, 1976], we propose to use the unsupervised classification ISODATA approach to segment the panchromatic image window into several (> 2) classes [Lee et al, 2002]. In this way, multiple gray values in the working window can be taken into account and the difficulty in determining the segmentation threshold is avoided. This approach helps to separate buildings from multiple non-building backgrounds and results in satisfactory segmentation results. It is especially beneficial in low contrast areas where building and background have similar gray values. After the image window is segmented into several classes, the class polygon that has the maximum intersection area with the building polygon obtained from classification will be selected as the building delineation result from the panchromatic image. An algorithm is developed that calculates the intersections of the building class polygon from classification results with each polygon obtained with unsupervised classification on the panchromatic image. However, the selected maximum intersection polygon may contain small holes due to the sensor noise and uneven reflection of the building roofs. These unrealistic small holes need to be detected and removed from the building boundary polygon. An algorithm is developed that explodes a complex polygon with rings to its simple component polygons so that only the building polygon boundary is remained for further generalization processing.



a. building class b. edge enhanced c. segmented b d. vector of c. e. max-polygon f. cleaned



a. building class b. edge enhanced c. segmented b d. vector of c. e. max-polygon f. cleaned

Figure 3. Two examples (in two rows) of building delineation process

Figure 3 presents sample results of building delineation process described above. Figure a shows the building class polygon obtained through classification. It is overlaid on the panchromatic image. The region is the working area defined by the extent window of the building class polygon. Figure b shows the enhanced image after a low pass smoothing filter is applied. This step is to ensure that the segmentation process will not return too many small meaningless polygons inside a building boundary. The segmentation result is shown is figure c, while figure d is the vector form of figure c. As shown in figure d, it contains many polygons that are potentially corresponding to the building class polygon (figure a). Figure e shows the polygon that intersects with the building class polygon with maximum overlap. The final delineation result is shown in figure f where the internal small holes are removed.

4. BUIDLING POLYGON GENERALIZATION

The objective of this step is to obtain accurate boundary and shape of the delineated building polygons. As shown in Figure 3, the straight boundaries of the delineated building polygons are zigzagged. Besides, corners of building polygons are closed unrealistically. The building polygons may not be completed correctly. Some parts are missed and some are wrongly extended to backgrounds due to the shadow of building and trees. To solve this problem, a vector based two-step generalization procedure is proposed. As the fist step, we use the conventional (distance-based) Douglas-Peucker (1977) approach to remove extra building points that cause the zigzag shape of a building boundary. In the second step, we propose a curvature or angle based generalization technique [Costa et al, 2001] to further remove the points that cause unrealistic sharp changes in the building polygon. This technique is outlined as follows.

As is shown in Figure 4, nodes of a polygon are ordered clockwise. The angle β_i at node of two consecutive edge vectors can be calculated by

$$\cos\beta_i = \frac{p_{i-1}p_i \bullet p_i p_{i+1}}{\left|p_{i-1}p_i\right| \cdot \left|p_i p_{i+1}\right|} \tag{1}$$

The cross product of the two edge vectors is calculated as follows

$$\Delta_i = \frac{p_{i-1}p_i * p_i p_{i+1}}{\left|p_{i-1}p_i\right| \cdot \left|p_i p_{i+1}\right|} \tag{2}$$

and can be used to obtain the interior angle α_i of two consecutive edge vectors at a node

$$\alpha_i = \pi + sign(\Delta_i)\beta_i \tag{3}$$

The convex and concave property of the node can then be determined by using the following determinant

$$\Delta_i \begin{cases} > 0 & \alpha_i \quad concave, \quad mouth \\ < 0 & \alpha_i \quad convex, \quad ear \end{cases} \tag{4}$$

A polygon node will be trimmed if its angle is beyond a certain threshold and its sign is different with its two immediate neighbouring nodes (one forward, one backward). In this way, sharp and unrealistic angles remained from the distance-based generalization can be removed.

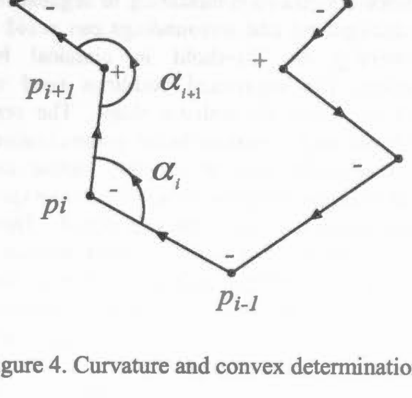


Figure 4. Curvature and convex determination

To generalize the building polygons, the distance-based approach and the angle-based approach are used sequentially. First, the nodes of a polygon are significantly reduced by using the distance-based generalization. Usually the reduction factor for the number of nodes can be as high as 2, namely half of the nodes in the delineation results can be removed. During this step, in order not to remove the building corner points, the generalization is carried out conservatively by giving a small distance threshold of not larger than one pixel (1 meter in this study). The angle-based generalization follows the distance-based approach. For each node in a building polygon, its interior angle is calculated using Equation (3). Beside, each node will be labelled according to Equation (4) as concave and convex. A node will be removed or trimmed if it alternates the convex and concave property of its two immediate neighbours and is beyond a given threshold region, which is usually selected to keep right (90 degree) angles of a building polygon and eliminate its sharp turns. Sample results of the two step generalization are presented in Figure 5. In this figure, blue (at the bottom) is the original delineated building polygon, the red is the distance-based generalization polygon, the green one is obtained from angle-based generalization applied on the distance-based generalization results.

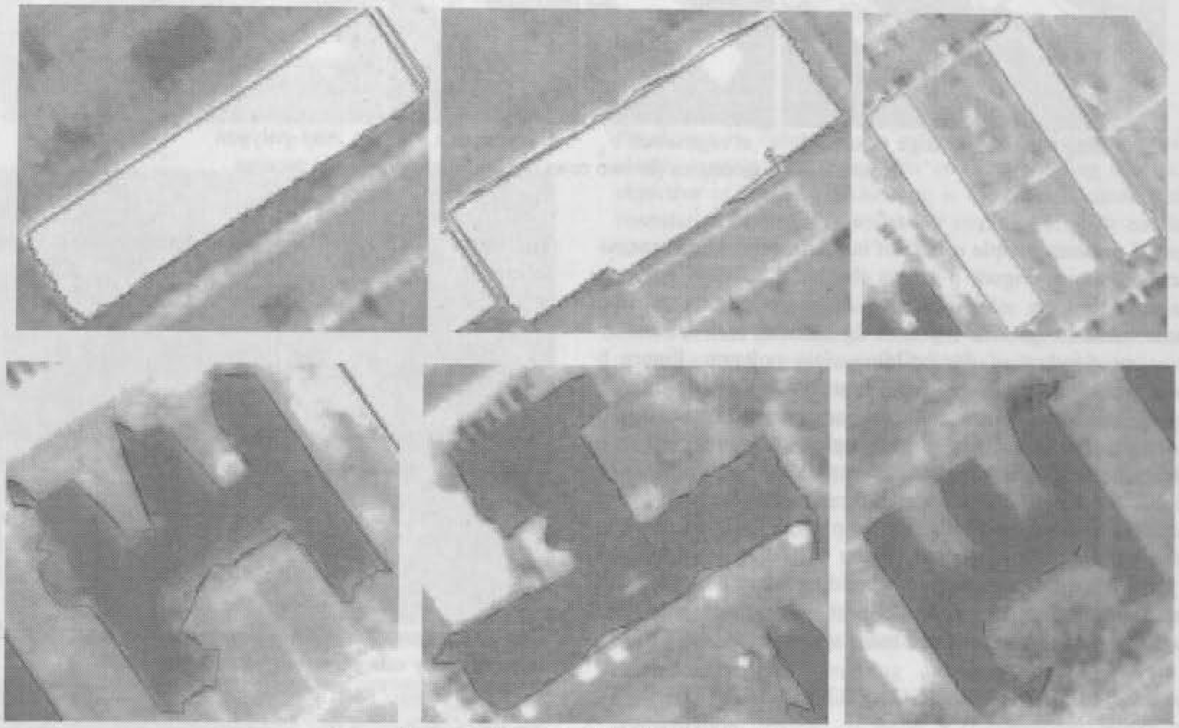


Figure 5. Generalized building polygons
Blue: delineated; Red: distance-based; Green: curvature-based (applied to the Red result)

5. CONCLUDSIONS

The high resolution of recent IKONOS images makes the building extraction possible and is attracting increasing research interests. To automate this process and obtain reliable results, multiple data sources and advanced techniques are needed. The classification guided approach can utilize the properties of both multispectral and panchromatic images. The use of unsupervised classification/clustering to segment building class from its background and surroundings can avoid the difficulty in determining the threshold in classical binary image segmentation. The segmented buildings need to be further processed to obtain its realistic shape. The combination of distance-based and curvature-based generalization approaches provide a powerful tool to simplify, refine and shape the delineated building polygons obtained from image classification and segmentation. The distance-based Douglas-Peucker approach is effective to eliminate a large number of redundant nodes from the building polygon, whereas the subsequent curvature-based approach can then eliminate sharp and unrealistic angles based on polygon shape analysis. Further study will be focused on expanding this vector based shape analysis approach by introducing generic models, including straight and parallel lines and right angle to square the building polygons. In this way, the effect caused by shadows and trees can also be remedied. The vector-dominant characteristic of the proposed approach will enable us to integrate it into GIS packages so that the powerful GIS vector analysis and operation capabilities can benefit the building extraction process, while the extracted buildings will be directly stored in GIS for database creation and visualization.

REFERENCES

Baltsavias, E., Pateraki, M., Zhang, L., 2001a. Radiometric and geometric evaluation of Ikonos GEO images and their use for 3D building modeling, Proc. Joint ISPRS Workshop "High Resolution Mapping from Space 2001", Hannover, Germany, 19-21 September.

Baltsavias, E., Gruen, A., Van Gool, L. (Eds.), 2001b. Automated Extraction of Man-Made Objects from Aerial and Space Images (III), Balkema Publishers, Lisse.

Biehl, L. and D. Landgrebe, 1996. "MultiSpec - A Tool for Multispectral-Hyperspectral Image Data Analysis", Pecora 13, Sioux Falls, South Dakota, August 1996.

Braun, C., Kolbe, Th., Lang, F., Schickler, W., Steinhage, V., Cremers, A.N., Förstner, W., Plumer, L., 1995. Models for Photogrammetric Building Reconstruction. Computer & Graphics, Vol. 19, No. 1, pp. 109-118.

Costa, Luciano da Fontoura, Cesar, Roberto Marcondes Jr., 2001, Shape Analysis and Classification, Theory and Practice, CRC Press.

Dial, G., 2000. IKONOS Satellite Mapping Accuracy, CD-ROM Proceedings of the ASPRS Annual Conference.

Dial, G., Gibson, L., Poulsen, R., 2001. IKONOS satellite imagery and its use in automated road extraction. In: Automated Extraction of Man-Made Objects from Aerial and Space Images (III), E. Baltsavias, A.Gruen, L. Van Gool (Eds.), Balkema Publishers, Lisse, pp. 349-358.

Douglas, D., Peucker, T., 1973. Algorithms for the reduction of the number of points required to represent a digitised line or its caricature, The Canadian Cartographer, Vol. 10, No. 2, pp. 112-122.

Förstner, W., Pluemer, L., 1997. Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, Birkhaeuser.

Förstner, W. 1999. 3D-City Models: Automatic and Semiautomatic Acquisition Methods, in Dieter Fritsch / Rudolf Spiller (Eds.), Photogrammetric Week '99, Wichmann, Heidelberg, pp. 291-303.

Fraser, C., Baltsavias, E., Gruen, A., 2001. 3D building reconstruction from high-resolution IKONOS stereo imagery. In: Automated Extraction of Man-Made Objects from Aerial and Space Images (III), E. Baltsavias, A. Gruen, L. Van Gool (Eds.), Balkema Publishers, Lisse, pp. 325-338.

Gruen, A., Kuebler, O., Agouris, P., Eds, 1995. Automatic Extraction of Man-made Objects from Aerial and Space Images, Birkhaeuser, Basel, Switzerland.

Gruen, A., Baltsavias,E., Henricsson,O. (Eds), 1997. Automatic Extraction of Man-made Objects from Aerial and Space Images, Birkhaeuser, Basel, Switzerland.

Haala, R., 1999. Combining multiple data sources for urban data acquisition, in Fritsch and Spiller, 1999, pp. 329-340.

Huertas, A., Kim, Z., Nevatia, R., 2000. Multisensor Integration for Building Modeling, Proceedings of CVPR.

Huertas, A., Nevatia, R., Landgrebe, D., 1999. Use of Hyperspectral Data with Intensity Images for Automatic Building Modeling, Proceedings of The 2nd International Conference On Information Fusion, Sunnyvale, CA, July.

Lee, D.S., Shan, J., Bethel, B., 2002. Classification-guided building extraction from IKONOS imagery, CD-proceedings of American Society for Photogrammetry and Remote Sensing, Washington DC.

Mather, P., 1999. Computer Processing of Remotely–Sensed Images: An Introduction, 2nd Edition, John Wiley & Sons, Chapter 8.

Mayer, H., 1999. Automatic object extraction from aerial imagery – a survey focusing on buildings, Computer Vision and Image Understanding, Vol. 74, No.2, May, pp. 138 – 149.

McKeown, D., Jr., Cochran, S., Ford, S. McGlone, C., Shufelt, J., Yocum, D., 1999, Fusion of HYDICE Hyperspectral data with panchromatic feature extraction, IEEE Transactions on Geoscience and Remote Sensing, Vol. 37, No. 3, May.

Otsu, N., 1976, "A threshold selection method from gray-level histograms," IEEE Transactions of Systems, Man, and Cybernetics, vol. SMC-9, pp. 62-66, January.

Sohn, G., Dowman, I.J., 2001. Extraction of buildings from high resolution satellite data, Proceedings of the Third International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images, 10. June - 15. June, Centro Stefano Franscini, Monte Verita, Ascona, Switzerland.

Tao, C., Hu, Y., 2001, A Comprehensive Study of the Rational Function Model for Photogrammetric Processing. PE&RS, Vol. 67, No. 12, pp1347.

Toutin, Th., Chénier, R. and Carbonneau, Y., 2001. 3D geometric modelling of Ikonos GEO images. Proc. Joint ISPRS Workshop "High Resolution Mapping from Space 2001", Hannover, Germany, 19- 21, September.

Braun, C., Kolbe, Th., Lang, F., Schickler, W., Steinhage, V., Cremers, A.B., Förstner, W., Plümer, L., 1995. Models for Photogrammetric Building Reconstruction. Computers & Graphics, Vol. 19, No 1, pp. 109-118.

[illegible], 2001. [illegible] Analysis and Classification, Theory and Practice.

Dial, G., 2000. IKONOS satellite mapping accuracy. ASPRS Proceedings of the ASPRS Annual Conference, Washington.

Dial, G., Gibson, L., Poulsen, R., 2001. IKONOS satellite imagery and its use in automated road extraction. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), E. Baltsavias, A. Gruen, L. Van Gool (Eds.), Balkema Publishers, Lisse, pp. 357-367.

Douglas, D., Peucker, T., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer, Vol. 10, No. 2, pp. 112-122.

Förstner, W., Plümer, L., 1997. Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, Birkhäuser.

[illegible]

[illegible]

Gruen, A., Baltsavias, E., Henricsson, O. (Eds.), 1997. Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser, Basel, Switzerland.

Gruen, A., Kuebler, O., Agouris, P. (Eds.), 1995. Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser, Basel, Switzerland.

Haala, N., 1999. Combining multiple data sources for urban data acquisition. In: Fritsch and Spiller, 1999, pp. 329-340.

Huertas, A., Kim, Z., Nevatia, R., 2000. Multisensor integration for building modeling. Proceedings of CVPR.

Jaynes, C., Riseman, E., Hanson, A., [illegible] Proceedings of The 2nd International Conference on [illegible]

Lee, D.S., Shan, J., Bethel, J.S., 2002. Class-guided building extraction from Ikonos imagery. Proceedings of ASPRS Annual Conference, Washington DC.

[illegible]

Mayer, H., 1999. Automatic object extraction from aerial imagery - a survey focusing on buildings. Computer Vision and Image Understanding, Vol. 74, No. 2, pp. 138-149.

McKeown, D., Cochran, S., Ford, S., McGlone, C., Shufelt, J., Yocum, D., 1999. Fusion of HYDICE hyperspectral data with panchromatic imagery for cartographic feature extraction. IEEE Transactions on Geoscience and Remote Sensing, vol. 37, No. 3, pp. 1261-1277.

Otsu, N., 1979. "A threshold selection method from gray-level histograms", IEEE Transactions of Systems, Man, and Cybernetics, vol. SMC-9, pp. 62-66.

Sohn, G., Dowman, I.J., 2001. Extraction of buildings from high resolution satellite data. In: Third International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images, [illegible]

Tao, C.V., Hu, Y., 2001. A comprehensive study of the rational function model for photogrammetric processing. PE&RS, Vol. 67, No. 12, pp. 1347-1357.

Toutin, Th., Chénier, R. and Carbonneau, Y., 2001. 3D geometric modelling of Ikonos Geo images. Proc. Joint ISPRS Workshop "High Resolution Mapping from Space 2001", Hannover, Germany, 19-21 September.

WG IV / 4:

SPATIAL DATA INFRASTRUCTURES

DESIGN, MODELING AND DATA PREPARATION OF MINES AND METALS GEOGRAPHICAL INFORMATION SYSTEM

A. Mansourian*, M.J.Valadan Zoej**

Department of Geodesy and Geomatics Engineering, K.N.Toosi University of Technology, P.O.Box: 15875-4416,
Postcode:19697, Tehran, Iran
*Mansourian@edri.net, **Valadan@ce.kntu.ac.ir

Commission IV, WG IV/4

KEY WORDS: GIS, spatial data infrastructure, mines and metals, data preparation

ABSTRACT

In September 2000 Ministry of Industry and Mines (MI&M), which is responsible for exploration, exploitation and management of mines as well as managing Industrial affairs in Iran, decided to build up a GIS in such a way to manage and control different tasks related to its Mines and Metals (M&M) Department. In November 2000, Department of Geodesy and Geomatics Engineering (GGE) of K.N.Toosi University of Technology has been selected to carry out this project. This project consists of two main stages. In the first stage a DBMS (DataBase Management System) is constructed and in the second stage applications of different parts of M&M Department will be added to it. This paper outlines the first stage with emphasizing on preparing and constructing the spatial database for M&M GIS, which is one of the fundamental requirements of MI&M.

1- INTRODUCTION

Before starting M&M GIS project a feasibility study was carried out to find the potential of MI&M for construction of a GIS. The results of this study showed that no powerful spatial DBMS for their appropriate spatial data infrastructure such as standard and specification for designing and implementing the system was available. So it was decided to carry out the project into two main phases.

The first phase involves standardization and preparing fundamental data, which is required for M&MGIS system, and building a DataBase Management System (DBMS). The result of this phase is a base GIS system containing those data and common GIS functionalities that almost all of the sections in the ministry need them.

The second phase involves adding required data and GIS functionalities of each section of the ministry into the DBMS. So applied GIS of each section of the ministry will be built separately based on prepared DBMS in the first phase. This paper outlines the first phase of the project.

In order to supervise the first phase of the project, the Steering Committee of Mines and Metals GIS (SCOMMG) consisting of GIS, mapping and geology experts was organized. For managing and carrying out the project better and more sophisticated, the SCOMMG broke the project into different subproject as follows:

- Recognition
- Conceptual modeling
- Selecting GIS environment
- Standardization
- Preparing methods and rules
- Logical modeling
- Physical modeling
- Designing user interface
- Data preparation and structuring
- Software development and implementing the system

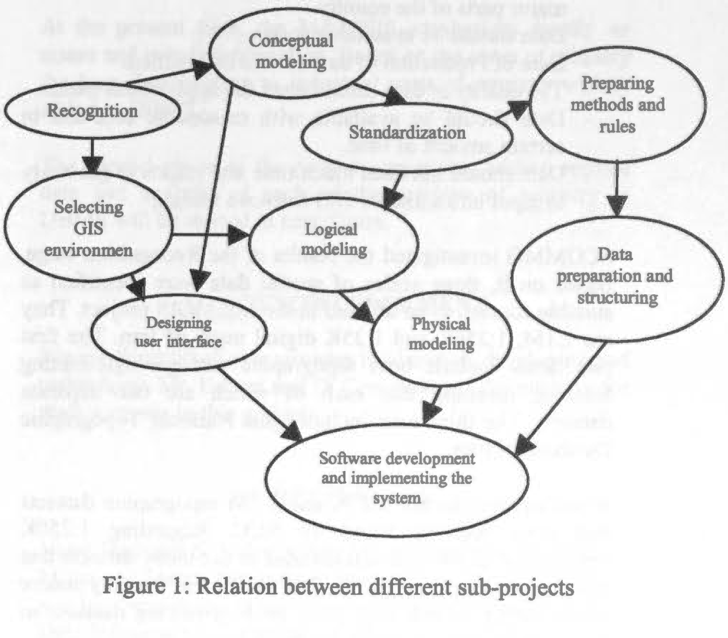


Figure 1: Relation between different sub-projects

2- RECOGNITION

This subproject includes recognition the availability of fundamental data in the ministry and other organization out of the ministry, in a GIS point of view.

The major departments of ministry dealing with spatial data such as the National Geological Survey Organization (NGSO), the Deputy of Exploitation, the Directorship of

Exploration, the Bureaus of Counties and etc. were visited. Different existing spatial data ranging maps, aerial photographs, satellite imageries as well as attribute data relating to spatial data were investigated. Also those range of requirements that a GIS can reply to each of the departments were investigated and documented.

Existing spatial data in different organizations, out of ministry, were also investigated and evaluated with respect to this project requirement. The main organizations include National Cartographic Center (NCC) which produces 1:25,000 scale topographic maps of whole Iran as well as 1:1,000,000 topographic database, National Geography Organization (NGO) which had produced 1:50,000 and 1:250,000 scale topographic map of the whole country, Iranian Remote Sensing Center (IRSC) which has updated some of the layers of 1:250,000 scale topographic maps using satellite imageries.

3- CONCEPTUAL MODELING

This stage consists of selection of appropriate scale (scales) of spatial data for storing in the system, compilation of the information framework, compilation of the information cycle and defining and representing the relation between entities. For building a base GIS, one has to consider the potential sources of data and what data are available and can be used with minimum level of preparation and processing.

The most important factors that were considered in selecting the fundamental datasets were:

- Data should be available for entire country or at least major parts of the country.
- Data should be in suitable scales.
- Date of Production of data should be justified.
- The quality of data should meet the applications needs.
- Data should be available with reasonable cost and in certain amount of time.
- Data should not need much time and efforts to get ready to input into a known GIS software format.

SCOMMG investigated the results of the Recognition stage. Based on it, three scales of spatial data were identified as suitable sources to be utilized in the M&MGIS project. They are 1:1M, 1:250K and 1:25K digital maps of Iran. The first two items include both topographic and geologic-mining features, meaning that each of which are two separate datasets. The third item includes just National Topographic Database of Iran.

It was decided to use 1:25K and 1:1M topographic datasets that have been produced by NCC. Regarding 1:250K topographic datasets, it was decided to use those datasets that have been updated by IRSC. The NGSO was the only source of producing 1:250K and 1:1M geology-mining datasets so their datasets were selected for this project.

Based on the requirements of M&MGIS project the framework of information at each scale was compiled so the required features as well as their associated attributes were selected. The information cycle, which clarifies those departments that are responsible for producing, updating, storing and using of each of the features, was compiled.

At the end, the relationships between entities including spatial-topological relationships and spatial-non-topological relationships as well as non-spatial relationships were defined and represented in a network diagram.

4- SELECTING GIS ENVIRONMENT

An expert group evaluated different commercial GIS software under supervision of SCOMMG. Comparing the M&MGIS requirements and the capabilities and specifications of each of the evaluated software, Arc/Info was selected to be used as GIS environment. Also SQLServer was selected for storing attribute data, which will be linked to Arc/Info. It was decided to use Visual Basic for programming purposes.

5- STANDARDIZATION

The great efforts have been done for unification of data sets specifications. This was because of the project dealt with different data sets from different sources. The main issue in this regard was the adaptation of standards and specifications in such a way that one can relate between similar features in different scales and data sets without facing to any problem. It was obviously easy to add extra layers to an existing data set with the same scale. But if one intends to mix several features in different scales, he or she may meet a lot of problems regarding logical consistency, different form of geometry, having the same features in more than one scale with different specifications (graphic symbology, definition, extents, etc) and so on.

For the project, first of all, a set of features in topography and each specific application in the Ministry, including geology survey, have been identified by SCOMMG. Two data sets, namely, 1:25K and 1:1M topographic databases had their own standards and specification. The rest did not have standards but there were a lot of documents concerning definitions, specifications, and symbology. Since 1:25K data set was accompanied by the most complete sets of standards and specification and it was the largest scale amongst the other, SCOMMG decided the standard set of 1:25K to be the basis for adapting and developing of the other specifications for the selected data sets. Subsequently, all the specifications were put together to develop a single set of standards, including all features in all categories, applications and scale. This set includes both topographic and mining features at the scale of 1:25K, 1:250K and 1:1M. Features are grouped into two main classes, namely, topography and mine & metals. The standards set also includes feature name, feature code, class and sub-class, geometry (the shape of feature), graphic symbology, and non-spatial attributes of each feature. The set is called "the Mines and Metals (M&M) GIS Standard"

6- PREPARING METHODS AND RULES

For producing, updating, storing, editing, structuring and quality control of spatial data as well as attribute data, required methods and rules were prepared and documented.

7- LOGICAL MODELING

In this stage, the conceptual model was documented in a manner to be implemented easily in the system. It contains normalization and denormalization of tables, defining required tables and their associated fields, defining primitive keys and foreign keys as well as defining the domains of the fields in each scale separately.

8- PHYSICAL MODELING

In this sub-project the result of logical modeling was implemented in Arc/Info and SQLServer. It contains creating tables, associated fields with tables, primitive keys and foreign keys as well as linking tables together based on logical model in SQLServer.

9- DESIGNING USER INTERFACE

After holding several meetings and evaluating the requirements of ministry as well as the technical knowledge of GIS users in ministry it was decided to develop four software in one frame by customizing Arc/Info or developing software using Arc/Info functionalities. The first software, which was named General Mine GIS Software (GMGS), should have a user-friendly interface, which gives the users the capability of creating reports, doing required analysis easily as well as creating special thematic maps and so on. The other software will be used for storing and updating spatial data, storing and updating attribute data and creating backup.

In this sub-project appropriate user interfaces were designed for each of the software.

10- DATA PREPARATION AND STRUCTURING

As it was mentioned, different datasets from different organizations were used in this project in which each of the datasets had their own quality and structure. Data to be applied in any geographic database must be appraised in terms of quality. Quality comprises some components, namely, accuracy, logical consistency, resolution, completeness, time, lineage, and so on.

Various tests had been carried out by SCOMMG on the selected data sets to examine the quality parameters. For example, the date of production and the level of processing on digital maps were two important characteristics that had been assessed by SCOMMG.

1:1M and 1:25K topographic data sets had metadata and therefore evaluated easily. The 1:1M topographic maps fulfilled the conditions and were introduced as the suitable sets. 1:25K topographic data had some problems such as: existence of overshoots and undershoots in maps due to the specification and limitation of Microstation in removing these errors. Also no unique code is assigned to features, which have been continued in different map sheets. Specifying a separate database to each map sheet and etc were another problems.

1:250K geologic-mining maps had some problems too. Having no good geometric accuracy, difference in interpretation of geological features that lied in the edge of adjacent map sheets, using no standard in cartographic representation of features in different map sheets and so on were some of the problems of these maps.

1:250K topographic maps were the only set that did not come with any metadata or ancillary information. As it stated before, they have been produced several years ago just as paper map and for long time no updating has come to pass. The only thing that helped SCOMMG was list of map features and layers of digital maps. But since this set was the only set that could cover the entire country, SCOMMG decided to use the set with extra steps for processing and making them ready for adaptation to other sets and for being used in the project.

Except 1:1M scale topography and geology maps, the other datasets needed to be structured and ready for using in M&MGIS system.

In this stage, based on the prepared specifications, methods and rules, all of the datasets, which had not an appropriate structure, were edited and became ready for using in M&MGIS system.

11- SOFTWARE DEVELOPMENT AND IMPLEMENTING THE SYSTEM

Based on the designed user interfaces and functionalities, required software is now under customization and developing by a group of experts in software engineering. After that the prepared and structured data will be integrated with software. The prepared package will be tested and then will be installed in different departments of ministry.

12- FUTURE TRENDS

At the present time, the M&MGIS emphasizes mostly on mines and metals information. Based on the order of ministry the base data relating to industrial parts of ministry will be added to DBMS.

The second phase of the project concerning adding applied data and analysis of each of the sessions of ministry to DBMS will be started in near future.

13- ACKNOWLEDGEMENT

We are thankful of vise minister in planning, developing and technology, Mr. Eslami and IT Committee of the ministry for their supports in this project.

14- REFERENCES

Noori Bushehri S, Valadan zoej MJ, Mansourian Ali (2001) Mines and Metals Geographic Information System, Design and Implementation in the Context of Iran's National Spatial Data Infrastructure, International Symposium on SDI, Melbourne, Australlia

M.J.Valadan Zoej, A. Mansourian (2002) Design and Implementation of a Seamless and Multi-Scale Spatial Database for Iranian Ministry of Industry and Mines", MIS 2002, Wessex Institute of Technology, Southampton, UK.

AN ON-LINE SERVICE TO ACCESS AND VISUALIZE THE DATA ALIGNMENT LAYER

B. Ameri, A. Shankaie, R. Orth

MacDonald Dettwiler and Associates (MDA)
13800 Commerce Parkway, Richmond, BC. Canada
Tel: (604) 231-2308, Fax: (604) 278-2117
bameri@mda.ca, ashankai@mda.ca, rorth@mda.ca

Commission IV

KEY WORDS: Web mapping, On-line services, Spatial Search, HTML/XML, RDBMS

ABSTRACT:

The CGDI (Canadian Geospatial Data Infrastructure) currently supports access to the CDAL (CGDI Data Alignment Layer) data via a public Web site, however, the process to retrieve the data is designed for human interaction and is not available programmatically as an online service. Furthermore, retrieval of individual, or an arbitrary range of points is not currently supported by the CDAL, and instead complete data sets have to be requested, where a data set consists of a regular area of points based primarily on the NTS (Canadian National Topographic Sheet) system. A data set can be retrieved only one at a time by specifying its data set name or number.

This paper describes a new service that supports more general access and visualization of the CDAL data. The new CDAL system is an online service and provides easy access to a database of ground control points. The clients of this service are able to obtain responses in either HTML or XML format and to request CDAL data in either a textual or graphical form. The service accepts HTTP requests that permit a geographic region of interest to be specified as the search criteria. Only those points found within the region of interest will be returned. The second service, built on the first one, uses the textual information associated with CDAL ground control points to deliver rasterized image chips rendered for visualizing points of the CDAL.

1. INTRODUCTION

This paper describes a new service that supports more general access and visualization of the CDAL data. The new CDAL service (CDAL 2001) is an online service and provides easy access to a database of ground control points originating from the 1:50,000/1:250,000 Canada national topographic base map series (NTDB data sets), which is compiled by the Center for Topographic Information (CTI)-Sherbrooke. All that the user needs to access CDAL service are an HTML enabled browser (e.g., IE or Netscape) and Internet connection.

The new CDAL service is an online bilingual service and provides easy access to a database of 7 million ground control points. The database stores textual items in both English and French. The contents of these two data sets are identical, with exact geometry but different attributes in terms of language.

Using the existing CDAL service provided by CTI-Sherbrooke (Parent 1999, CDAL 2000) the user can only retrieve data sets based on a data set number, which corresponds to different areas where the control points belong. This method of information retrieval is not efficient since there are cases where the user wishes to retrieve the data corresponding only to a small region of interest (or even a region larger than a typical area, for that matter).

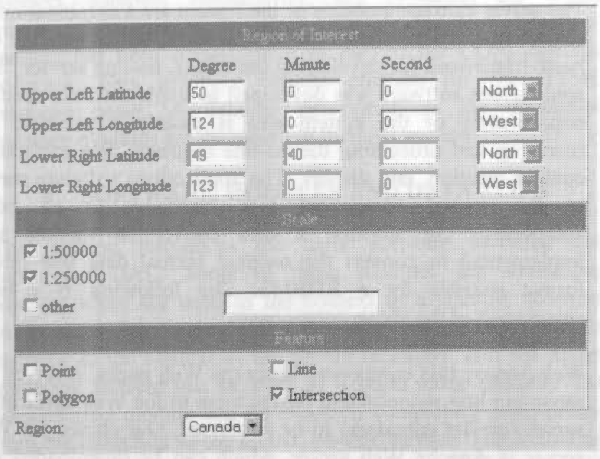
Another drawback of the current service is that the user usually receives too many records and the main task of examining the data and extracting the more appropriate data items will be left to the user as a manual operation. The new CDAL service will allow for specifying a region of interest as well as the type of control point (feature type) for information retrieval. The client will be able to obtain responses in either HTML or XML format. An image chip featuring each identified control point in the region of interest will be displayed in GIF format. Also, another image chip in GeoTIFF format will be available if the user wishes to fetch and store it on their local machine. Such a service is of economic benefit e.g., for the production of geo-coded satellite/aerial imagery, and is geared towards clients who need to use the CDAL for activities such as ground control point marking where ground control is often expensive and difficult to obtain.

An overview of the new CDAL service is presented in the next section. Section 3 discusses the system architecture. The technology used in this project, particularly existing COTS product, and a detailed discussion on the development of the system and data-flow are given in sections 4 and 5 respectively, followed by a conclusion in section 6.

2. OVERVIEW

CDAL is an online bilingual service and provides two new online services for the CGDI:

Spatial Search engine: The CDAL service allows the user to perform a spatial search and retrieve all, or only selected categories of the CDAL entities that belong to the specified geographical region. The search mechanism is capable of accepting the user-defined parameters, find the required element(s) based on the specified criteria, and pass the result into the visualization module. Figure 1 shows the input area of the CDAL front page where the user can specify the search criteria.



Region of Interest

	Degree	Minute	Second	
Upper Left Latitude	50	0	0	North
Upper Left Longitude	124	0	0	West
Lower Right Latitude	49	40	0	North
Lower Right Longitude	123	0	0	West

Scale

☑ 1:50000
☑ 1:250000
☐ other

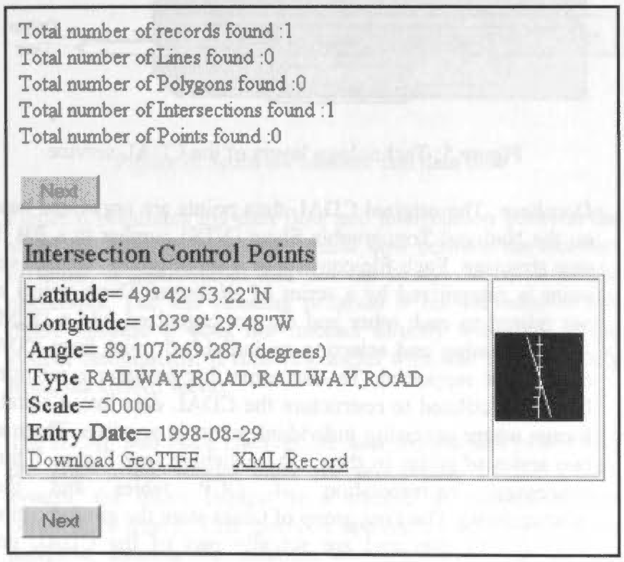
Feature

☐ Point ☐ Line
☐ Polygon ☑ Intersection

Region: Canada

Figure 1: CDAL front web page, where a query can be formed and submitted

The input area is divided into three main sections. The first section defines the Region Of Interest (ROI). The user specifies the bounding rectangle of the area for which a search should be conducted. The bounding rectangle is defined by entering the geographical co-ordinates of its upper left and lower right corner. The second section is for specifying the scale. The control point data stored in the database are derived from two different map scales (1:50000 and 1:250000). The user has the option of choosing either or both of the scales. And the third section is where the user selects the type of control point(s) to retrieve from the database. There are four types of control points to search. These are polygons, lines, intersections, and points. The most common example of a polygon control point is a lake. A bridge is an example of line control points and a well or a water tank can be considered a point. Intersections are comprised of 3 or 4 branches of roads, railways or highways (see 5.1 for more details).



Total number of records found :1
Total number of Lines found :0
Total number of Polygons found :0
Total number of Intersections found :1
Total number of Points found :0

Intersection Control Points

Latitude= 49° 42' 53.22"N
Longitude= 123° 9' 29.48"W
Angle= 89,107,269,287 (degrees)
Type RAILWAY,ROAD,RAILWAY,ROAD
Scale= 50000
Entry Date= 1998-08-29

Figure 2: Snapshot of CDAL front page containing the outputs

Visualization and Rendering: The CDAL service is able to receive, reformat and visualize the results of the search process in a textual (HTML/XML) and graphical (GIF/GeoTIFF image chips) format. Assuming that a query is submitted and a set of ground control points within the specified search area is found, then a rendering function will be triggered, which reads the textual information of the CDAL data elements and generates two contextually identical image chips. These image chips consists of a thumbnail GIF image for displaying on the front web page and a GeoTIFF image chip stored on the server ready for download by the user. Figure 2 shows an example of the textual information and the corresponding image chip displayed on the output area of the CDAL front page. At the beginning of the display page is a summary of the accumulated total number of records as well as the total number of records found for each selected feature type. The user can view the text information and the associated GIF image chip as the standard output. However, the user may perform extra actions for retrieving the GeoTIFF image chip, or the XML record of the result.

CDAL output can be used in combination with other systems to perform a specific task such as image geo-coding. It is important to note that CDAL does not interface automatically to any other system, and there will always be some level of human intervention to use CDAL output as an input to another system.

An immediate application of the service is the incorporation of the ground control points into an efficient and cost-effective geometric correction of Earth observation data. The textual descriptions provided by CDAL can be extremely helpful in identifying and marking control points. There is an overall increase in operator efficiency because control does not have to be digitized from paper map products, but rather retrieved from CDAL.

3. SYSTEM ARCHITECTURE

Figure 3 illustrates a context diagram for the CDAL system including online operation and conversion/update operation. It mostly focuses on the external interfaces of the system. During the online operation, the user communicates with the system through an HTTP link. Upon receipt of a request (query) from the user, the application runs the query against the CDAL database and returns the appropriate information to the user.

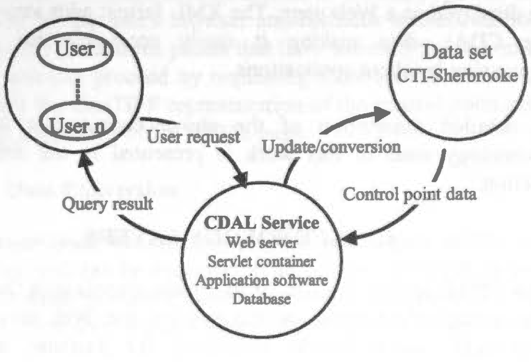


Figure 3: CDAL system context diagram

Conversion/update operation takes place whenever it is required to extract new information from the CGDI data set and insert it into CDAL database. It is not part of the online service and can be considered an administrative operation. The data stored on CGDI data set is in ASCII format. The conversion/update module extracts the control point data, converts it to suitable format for CDAL database and inserts it into CDAL database. This mechanism is performed once at the beginning of the CDAL operation (conversion) and thereafter, in a regular basis (see section 5.1 for more details).

Figure 4 illustrates the on-line operations and the inter-relationships between CDAL components. Operations depicted by dashed line are optional operations that may or may not occur depending on the user request.

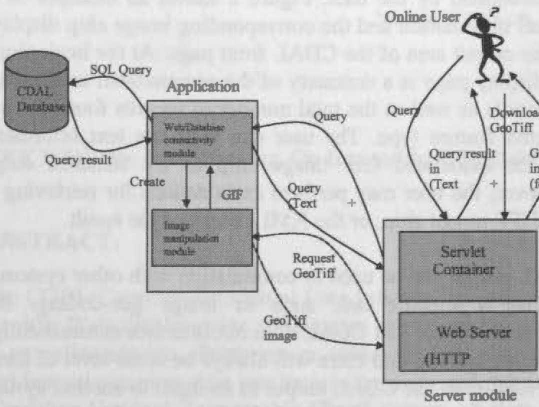


Figure 4: Operational overview of the CDAL service

The user requests the Web front page of the CDAL service by entering the CDAL URL into the web browser. Then s/he specifies a set of constraints to form a query and submits the query to the CDAL server. Upon receipt of the query, the application connects to the CDAL database to retrieve the data that satisfy the query. The result is prepared in Hypertext Markup Language (HTML) format and is sent back to the client.

Apart from the regular information that the user receives, s/he may also wish to receive additional information such as a GeoTIFF image featuring a specific control point (dashed lines). This image chip can be used for different application such as image geo-coding. By default, the information is provided in HTML format but the system allows the user to receive the information in eXtensible Markup Language (XML) format as well. The HTML format is useful if the response is to be displayed as a Web page. The XML format adds structure to the CDAL data making it easily accessible for further processing by client applications.

A detailed description of the above components and the technology used in this work is presented in the following section.

4. TECHNOLOGY LAYERS

The CDAL service consists of four main components that form the technological layers of the system: the Web server, the application server (servlet container), the database, and the application software (see Figure 5).

These are the minimum system components to fulfil all the requirements for a system such as CDAL. The web server and servlet container are used primarily for web-enabling the application. By web-enabling, we mean providing the service over the Internet so that a remote client can connect to the CDAL server, submit a query based on some pre-determined constraints, and receive the appropriate information in the desired format.

The server side components of the system are a combination of commercial off-the-shelf (COTS) products, which provide the basic framework for developing the CDAL on-line service. The applications software was developed to implement the specific requirements of the system and is mainly responsible for receiving and processing the on-line requests, data extraction, content creation, and display. The applications software serves as the central component that interfaces to all other components. In addition a data conversion module was implemented to convert the original textual data sets into a format suitable for a RDBMS. The following is a brief description of the above components.

Web Server: this component hosts the Web pages, receives the incoming http requests and passes them to the Web application server (servlet container) to be processed. Our choice for Web server is Apache Web server, which can be downloaded and installed with no cost and has proved to be a reliable. Currently, close to 70% of Web servers around the world are using Apache (Apache 2001).

Application server (servlet container): This component is used to create dynamic Web pages where the flow of information is bilateral. That is, the user can send information to the Web server, which can be used to provide the content requested by the user. In CDAL context, the user specifies the criteria for a query, which will be received by the server and will be used to perform the appropriate action (Tomcat 2001)

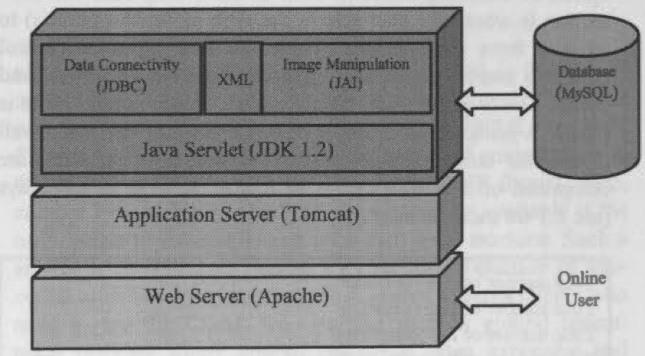


Figure 5: Technology layers of the CDAL service

Database: The original CDAL data points are organized based on the National Topographic Sheet (NTS) number in a flat file data structure. Each file consists of a set of points, where every point is categorized by a series of attributes. These points are not related to each other and are not organized into a DBMS. Also, accessing and selecting an individual point from CDAL data is not supported. To overcome this problem a relational DBMS is utilized to restructure the CDAL data into a suitable format where accessing individual points is possible. There are two series of tables in the CDAL database (see Figure 7 for a conceptual representation of GCP tables and their relationships). The First group of tables store the ground control point (GCP) data and are actually part of the CDAL core operation, whereas the other table maintains a history log that keeps track of the incoming queries, and is mainly used for bookkeeping and management purposes.

Application software: In addition to the above layers, a software layer must be implemented to deploy the CDAL specific tasks. The elements of this layer are image manipulation, XML support, database connectivity, and capturing the online queries. Additionally, an HMI (Human Machine Interface) module has been implemented as part of the CDAL-specific software. This involved content creation for the HTML response pages. A detailed description of the developed software modules are discussed in the next section.

5. SOFTWARE MODULES AND DATA FLOW

This section introduces the main software modules that comprise CDAL. Some of these modules are standalone whereas some are used in the context of a specific software component. For example the database connectivity module operates in the context of the servlet container. But the image rendering module can be seen as a stand alone module. The bookkeeping module can be executed completely independent from the operation of the CDAL on-line service. The following is a list of all software modules that are implemented in accordance with the software components.

1- HTML web front page
2- Image rendering
3- Database connectivity
4- Result web page (HTML (default) or XML)
5- Bookkeeping / statistics
6- File cleaning (to delete all the temporary files such as image chips, XML files, etc.)

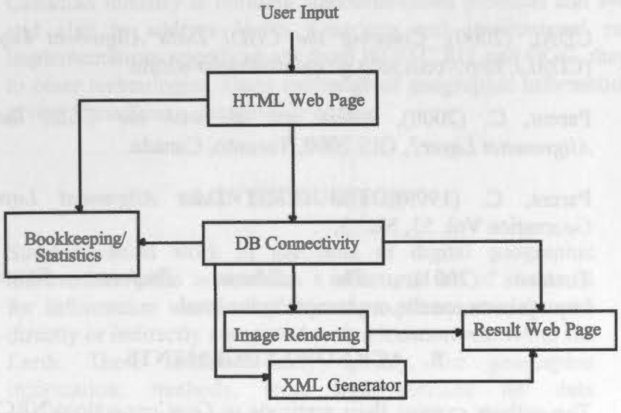


Figure 6: Software Module and data flow

Figure 6 illustrates the data flow and interrelation between the above modules. It shows the control flow between the modules in terms of input/output and not the timing sequence of the operations. The file cleaning program is not included in this figure because it does not interact directly with any of the system components. It runs in parallel with, but independently from, the CDAL server.

The *HTML web page* receives the user request and passes it to the appropriate servlet class for processing. In the CDAL system, the user's request is a number of constraints that comprise a query. This query is applied to the control point database, causing the database connectivity module to start operating. This is implemented using Java Database Connectivity (JDBC) functions, which are called from within a Java servlet class.

Another module that receives input from the front Web page is the *bookkeeping and statistics* module. The main responsibility of this module is to keep track of all hits to the Web page and the actions performed on the database, and to store the information for future reference. As the Web front page receives requests from users, it notifies the bookkeeping module of the requests so that it updates a counter and stores information (e.g., the IP address of the remote user).

The *database connectivity* module provides input for four modules, namely the bookkeeping module, the image rendering module, the XML generator, and the result Web page module. The database connectivity module can be considered the core of the CDAL system where all the useful information is extracted. This module notifies the bookkeeping module about the number of records returned for each request. It also provides the image-rendering module with information required for drawing the appropriate feature. The XML module uses the output from the Database Connectivity module to create the XML equivalent of the record. Finally, the output of the database connectivity module is used to create the text information to be displayed in the result Web page.

The main responsibility of the *image-rendering* module is to create image chips that graphically represent the control point data. Currently, two image chips with GIF and GeoTIFF formats are created. The GIF format is included in the result Web page whereas the GeoTIFF image is available for download. The module reads the geographical information such as geographical co-ordinates of the resulted point, which are stored in a text file as metadata, and then attaches them to a standard TIFF image chip, thus creating a GeoTIFF equivalent of the original GIF image.

The *XML generator* module is a static class that is responsible for creating an XML file for every control point that is contained in the result, using the specified Document Type Definition (DTD). There is a hyper link to the XML file for every control point in the result Web page. The user has the choice of downloading the XML file, or if the browser is XML enabled, the file could be viewed on the browser. The contents of XML tags are dependent on the language of choice but the tags themselves are not bilingual.

The *result web page* module creates an HTML page that can be viewed by the user's browser and includes textual information regarding the control points that have satisfied the user criteria. The user can proceed by requesting more specific information, namely the GeoTIFF representation of the control point and the XML equivalent of the HTML page.

5.1 Data Conversion

As mentioned before, this process is not part of the online service and can be executed in an off-line operation as part of administrative process. However, it is a major part of the CDAL application software and therefore it is discussed separately.

The original data set is currently stored in text format in two separate files, English and French language, and need to be converted into relational data in accordance with the proposed data model (see Figure 7). It is an important issue since the same implementation of conversion is used later for data updates when data entry is made into the textual data set. The data conversion is a two-step operation. The first step is inserting new records into the database by reading from the

English data set. This is seen as the record generation step. During this step, the French fields in the record will remain blank to be later filled in the second step. The second step is reading through the French data set, retrieving the French-specific fields, and updating the French fields in the corresponding record of the database. A unique data item is used to find the corresponding records in the French and English data sets.

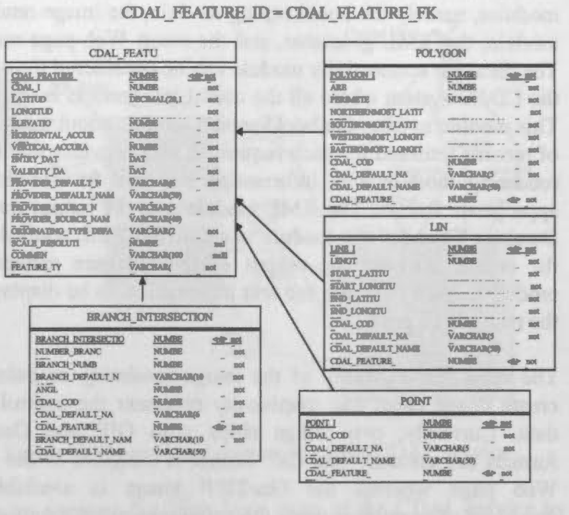


Figure 7: The CDAL data model for control points

Figure 7 illustrates the data model utilised to design the required tables and create the CDAL database. The proposed data model is in fact a modified version of the data model suggested by CTI (Centre for Topographic Information at Sherbrooke). The data are organised in a number of tables based on the type of the control point. In general, the CDAL data set consists of four classes of control points (Parent 2000):

- *Intersection Point (IN):* Point is derived from an intersection feature in the source data set, its position corresponds to the point where at least three branches (roads, railways, highways, or access roads) meet. A branch is a linear feature that has one of its extremities at the intersection.
- *Point originating from a Point feature (PP):* Point is derived from a point feature in the source data set. Its position corresponds exactly to that of the originating point.
- *Point originating from a Linear feature (PL):* Point is derived from a linear feature in the source data set. Its position corresponds to the geometric centre of the originating line.
- *Point originating from a Polygon (PS):* Point is derived from a polygonal feature in the source data set. Its position corresponds to the centre of mass of the originating polygon.

Accordingly, there are 4 tables (special tables) that store the information exclusive to a particular type of control point. These tables are Polygon, Line, Point, and Branch_Intersection. There is also one main table, CDAL_Feature that holds the common attributes of control points regardless of their type. This table is indexed on Latitude and Longitude fields.

The input text file contains data for a specific type of control point. Therefore, each run of the conversion program results in the insertion of records into the main table CDAL_Feature and one of the special tables (Point, Line, Branch_Intersection, and Polygon).

6. CONCLUSION

This paper discussed the design and implementation of a new online service that serves two purposes. The first permits searching of the CGDI Data Alignment Layer (CDAL), using a region of interest, through a Hypertext Transfer Protocol (HTTP) interface. That is, CDAL points are returned that fall in any bounding rectangle, specified by latitudes and longitudes. This service directly benefits the CGDI by providing a new CDAL access mechanism. The second service, built on the first, is a rendering service, which will visualize requested points of the CDAL. Such a service is of economic benefit for the production of geo-coded satellite/aerial imagery, where ground control is often expensive and difficult to obtain.

The operation concept developed in this project can be used in similar projects that aim at leveraging the power of the Internet to create on-line client/Server geo-spatial applications.

7. REFERENCE

Apache (2001), *Apache: HTTP Server Project*, http://httpd.apache.org/

CDAL (2001), *CDAL: A Service to Access and Visualize Data Alignment Layer*, http://cdal.mda.ca/main.html

CDAL (2000), *Creating the CGDI Data Alignment Layer (CDAL)*, http://ccdi.icdg.gc.ca/html/main-e.html

Parent, C. (2000), *Where are we with the CGDI Data Alignmenet Layer?*, GIS 2000, Toronto, Canada.

Parent, C. (1999), *The CGDA Data Alignment Layer*, Geomatica Vol. 53, No. 3.

Tomcat (2001), *The Jakarta Project: Tomcat*, http://jakarta.apache.org/tomcat/index.html

8. ACKNOWLEDGEMENTS

The authors express their gratitude to GeoConnections/NRCan (GeoInnovations program) for partial funding support of this project. RADARSAT International helped to define system requirements and assisted the project by continued evaluation of the system. We acknowledge the Center for Topographic Information-Sherbrooke for their support and providing the CDAL data set.

CANADIAN GEOSPATIAL STANDARDS IN ACTION

Fadaie [1], K., Habbane [2], M., and Tolley [3], C.
Natural Resources Canada, Canada Centre for Remote Sensing
[1] fadaie@nrcan.gc.ca, [2] mhabbane@nrcan.gc.ca, [3] ctolley@nrcan.gc.ca

Working Group IV/4

KEY WORDS: Standards, Geographic information, Geospatial data, architecture, interface, data warehouse, interoperability

ABSTRACT:

The Canadian Geospatial Standards in Action is an initiative that impels the progression of geographic information technology by showcasing Canadian standards-based products as part of an integrated geospatial architecture. This initiative reflects the strategy taken by both the International Organization for Standardization, Technical Committee 211 on Geographic Information and Geomatics (ISO/TC 211), and by Open GIS Consortium (OGC). ISO/TC 211 provides a suite of abstract standards specifications that defines framework, geospatial data models, data administration, and services, while OGC delivers spatial interface specifications that are openly available for global use and that can be implemented in products. Geospatial Standards provide connectivity amongst the user interfaces, data servers and storage warehouses in the three-tiered geospatial architecture by defining a minimum set of practices, protocols, and specifications. This allows for increased and timely access to a vast quantity of geospatial data in a usable format.

The Canadian Geospatial Standards Architecture as defined in this paper is broken into three component tiers that provide maximum portability across computer operating systems. The client tier displays information and processes graphics, communication, keyboard input, and local applications. The applications service tier encompasses a set of shareable, multi-tasking components that interact with clients, peer services, and the other data source tiers. The data source tier has the data and metadata configurations and environments.

Canadian industry is building standards-based products and system components in support of the Canadian national requirements and also to address North American and international requirements. Once Canadian products encompass the standards implementations specifications from ISO/TC 211 and OGC, they can provide their expertise worldwide in addition to gaining access to other technologies. Open exchange of geographic information improves collaboration in the geospatial data industry, ultimately giving Canadians a competitive edge.

1. INTRODUCTION

Standardization work in the field of digital geographic information aims to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth. These standards may specify, for geographic information, methods, tools and services for data management (including definition and description), acquiring, processing, analyzing, accessing, presenting and transferring such data in digital/electronic form among different users, systems and locations. The work links to appropriate standards for information technology and data where possible, and provides a framework for the development of sector-specific applications using geographic data.

Through an iterative process and by building on previous standardization work, it is possible to develop new standards that guide (prescribe) new technology but support existing systems and data. The iterative process of standards development builds upon existing standards and refines them to develop new, broader standards.

International Standards in Geographic Information and Geomatics have been developed under the umbrella of the International Organization for Standardization Technical Committee 211 on Geographic Information and Geomatics (ISO/TC 211). ISO/TC 211 provides a suite of standards that defines framework, geospatial data models, data administration, and services.

The purpose of ISO/TC 211 is to develop mature concepts for GIS components. The "first generation" GIS standards being created within ISO/TC 211 are, for the most part, written at an "abstract specification" level. It is the goal of an ISO/TC 211 standard to communicate the objects, relationships, behaviors, and attributes associated with the particular component at hand. OGC's (Open GIS Consortium) mission is to deliver spatial interface specifications that are openly available for global use and that can be implemented in products. OGC manages both Abstract Specifications and a growing suite of Implementation Specifications.

2. GEOSPATIAL STANDARDS

2.1 Standards Program of Work

Scientists in Canada are conducting research in developing Geomatics Standards through the Canadian General Standard Board Committee on Geomatics (CGSB CoG), Canadian Advisory Committee to ISO/TC 211 (CAC to ISO/TC 211), and the International Organization for Standardization Technical Committee on Geographic Information and Geomatics (ISO/TC211).

Table 1 stresses the ISO/TC 211 work program as updated of March 2002. New working groups have been added to include new projects such as sensor and data models for imagery, location based services and registry of geographic information items.

Table 1. ISO/TC 211 Work Program – March 2002

ISO Standard Number and Title	Most recent document available on the TC 211 Web Site	Availability of DIS
WG 1 – Framework and Reference Model		
19101 Reference model	DIS 19101	2000-10
19102 Overview	Deleted	Deleted
19103 Conceptual schema language	211N1082	TS 2001-12
19104 Terminology	211N1130	2001-12
19105 Conformance and testing	ISO 19105: 2000	1999-08
19121 Imagery and gridded data	ISO/TR 19121:2000	TR 2000-10
19124 Imagery and gridded data components	211N1017	Review Summary 2001-01
WG 2 – Geospatial Data Models and Operators		
19107 Spatial schema	DIS 19107	2001-06
19108 Temporal Schema	DIS 19108	2000-11
19109 Rules for application schema	DIS 19109	2002-01
19123 Schema for coverage geometry and functions	211N1227	2002-08
WG 3 – Geospatial Data Administration		
19110 Feature cataloguing methodology	DIS 19110	2001-07
19111 Spatial referencing by coordinates	DIS 19111	2000-12
19112 Spatial referencing by geographic identifiers	DIS 19112	2001-10
19113 Quality principles	DIS 19113	2001-02
19114 Quality evaluation procedures	DIS 19114	2001-08
19115 Metadata	DIS 19115	2001-09
WG 4 – Geospatial Services		
19116 Positioning services	211N1132	2002-02
19117 Portrayal	211N1114 Proposed text for DIS	2001-11
19118 Encoding	211N1136 Proposed text for DIS	2002-02
19119 Services	DIS 19119	2001-11
19125-1 Simple feature access - Part 1 Common architecture	DIS 19125-1	2000-11
19125-2 Simple feature access - Part 2 SQL option	DIS 19125-2	2000-11
19125-3 Simple feature access – Part 3 COM/OLE option	211N940/ 997	2001-
19128 Web map server interface	211N939/ 996	2002-06

ISO Standard Number and Title	Most recent document available on the TC 211 Web Site	Availability of DIS
WG 5 – Profiles and Functional Standards		
19106 Profiles	211N1134 Proposed text for DIS	2001-12
19120 Functional standards	ISO/TR 19120:2001	TR 2001-07
WG 6 – Imagery		
19129 Imagery, gridded and coverage data framework	211N1176	TS 2003-07
19130 Sensor and data model for imagery and gridded data	211N1167	2003-09
WG 7 – Information Communities		
19120/A1 Functional standards – technical amendment	211N835/ 888	TR 2002-04
19122 Qualifications and certification of personnel	211N573	TR 2003-02
19126 Profile – FACC data Dictionary	211N834/ 887	2002-12
WG 8 – Location based services		
19132 Location based services possible standards	211N1088/ 1139	RS 2002-01
19133 Location based services tracking and navigation	211N1089/ 1140	2002-05
19134 Multimodal location based services for routing and navigation	211N1117/ 1174	2003-05
WG 9 – Information Management		
19127 Geodetic codes and parameters	211N1161	TS 2002-11
19131 Data product specifications	211N1087/ 1138	2003-09
19135 Procedures for registration of geographical information items	211N1122/ 1175	2002-12

2.2 Standards in Action Demonstration

ISO/TC 211 resolved in several resolutions (e.g. Resolution 160) to arrange a workshop on standards implementation experiences. The first workshop was held on Wednesday, March 7th, 2001, prior to the plenary in Lisbon. ISO/TC 211 invited the national bodies of Australia, Belgium, Canada, Germany, Russia, Switzerland, USA, and other member bodies to prepare presentations and to assess the applicability of standards testbeds at the workshop in Lisbon.

The National Body for Canada, lead by the Canada Centre for Remote Sensing of Natural Resources Canada, has created a demonstration site to showcase the components of a standards-based distributed spatial architecture. The components are based on and related to the ISO/TC 211 19100 suite of standards.

The Standards In Action (SIA) demo comprises several implementations as follows: Distributed navigation, data access, viewing, and downloading as an implementation of Simple Feature Access (ISO 9125-1 & 19125-2), Metadata (ISO 19115), and Spatial Referencing by Coordinates (ISO 19111); Geospatial Standards-based data products, related to ISO 19107 (Spatial Schema), ISO 19110 (Feature Cataloging Methodology), and ISO 19111 (Spatial Reference by Coordinates); Multilingual support for meta-content as per ISO 19115 (Metadata); Open Geospatial Data Store Interface (OGDI), related to ISO19119 (Services); Geographic Markup Language (GML), related to the ISO 19118 (Encoding); Electronic Nautical Charts (ENC), related to the ISO 19117 (Portrayal); Self-Defining Structure (SDS) and Helical Hyperspatial Codes (HH Codes) as a type of coverage function related to the ISO 19123 (Schema for Coverage Geometry and Functions), 19124 (Imagery and Gridded Data Components), and ISO 19129 (Imagery, Gridded and Coverage Data Framework).

The demonstration provides a concrete example of integration of standards-based products that results in access to a wide array of information sources and services over the Internet. This demonstration has been given in several venues in different countries and has generated significant interest in those products demonstrated. It is an example of a success story for government, private sector and consortia such as 3i, working together to create and demonstrate Canada's ability to deliver state-of-the-art solutions.

2.3 Standards Architecture

Geospatial Standards provide connectivity between the user interfaces, data servers and storage warehouses in the three-tiered geospatial architecture by defining a minimum set of practices, protocols, and specifications. The Canadian geospatial standards architecture is broken into the following component tiers:

- Client Tier: the client component displays information and processes graphics, communications, keyboard input, and local applications. These interfaces provide maximum portability across computer operating systems;
- Applications Services Tier: a set of shareable, multi-tasking components that interact with clients, peer services, and the data source tier;
- Data Source Tier: the data and metadata configurations and environments.

This architecture stresses the components of the overall storage and query architecture with the Oracle/CubeSTOR cartridges as a distributed database environment. The application server tier components include the Feature Manipulation Engine (FME), the Open Geospatial Datastore Interface (OGDI) server, the CubeServ server, ArcIMS server, and Map Manager Server, with associated Application Programming Interfaces (APIs).

Clients link to CubeServ/ArcIMS/Map Manager Server and other OGC compliant Web Map servers through the APIs using five different protocols and mechanisms:
Geographic Library Transfer Protocol (GLTP) for the OGDI; Hypertext Transfer Protocol (HTTP) for the Open GIS Consortium (OGC) Get Protocols; Tomcat Java server to interpret calls from the Unified Service Interface (USI) and the ArcView-USI extension; Sequel Query Language (SQL)/Open Database Connectivity (ODBC) to interpret queries from Helical Hyperspatial Viewer; and Parsed html to interpret calls from Service Manager, ArcIMS, Percipio, and MetaMiner.
Clients also use metadata tools to populate and update databases and Z39.50 server protocol to connect to different FGDC clearinghouse. Clearinghouses could be created and managed using MetaMiner from Compusult. Several Canadian companies have developed tools for managing metadata. Some allow documentation of data and others provide user access to the data, using protocols such as ANSI Z39.50. As a component system, these products catalogue/organize, locate, and showcase data. At the data source tier, data warehouses receive, archive, catalogue, maintain and disseminate data holdings. Metadata search tools harvest from the data sources tier, providing data connectivity to relational database management systems using Z39.50 protocol. The search tools provide search specification fields, as well as result processing. Once data is collected and documented using metadata tools, it can be displayed, linked, distributed and published on the client tier. Currently, Canadian products support the US FGDC metadata standard and the Canadian 171/3 metadata standard, but they are being adapted to support the full ISO 19115 metadata standard and the joint Canada/US implementation specification.

The overall architecture is related and implements OGC implementation specifications and the ISO/TC 211 suite of standards, namely: ISO 19110 Geographic Information - Feature cataloguing methodology; ISO 19111 Geographic Information - Spatial referencing by coordinates; ISO 19115 Geographic Information – Metadata; ISO 19117 Geographic Information – Portrayal; ISO 19118 Geographic Information – Encoding; ISO 19119 Geographic Information – Services; ISO 19123 Geographic Information - Schema for coverage geometry and functions; ISO 19125-1 Geographic Information - Simple feature access, common architecture; ISO 19125-2 Geographic Information - Simple feature access, SQL option; and ISO 19128 Geographic Information - Web map server Interface.

Standardized products for Web-mapping and imagery, gridded and coverage data are being developed, and are embedded within the geospatial standards architecture. Web based services provide access to geographic information by way of interpretable infostructures. Canadian products support web-mapping services, including grid coverage and simple feature access within a database schema that handle multi-dimensional datasets. Map servers provide user access to spatial warehouses, while allowing OGC compliant servers to cascade and access images and additional data from other OGC compliant map servers.

2.4 Imagery

The area of Imagery and Gridded Data (I&GD) is one of the most challenging within ISO/TC 211. While most data is organized in simple grids, there are many different traversal methods for grids and structures that support the distribution of attributes over an area. In addition, the data represents the sensors from which it was collected or, in the case of synthetic data, the source from which it was generated. Sensor models and associated georeferencing are an important aspect of Imagery, Gridded and Coverage Data (IGCD) geographic information.

The standards and industrial specifications for Imagery, Gridded and Coverage Data (IGCD) vary in a different way. Almost all have different encoding mechanisms that may be needed for various exchange media or storage mechanisms. However, in general, one encoding mechanism can be converted to another. The more fundamental differences occur at the "Content Model" level. But a well-defined number of content structures, together with as much common metadata as possible, will result in a degree of alignment between the different standards.

It is not possible to endorse one standard or industrial specification, or to come up with a new and more comprehensive standard because of the very large volumes of data that exist in the various formats already in use. Nor is the solution to build a very flexible all-encompassing standard with a broad array of options; data sets can use incompatible subsets of the same overall standard.

The direction must enhance compatibility while at the same time retain compatibility with the widely used existing industrial standards. This is very complicated and can only be partially achieved. However, it is easier to build converters between a small numbers of recognized choices than between an open-ended spectrum of incompatibilities.

The Imagery architecture in Canada aims to harmonize the structure and the way Canadian companies such as PCI Geomatics, CARIS, Helical Systems, and CubeWerx have been handling imagery and gridded data. The architecture encompasses either direct access to imagery data sets using standards of the shelf applications like Geomatica Focus or HH Viewer, or Web mapping services with OGC compliant clients and Java Viewers. The mid-architecture also includes Web servers such as CubeServ, Spatial Fusion or Spans servers. The Data itself are stored in either a RDBMS systems that include Cartridges (i.e. CubeStor Spatial, Oracle Spatial Fusion, Oracle GeoImage) or Flat files with an encapsulation schema that handles multiple attributions for the same geometry. This encapsulation schema also embraces what is known as Helical Data Fusion", which corresponds to the association of several attributions sharing the same geometry. In fact, there are several distinct types of data fusion, for example, the data corresponds to different attributes associated with the same geometry, within one architecture or file. In others, the data consists effectively of repeated measurements of different types of attributes that are assembled together using overlay techniques, which were formerly known as data" compilation or data assimilation. Such manipulations are mostly seen in Geographic Information Systems (GIS) available on the market.

The data is also stored in two manners: As a relational tables' record as a "BLOB" (e.g. GeoTIFF in an ORD Image class) or a Grid Value Cell (set of records in CubeStor or Spatial Fusion). As a set of record (Grid Value Cell "Grid", Point Value Pair "TIN") in a Flat file (Self Defining Structure Architecture).

2.5 Strategy for Standardization

There are four levels in standards and standardization that could be identified:

Level 1 deals with "Abstract" standards, meaning that a set or suite of standards, ISO/TC 211 per se includes models (e.g. UML diagrams) that describe pieces of information and how they are related to each other (see for example the Reference Model, ISO 19101). These are high level standards where an organization such as Natural Resources Canada should be involved by attending and participating in building those standards.

Level 2 deals with "National or Regional" standards, meaning that National bodies (e.g. Canada and the US) should develop their own standards based on ISO/TC 211 and other standards. National or regional standards are in fact profiles, or subsets of International Standards like ISO/TC 211 suite. National Bodies should then adopt/adapt and implement these standards. One of these standards is ISO 19105 metadata, which defines metadata elements, provides a schema, and establishes a common set of metadata terminology, definitions, and extension procedures. This standard is being implemented within infrastructure components such as CGDI and the FGDC clearinghouse. ISO/TC 211 urges countries and organizations to implement and test the standards and to report the results to ISO/TC 211 so that the ISO/TC 211 standards can be refined.

Level 3 deals with "Specific" standards, meaning that an organization such as Canada Centre for Remote Sensing of Natural Resources Canada should be involved in building a specific ISO work item as Imagery, Gridded and Coverage Data Framework (ISO 19129) and Sensor and Data Models for Imagery and Gridded Data (ISO 19130). Those upcoming standards handle remote sensing imagery, and other requirements such as metadata and quality elements for gridded data (e.g. ortho-rectified imagery, and the Digital Elevation Models). This is a specific requirement for handling (archive, document, index and diffuse) large data sets such as remote sensing imageries.

Level 4 deals with "Business and Marketing", meaning to develop a business and marketing plan for standard-based technology, and encourage members to market all Draft International Standards as soon as they are available. When implemented by a data producer, these International Standards will:

1. Provide data producers with appropriate information to characterize their geographic data properly;
2. Facilitate the organization and management of metadata for geographic data (vector and Grid);
3. Enable users to apply geographic data in the most efficient way by knowing its basic characteristics;
4. Facilitate data discovery, retrieval, and reuse. Users will be better able to locate, access, evaluate, purchase, and utilize geographic data; and
5. Enable users to determine whether geographic data in a holding will be of use to them.

2.6 Ongoing Research And Applications

Canada is a leading nation in both geospatial science and technology development. As Canada moves further into the knowledge-based economy, many organizations will require better access to geospatial data for decision-making. Geospatial information in the framework of ISO/TC 211 leads to greater productivity because it is used for a range of purposes, including analyzing markets, deciding where to locate new plants and offices and improving transportation systems. Canadian companies continue their commitment to standards development through the Open GIS Consortium (OGC). Most of the OGC's work is aligned with ISO/TC 211. The standards specifications will make the key issue of interoperability much easier for businesses, citizens, and governments to find, view, pan, zoom, overlay, and query geographical images and maps on the Worldwide Web. The standard specifications passed by the members of the OGC will allow maps and map queries to become a significant component of the Web.

There is a need for standardized technologies both at the content and the encoding levels, while allowing functionality between existing products. A standards-based architecture can provide end users with a direct, mandated link to geomatics and remote sensing technologies and data sets. This can provide decision-makers with a vital tool for effectively addressing issues such as local emergencies.

2.7 Conclusions

Canadian industry is building standards-based products and system components in support of the Canadian national requirements and also to address North American and international needs. Many of the products address the storage, access and display of geographic information, and Canadian companies have been active participants in the OGC test bed projects, through the Information Interoperability Institute (3i), a Canadian technology consortium. Once Canadian products encompass the standards implementations specifications from ISO/TC 211 and OGC, they can provide their expertise worldwide in addition to gaining access to other technologies. Open exchange of geographic information improves collaboration in the geospatial data industry, ultimately giving Canadians a competitive edge.

3. SELECTED BIBLIOGRAPHY

3.1 References from Websites

http://www.ccrs.nrcan.gc.ca/ccrs/data/standards/welcom_e.html

http://www.isotc211.org

4. ACKNOWLEGMENTS

The authors would like to thank our Canadian experts in geospatial Standards. This initiative is the result of work of experts such as Henry Kucera (Swiftsure Spatial Systems), Herman Varma (Canadian Hydrographic Service), Doug, C. O'Brien (IDON Technologies), Yves Hudon (Conseil du trésor du Québec), Jean Brodeur (CTIS), Jeffery Stockhausen (Helical Systems), and Michel Mellinger (3i).

5. TRANSMITTAL AND FURTHER INFORMATION

5.1 Further Information on the Standards in Action initiative

If you have questions about the standards in action initiative, please contact us at Fadaie@nrcan.gc.ca

AN EFFICIENT DATA MANAGEMENT APPROACH FOR LARGE CYBERCITY GIS

Qing Zhu, Xuefeng Yao, Duo Huang, Yetin Zhang

National Key Lab of Information Engineering for Remote Sensing, Surveying and Mapping,
Wuhan University 430079, P.R. China
E-mail: zhuq@rcgis.wtusm.edu.cn

Commission IV, WG IV/4

KEY WORDS: CyberCity, Spatial Database Engine, R+_trees, Object-relational Database.

ABSTRACT:

CyberCity facilitates the processes of urban planning, communication system design, control and decision making, tourism, etc. However, the high efficient database management has become a bottleneck of CyberCity applications. This paper proposes an efficient approach to manage the integrated databases of large CyberCity. This approach consists of following three schemes: At first, a special R+_tree index was designed to accelerate spatial retrieving. The spatial index of CyberCity includes three different types of indexes, i.e. 3D object index, DEM index and image index. The whole city is divided into rectangular regions, and geometries are then classified into the regions by the center of the rectangular bounding box of each geometry. We call it a R+_tree index because among the bounding boxes of local regions has no intersection. And among all the leaf nodes of the R+_tree (geometry records) there is no repetition. Secondly, different data compression algorithms are adopted to compress the digital elevation models, 3D vector models and images, such as LZ77 lossless compression algorithm for compression of vector data and JPEG compression algorithms for texture images. Thirdly, in order to communicate with the Oracle8i database, the CyberCity GIS spatial database engine (SDE) is designed. At last, based on the SDE prototype a case study is presented. It is hopeful to satisfy the requirement of real time applications of CyberCity GIS. It is proved to establish the efficient spatial index and to adopt proper compression methods as well as to extend the data retrieve strategy of commercial ORDBMS are significant for large CyberCity GIS.

1. INTRODUCTION

The term 'CyberCity' is used to represent the virtual representation of a city that enables a person to explore and interact, in cyberspace, with the vast amounts of environmental and cultural information gathered about the city (Gruber, 1999; Zhu et al., 2000). It facilitates the processes of urban planning, communication system design, control and decision making, tourism, etc. The CyberCity not only shows data in three dimensions, i.e. the 3D city models in most cases, but also presents photorealistic surface description. Therefore, the description of surface character and material parameters, including geometry, photo texture and additional information, are the contents of a CyberCity database, and this would result in the CyberCity database of an entire city facing an amount of some hundred gigabytes of data. Even very complex and large scenes are organized as a collection of files. The access efficiency of the file system is too low to afford the development of the CyberCity, the database management system has been required to manage the huge urban data of CyberCity. The high efficient database management has therefore become a bottleneck of CyberCity applications (Kofler *et al.* , 1996). The relational databases have been established as the most important database technology during the last 30 years, which are widely used in nowadays 2D geographic information systems (GISs). But to the 3D GIS, the disadvantage of relational database has become more and more obvious. Because the CyberCity has various data types, which include many numeric item data, string data, large unstructured data such as texture images of surfaces, structured vector data of 3D objects, and large size of terrain orthoimages. The object-oriented database is still not mature to organize all such kinds of data and their relations, so we choose the object-relational database, such as Oracle 8i, which has the mechanism to organize the individual data into some kind of objects, it reserves the advantage of relational database and adapts the idea of object-oriented into it. It can not only handle the complex relations of the 3D data, but also is good at forming the data into objects which should be treated as a whole logically. On the other hand, the access efficiency of Oracle8i is optimistic and more suitable for CyberCity GIS.

This paper is organized as following: after briefly introducing the architecture of CyberCity geographic information system (CCGIS), the R+_tree spatial index method, data compression algorithms and spatial data engine (SDE) based on object-relational database Oracle8i are proposed respectively. At last, based on the

SDE prototype a case study and conclusion will be presented.

2. THE ARCHITECTURE OF CCGIS

The construction of CyberCity requires 3D city models with realistic material or texture description and related attributes and to provide spatial query and analysis functions. There are usually three basic models involved in the CCGIS, including Digital Elevation Model (DEM), Digital Orthoimage Map (DOM) and 3D Object Models like buildings, roads, pipeline etc. Unlike the true 3D GISs needed in geological or oceanic applications, the main objective of current CyberCity GIS is to deal with the 3D objects as surface models (sometimes they are considered as 2.5D models). It does not matter how the 3D real world is mapped into spatial databases, the main point is the availability of 2.5D and 3D capability in GIS what should be realized in an efficient and robust manner (Fritsch, 1996). As shown in figure 1, the architecture of CCGIS consists of three layers, the top layer is the 3D real time applications such as 3D dynamic interactive visualization (Zhu, 1998). The data access depends on the mid layer, i.e. the CCGIS-SDE, which majors in the communication with the bottom layer, i.e. the database management system (DBMS), which manages the data of CCGIS in database. The database is an object relational database, i.e. Oracle 8.1.6. This SDE extends the data retrieve functionality, and provides an efficient data management approach including special R+_tree index, data compression schemes and fast access strategy as mentioned below.

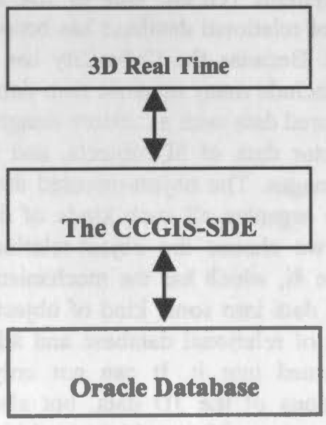


Figure 1 The architecture of CCGIS

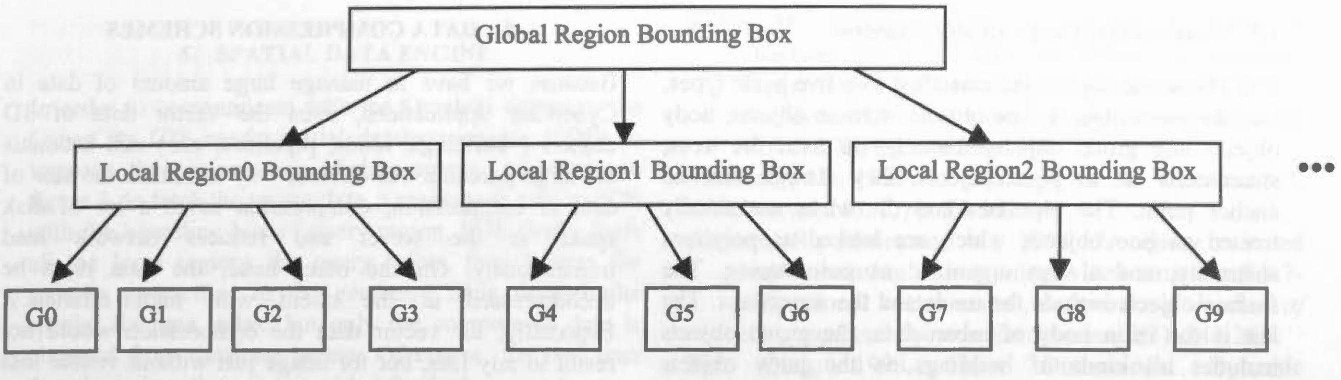
3. R+_TREE INDEX FOR FAST DATA RETRIEVE

As mentioned above, the data volume of a whole city is huge, if we want to retrieve data from the vast database without spatial index, this process would be very time-consuming, so proper spatial index must be built up to accelerate the data locating. The authors propose to use R+_tree as the basis for a CyberCity GIS to develop a data structure to spatially organize large amounts of 3D data. Herein we call it a R+_tree index because among the bounding boxes of local regions has no intersection. And among all the leaf nodes of the R+_tree (geometry records) there is no repetition, in order to keep the data consistence of the database, one geometry in one city is exclusive. As shown in figure 2, the spatial index of CyberCity includes three different types of indexes, i.e. 3D object index, DEM index and image index. The whole city is divided into rectangular regions, and geometries are then classified into the regions by the center of the rectangular bounding box of each geometry. A global table which manages all the rectangular regions of a city is established to locate the local regions with the bounding box information. Simultaneously several management tables are set up according to the number of local region, each local region has a management table to manage all the geometry tables in this rectangular region. To insert geometries into the geometry table, the proper local region and proper geometry table based on the center of the rectangular bounding box of each geometry is selected.

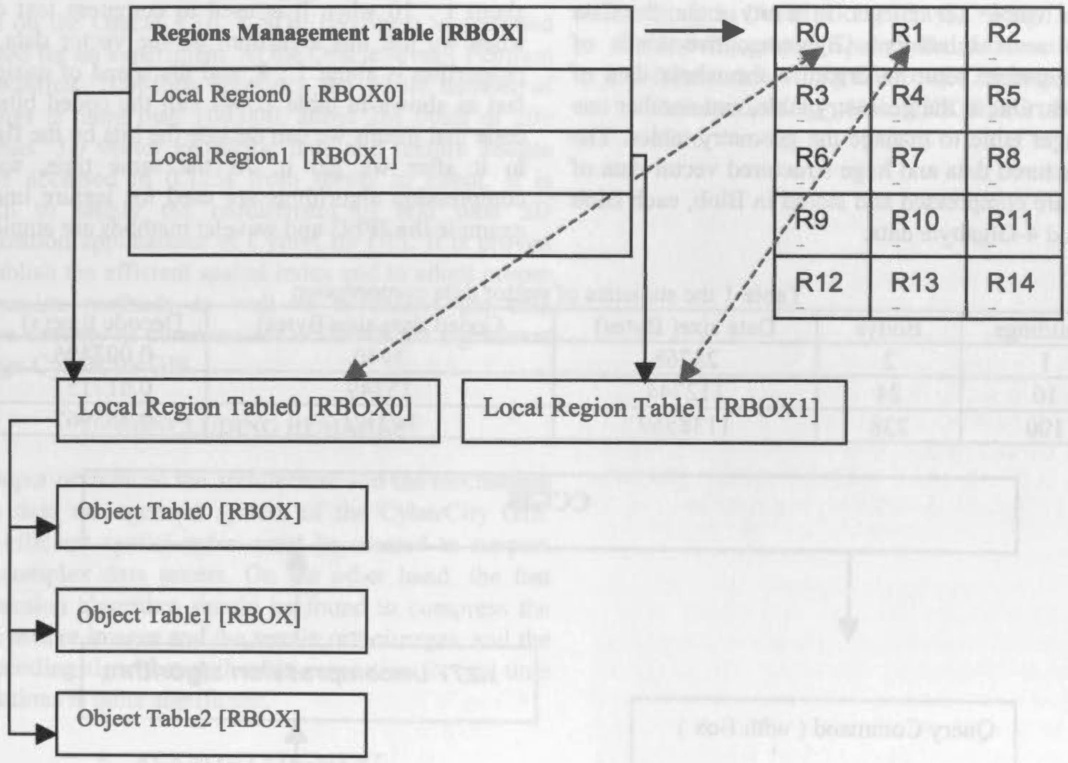
3.1 Management of large DEM

A so-called gridded manner algorithm was designed to handle large DEM as raster data. In gridded manner method, tile is the logical subdivision of the whole region and server as the foundation of data index. Block is the basic data storage and access unit. For 3D real time visualization applications, block is also the basic rendering and LOD (levels of detail) process unit. When blocks are stored every block is aligned with each other in tile and neighbour block and an overlay line data in four directions. Cell is the minimum subdivision of DEM data. It consists four original elevation points. All block within a specific grid interval constitute a layer. In another word layer is the DEM dataset of the same grid interval at a certain scale or resolution. LOD is an important concept. As far as DEM data is concerned, LOD relates to various grid interval DEM data (DEM pyramid). A large grid interval layer is created from the neighbouring minor grid interval layer. All layers have the same structure that is composed of blocks.

In object-relational DBMS, one layer data is stored in one table and the friend type of DEM data stream is large object binary (Lob). In modern DBMS, Lob is a popular data type. Though Lob is an unstructured data type, it can be trimed or appended arbitrarily and can store up to 4GB in some DBMS. So Lob is convenient to store large unstructured data, for instance, the image and video. In DEM layer table an elevation value in Lob field can consist of 2,3 or 4 bytes. When source data is submitted to DEM database, it is decomposed to blocks and each of them is indexed according to database original point and correspond (x, y) block width and height. This process


Global Region Bounding Box
Local Region0 Bounding Box
Local Region1 Bounding Box
Local Region2 Bounding Box
•••
G0
G1
G2
G3
G4
G5
G6
G7
G8
G9

(a) The concept of R+_tree


Regions Management Table [RBOX]
Local Region0 [RBOX0]
Local Region1 [RBOX1]
R0
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
Local Region Table0 [RBOX0]
Local Region Table1 [RBOX1]
Object Table0 [RBOX]
Object Table1 [RBOX]
Object Table2 [RBOX]

(b) The table structure

Figure 2 The R+_tree index

creates a series of float-formatted elevation data stream. Each block stream is the minimum data storage unit in Lob field. After all source data is stored in DEM database the initial DEM layer table is created.

The most advantage of this kind of DEM database is that the DEM data can be accessed arbitrarily based on block unit. Because of the limited data size of a block, the gridded manner means that we can roam in an unlimited DEM data freely, irrespective of region area or data size.

3.2 Management of large DOM

The process of DOM data is similar with that of DEM. Both the original points, the block structures and the layer structures of DOM database are identical with DEM's. It is apparent that DOM database is bound up with DEM database. These features result in the same query scope in DEM database can get the DOM data in the same region within DOM database. Because the resolution of the orthoimage is usually higher than that of DEM at the same layer, the data volumes of each block at the same layer are then quite different. For example, a DEM block may consists of 1K nodes as 4K bytes elevation values, but a DOM block may involve 64K pixels as 196K bytes RGB color values.

3.3 Management of large 3D object models

The 3D vector objects are classified into five basic types, i.e., the point objects, line objects, surface objects, body objects and group objects. Usually we treat the trees, streetlamps etc as point objects, they always have an anchor point. The pipelines and the wires are usually treated as line objects, which are looked as polylines abstractly, and always organized as point string. The surface objects include the roads and the waterways. The last is the main body of urban data, the group objects includes all kinds of buildings as the body objects generally are treated as the element of group objects, one group object represent one building, it can be composed of one to several body objects. We use the five types to organize the whole 3D objects. In a city scale, the data amounts to several hundred Gigabytes, two kinds of tables are therefore setup to organize the whole data of the five types. One is the geometry table, and another one is the manager table to manage the geometry tables. The large unstructured data and huge structured vector data of 3D objects are compressed and stored in Blob, each Blob item can hold 4-Gigabyte data.

4. DATA COMPRESSION SCHEMES

Because we have to manage large amount of data in CyberCity applications, even the vector data of 3D objects (buildings, roads, pipelines, etc.) still accounts for large percents. An obvious way to reduce the size of data is compression, compression saves a lot of disk space at the server and reduces network load tremendously. On the other hand, the data must be decompressed at the client with high efficiency. Especially, for vector data the compression would not result in any loss, but for image just without visible loss of quality is ok. Therefore, we choose LZ77 compression algorithm for vector data, the LZ77 is a dictionary compression algorithm without loss. The zip proportion is about 1 : 10 when it is used to compress text data, but when we use this algorithm on the vector data, the zip proportion is about 1 : 8, and the speed of unzip is very fast as shown in table 1. We call the coded bits instant-code that means we can decode the bits by the flag stored in it after we get it. At the same time, some loss compression algorithms are used for texture images, for example the JPEG and wavelet methods are employed.

Table 1 the statistics of vector data compression

Buildings	Bodys	Data size(Bytes)	Coded data size(Bytes)	Decode time(s)
1	2	28768	3940	0.003496
10	24	112744	15349	0.017157
100	238	1138569	142321	0.155967

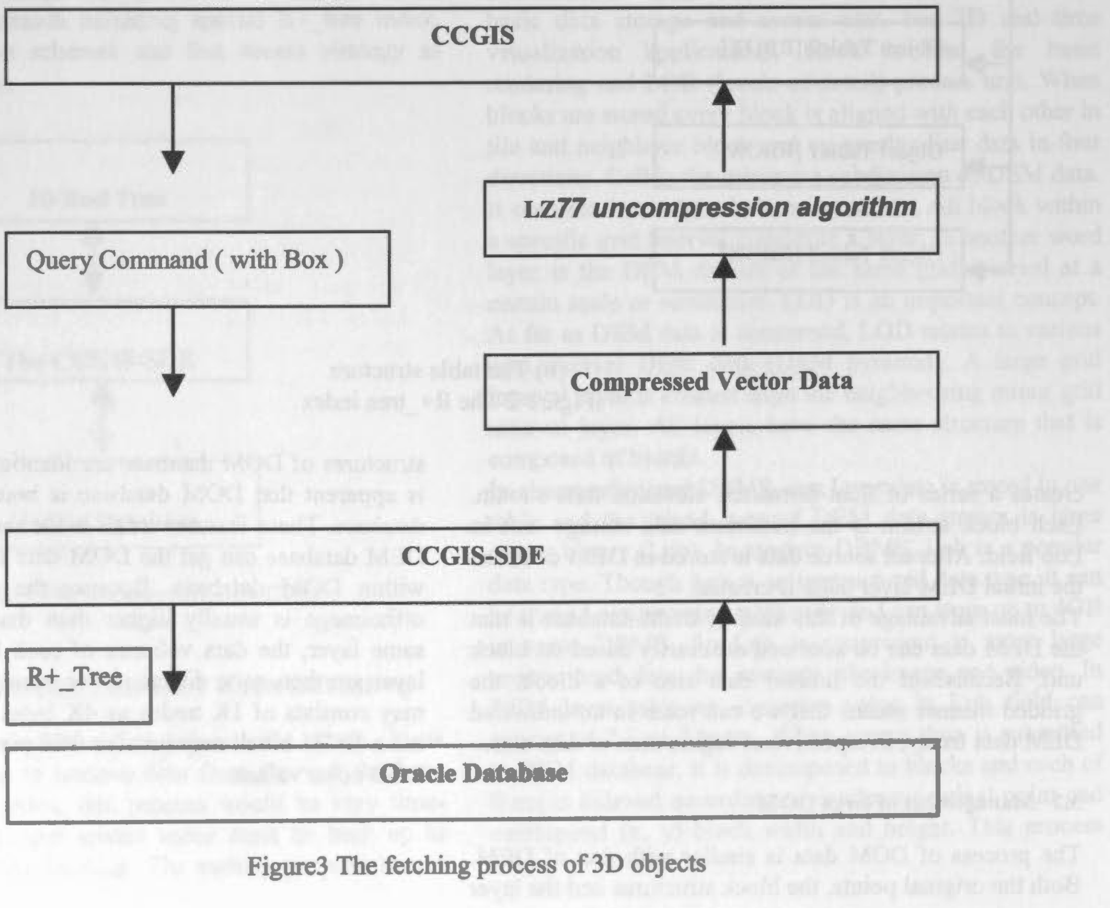


Figure3 The fetching process of 3D objects

5. SPATIAL DATA ENGINE

In order to communicate with the Oracle8i database, the CyberCity GIS needs spatial database engine (SDE) to improve the performance of data access. As shown in figure 3, to fetch the vector data, a message is sent to SDE with the bounding box of query region, SDE firstly finds all the local regions the query refers, then locates the specific geometries in the geometry table through the special R+_tree index, but only the compressed data is obtained from database hereto. Of course, the real time decoding algorithms are provided for further process.

6. CASE STUDY

Based on the Oracle 8.16, a SDE prototype is designed and used for an experiment. At the Oracle sever (Pentium III, 600Mhzs, 100Mbps network), the whole number of buildings is more than 100,000, about 2M bytes of 200 buildings' 3D vector geometric data and texture images can be accessed in 0.125s from server to client. It is hopeful to satisfy the requirement of real time 3D visualization applications of CyberCity GIS. It is proved to establish the efficient spatial index and to adopt proper compression methods as well as to extend the data retrieve strategy of commercial ORDBMS are significant for large CyberCity GIS.

7. CONCLUDING REMARKS

This paper introduced the architecture and the mechanism of the data management system of the CyberCity GIS. More efficient spatial index must be created to support more complex data access. On the other hand, the fast compression algorithm should be found to compress the object texture images and the terrain orthoimages, and the fast decoding algorithm with multi-resolution for real time applications is quite significant.

8. ACKNOWLEDGMENT

This project is supported by the NSFC (No.40001017), the Fok Ying Tung Education Foundation (No.71017) and the major state basic research program (No.G1999043801).

9. REFERENCES

Fritsch D., 1996, Three-Dimensional Geographic Information Systems—Status and Prospects, In: *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXI, Part B3 , pp. 215-221.

Gruber M and Wilmersdof E., 1997, Urban Data Management - A Modern Approach, *Computers , Environment and Urban System*, 21(2):147-158.

Gruber M. 1999, Managing Large 3D Urban Databases, *47th Photogrametric Week,* (Dieter Fritsch and Rudi Spiller, editors), Wichmann Verlag, Germany, pp.341-349.

Kofler M., Rehatschek H. and Gruber M., 1996, A Database for a 3D GIS for Urban Environments Supporting Photorealistic Visualization, Int. Archives of Photogrammetry and Remote Sensing, Vol. XXXI, B2, Vienna 1996.

Molenaar, M., 1992. "A topology for 3D vector maps", *ITC Journal,* No. 1, pp. 25-33.

Rikkers, R., Molenaar, M., 1994. "A query oriented implementation of a topologic data structure for 3-dimensional vector maps", *International Journal of Geographical Information Systems*, 8(3): 243-260.

Zhu Qing, 1998, Three Dimensional Dynamic Visualization Model, *Journal of Wuhan Technical University of Surveying and Mapping (in Chinese)*, 23(2):83-87.

Zhu Qing, Li Deren, Gong Jianya, Xiong Hanjiang, 2000, The Integrated Spatial Databases of GeoStar, In: *International Archives of Photogrammetry and Remote Sensing* , Vol. XXXIII, Part B4, pp. 1243 – 1246.

EXTENDING GEOINFORMATION SERVICES: A VIRTUAL ARCHITECTURE FOR SPATIAL DATA INFRASTRUCTURES

Javier MORALES, M. Mostafa RADWAN

International Institute for Geo-Information Science and Earth Observation (ITC)
P.O. Box 6, 7500 AA, Enschede, the Netherlands
(jmorales, radwan)@itc.nl

Commission IV, WG IV/4

KEY WORDS: **Abstraction, Design, Management, Modelling, Services, Spatial Infrastructures, Systems, Requirements**

ABSTRACT:

In the early 90's research on the concepts of spatial data infrastructures focused on ways to improve the capabilities of organizations to supply, share and discover the fundamental data required in their businesses. As a consequence geographic data has become more broadly available, however, continuous change in user requirements driven by factors like new technology, competition, etc., determines that customer satisfaction depends not only on the availability of data, but also in its suitability to the users' particular application domains. Through the integration of data, processes, operations and applications into an infrastructure capable of generating more specialised information services, highly suitable value added data can be obtained that suits specific user communities. Designing such geo-information systems, requires of an appropriate design methodology that supports abstraction, modularity and other mechanisms to capture the essence of the system and help controlling complexity. Numerous concepts to address this idea like system modelling, virtual enterprise, etc., have emerged in the last years. Such concepts properly applied to the geospatial information industry can take us beyond the SDI as just a mechanism for accessing geospatial data to include also demand oriented finished products and services. In this paper we describe an ongoing research project that aims at developing a methodology that enables the modelling of geo-information systems in an integrated perspective. Such methodology will help to develop the SDI as a set of interconnected companies that are integrated (virtual enterprise) and are sharing artefacts (data, value-added products, services) that are located along their distributed nodes and have an economic value. These companies are assembled to operate in a collaborative work perspective, to achieve business goals. The methodology aims at supporting the design process of the system, and also helps enforcing that geo-information systems are reliable, maintainable and compliant with the changing requirements.

1. INTRODUCTION

The goal of durable operationalization of Geo-information Science and Earth Observation will depend increasingly on evolving Spatial Data Infrastructures (SDI) and effective and efficient organizations (for example National Mapping Agencies and National Cadastres) that provide broadly needed fundamental data for a multiplicity of GIS application domains.

In the early 90's research has been focus on the concepts of Spatial Data Infrastructures (SDI) and on processes to improve the performance of organizations that supply the fundamental data in the SDI. Results of such research in SDI can be seen in (Groot R., et all, 2000). Meanwhile new developments in the industrial and service sectors are centring on the concept of Virtual Enterprises (VE). The objectives of the Virtual Enterprise concept are better customer satisfaction, reduced time-to-market and adaptation to changes in the continuous change of requirements. They are called "Virtual" because of their temporary nature, seizing certain, often short-lived business opportunities exploiting the opportunities offered by the ICT. VE is composed of functions provided by the participating organizations (enterprises) and structured and managed in such way that they present themselves to third parties as one enterprise. Applied to the geospatial information industry such concepts take us beyond the SDI as just a mechanism for accessing geospatial data to include also demand oriented finished information services. In turn this extended concept confronts us with the necessity to critically review how the conventional missions of the traditional Surveys will be affected and carried out by organizations that will operate largely beyond their traditional organizational borders. Good examples of the implications of these new concepts are "The National Map" (U.S. Geological Survey, 2001) and the "one-stop-shops" on the Internet to provide integrated public services.

Hence methodologies from manufacturing and service industry were studied to appraise their applicability to the new geoinformation production and dissemination environment. The research activity needs to continue along these lines to develop SDI as a set of interconnected companies that are integrated (virtual enterprise) and are sharing artefacts (data, value-added products, services) that are located along their distributed nodes and have an economic value to provide specialized services, this taking into account reference models and standards that provide a framework for specifying open, flexible and distributed systems.

Such integrated reference models will provide the focal point, around which the business operations in the SDI are designed, implemented, managed and improved and/or new business opportunities are identified. It will also support the assessment of SDI performance in 'totality' along the various operational dimensions, i.e. quality, time of delivery, cost, optimum use of resources, monitoring of changes in the surrounding environment as well as the capability to adapt at both the institutional, the organization's business and operational levels

Our starting points to develop methodologies to model the SDI as an integrated enterprise to achieve well-defined business goal focus on investigating:

- How can SDI (Spatial Data Infrastructure) operate as a set of interconnected companies that are integrated (virtual enterprise) and are sharing artefacts (data, value-added products, services) that are located along their distributed nodes and have an economic value? How can these companies be assembled to operate in a collaborative work perspective, to achieve business goals?
- How can the existing reference models and standards that are developed for open and distributed systems be used to describe and model various interactions (such as goals, processes, information, resources, rules as well as the surrounding changes) in such a SDI?
- How can the SDI integrated enterprise model be validated and used to implement requirements and to assess the SDI performance?

These broad domains will provide the context for the development of the "enterprise model" for a SDI. The result will be a mechanism for capturing knowledge of a SDI system in terms of simple components that can be assembled into large functional specifications. Our research concentrates on the definition, development and validation of the underlying principles that support the concept.

2. GEO-INFORMATION SYSTEMS

The change of emphasis in geo-information production from just data towards diverse products and services forces a change in the design of the system that handles such production. Diverse products and services imply that requirements are dynamic.

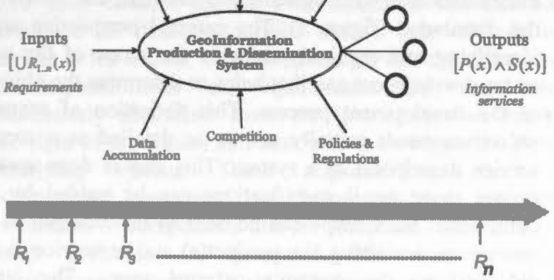


Figure 1. GIP environment & external drivers

To be effective in such conditions, one would prefer to have an adaptable system rather that having to re-engineer it every time. Figure 1 shows the environment surrounding geo-information provision systems. External factors like technology, competition, policies, etc. are the forces that drive the requirements and define the responsibilities (functionality) of the system at any point in time. Continuous change of requirements $R_{\{1..n\}}$ through time affects the system and subsequently impacts its behaviour and responsiveness. Possible outputs in terms of Products *[P]* or Services *[S]* are defined by the inputs *[R]* to the system. Inputs varying over time mean that the system has to adapt continuously in order to respond satisfactorily. We believe that if geo-information systems are to remain accepted as part of large scale business applications, the use of suitable design methodologies is crucial to attack this challenge.

By studying existing geo-information systems, we observe that most of them were design using data oriented design methods and some few using some basic forms of object concepts. They operate with an architecture suitable to acquire and store geographic data that can be retrieved at will. This architecture facilitates the production of certain maps but few diverse products and services. An interesting aspect however, is the existence of a mechanism for providing access to data produced by these systems. This data sharing mechanism is known as the Spatial Data Infrastructure (SDI). A SDI is defined as: the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. The GDI provides a basis for spatial data discovery, evaluation, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general (Douglas N.D., 2001).

We want to take advantage of the existence, understanding and maturity of this concept for data sharing and build upon it a system concept for the generation of a wider range of quality value-added information services as defined by the market. For this purpose we introduce the term GSI (Geo-information Service Infrastructure). A GSI is a system where specialised information products and services can be obtained by exploiting an infrastructure of interconnected data nodes (data repositories), data brokers, service providers, service brokers and clients. The benefits obtained from the GSI are based on a set of artefacts located along the distributed nodes which have an economic value; these artefacts can be assembled to perform operations within the infrastructure, resulting in an specialised artefact that has a value equal or larger than the value of the artefacts used. This behaviour can be regarded as a ``value-added discrete system". A service is created by integrating these artefacts and generating a functionality that satisfy a particular set of requirements. This flexible design based on connectable components is an appealing way of looking at geo-information provision systems but it imposes a great challenge from the design point of view. We dedicate the following sections of this paper to present a design methodology to support the development of sufficiently reliable, maintainable and conformant GSI systems.

3. A DEVELOPMENT METHODOLOGY

A very important consideration in the conceptual modelling of information systems is that models have to describe structural and behavioural aspects of a system (Graham I., 2001). Traditionally, the modelling of both aspects in geo-information systems has been treated separately, and in reality little or none focus has been put to the behavioural part of such systems. Other information based domains have already incorporated into their design methodologies techniques that integrate both aspects. Numerous techniques for this type of conceptual modelling have emerged in the last years like UML (OMG, 1999), IDEF (Mayer R. J., et all, 1995.), etc. Researches at the University of Twente (the Netherlands) have worked for the past several years in the creation of a design methodology (AMBER) to support the development of various types of system models. This method was created to support effective design of distributed systems (Ferreira Pires L., 1994; Quartel D., 1998). We have taken behavioural modelling ideas (AMBER), together with concepts from abstract data types and specifications of systems, to define a methodology devoted to modelling both structural and behavioural aspects of geoinformation systems and more specifically GSI systems.

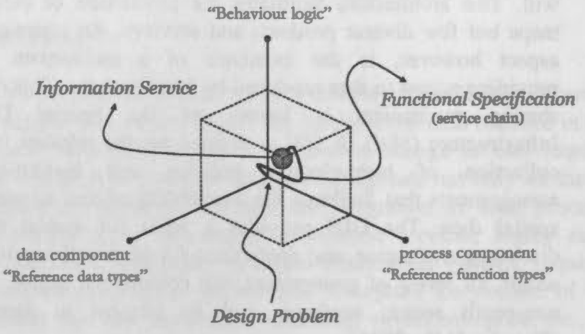


Figure 2. Modelling dimensions

The GSI system's principle of logical distribution provides a general, flexible and evolutionary approach. It offers advantages because it forces one to think in terms of objects, modularity, flexibility, structure and responsibility or, in other words, architecture. However, architecture implies the identification of components, independency, atomicity and, furthermore, interaction, relation and synchronisation between independent components. Such consideration forces one to look at those systems from different perspectives, and at different levels of detail (D'Souza D.F., et all, 1999). We have identified three views to consider for a proper description of the different aspects of a GSI system. Every one of the three axes of the diagram in figure 2 represents an aspect of the system. The data axis represents the information that is used, manipulated and/or generated by the system. The design concepts defined along this axis are used to capture information about *data components* of the system. Data components can be describe in terms of their structure, relationships and associations; the process axis represents the geo-processing capabilities of the system. Design concepts along this axis are used to describe *process components*. Process components are used to generate or modified data components. Process components can be applied to the data components but are independent of data objects themselves; the control axis represents design concepts that allow the definition of *service chains*. This includes the conditions and constraints that define the interactions between data and processes components. The origin of the tri-axes reference system represents the starting point of a processing flow that is activated by a design problem. Design problems are the various inputs (request) to the system that have to be converted into products and or services. The path taken through the cube is a functional specification of a service chain that satisfies a design problem by making use of a proper combination of artefacts. The execution of a predefined service chain generates a desired product or service.

In section 2 we explained the basic GSI concept. From that definition we identified the composing artefacts of a GSI system and they can be categorised as follows:

- Data, stored as datasets with proper documentation (metadata),
- Operations to perform transformations on datasets,
- Processes to generate non existing datasets,

These groups of artefacts constitute the foundation of the system (Figure 5). Existing datasets stored in data nodes, and processes and operations provided by the service nodes, have to be described properly within the corresponding view before they can be integrated into a service chain. One additional feature to consider here is the potential evolution in terms of the number of elements that the system may include. To address the above mentioned characteristics our structuring approach focuses in three perspectives: first the *external perspective* captures the responsibility of the system that is defined by the relations with the surrounding environment; next the system's constituent or forming elements (components) are described in the *internal perspective*, and then the possible interactions between the forming elements to achieve the various functionality are modelled in the *distributed perspective.*

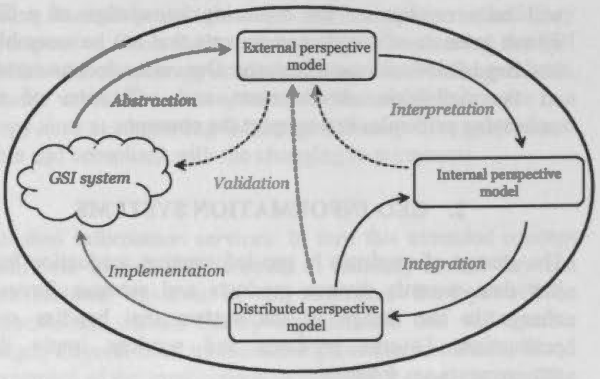


Figure 3. Modelling steps

We have defined three different levels of abstraction for the manipulation of GSI systems the external, the internal and the distributed (figure 3). The external perspective aims at identifying and explicitly delimits the scope of the system under development and that helps to determine the objectives of the development process. This definition of scope and objectives needs initially not be as detailed as a complete service description of a system. This step is done such that further more detail specifications can be guided by these definitions. Such scope can be defined for example as short statements describing the product(s) and/or service(s) to be delivered to the system's external users. The internal perspective focuses on the formal definition of the elements of the GSI system that fall within the selected scope and can satisfy the determined objectives. The output of this face is a set of system components of two types: the data-objects and process-objects; data-objects are modelled using data types which characterise their purpose; process-objects are modelled as complex actions depicting the data that they generate or require and the service that they perform. The distributed perspective aims at describing the internal system

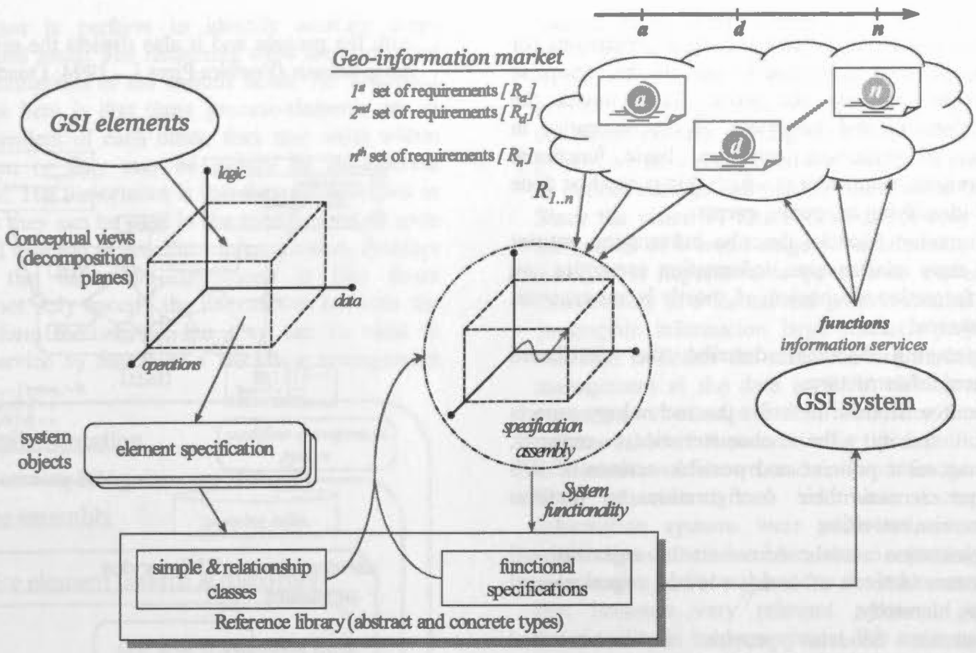

a
d
n
Geo-information market
1st set of requirements [R_a]
2nd set of requirements [R_d]
nth set of requirements [R_n]
GSI elements
$R_{1..n}$
logic
Conceptual views (decomposition planes)
data
operations
functions
information services
GSI system
specification assembly
System functionality
system objects
element specification
simple & relationship classes
functional specifications
Reference library (abstract and concrete types)

Figure 4. Development Methodology

structure in terms of composition of various components that interact to perform the system functions. This is regarded as system behaviour and shows how groups of interacting components work together to achieve an observable functionality of the system. Many different compositions are allowed to be define that may implement the same functionality. Or even functions that look the same but the quality of the generated services may vary.

To acquire the necessary information for the generation of the various system models it is crucial to follow a structuring method. Figure 4 depicts a structuring notion for GSI system; the top-left part of the figure shows how the system has to be analysed as a collection of elements that are in principle independent form each other. Using categorisation, those elements are arrange according to their correspondence to the three system views (figure 2). Every contributing element has to be described to enable its participation in different system functionality. For that purpose a formal definition of the elements is produced and stored in a reference library. At this point each element has its own definition and can now participate in different service chains that will deliver services to the clients. To define these service chains we used the responsibilities or requirements that are being continuously identified from the geo-information market (top-right of the figure) and that have to be supported by the underlying system. These responsibilities (sets of requirements) are used as design problems. Suitable interactions of components that are compliant to the different design problems are generated. This step is called specification assembly. During this step a path through the cube is defined it represents a behaviour definition (specification assembly). Every behaviour definition will be a sequential arrangement represented by the relations between the participating elements, including conditions and alternatives. The logic that drives these relations is based on pre and post conditions that are evaluated to identify the correct sequences. These functional specifications suitable to deliver particular services are stored in the reference library. Functional specifications can be reused in combination with additional elements to define specifications to define more complex functions. All this stored specifications can be enacted with a workflow management system or a similar tool to physically generate a desired service.

4. GSI ARCHITECTURE

The GSI is defined as a set of interconnected companies (GSI Nodes, Figure 5) that are integrated and sharing artifacts (business goals, strategies, data, processes, value-added products, systems, etc.) that are located along their distributed nodes and that have an economic value, to provide specialised services, all this within a framework of reference models and standards that provide the mechanism for specifying open, flexible and compliant systems.

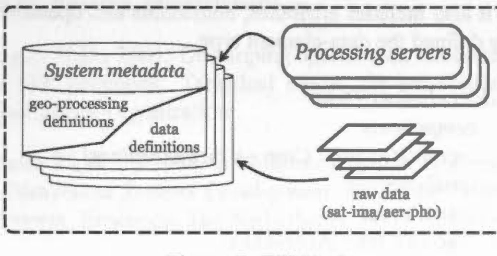

System metadata
geo-processing definitions
data definitions
Processing servers
raw data
(sat-ima/aer-pho)

Figure 5. GSI Node

The enterprise model for GSI consists of various models (each can be broken into more detailed sub-models), such as:

- Goals/Business Vision: describes the purpose of the business, what to achieve, the goal structure, and illustrates problems that must be solved in order to reach those goals;
- Production Model: represents the geometric and non-geometric features as well as design details of products, their logistics and life cycle;

- Process models: indicate the set of operations (or actions) to be performed to execute the enterprise activities and do the work. They model the functionality and behaviour of the enterprise in terms of processes, activities, basic functional operations, sequences in which things must be done and identifying triggering events;
- Information Models: describe information entities and their relationships, information semantics and the federation mechanism of, mostly heterogeneous, databases;
- Engineering Models: describe the distributed system infrastructure;
- Resource Models: describe the technology aspects and specify the characteristics, capacity, management policies and possible actions of H/S resources and their configuration to perform enterprise activities;
- Organization models: document the organizational structure in terms of decision levels, organizational units, hierarchy;
- Economic Models: provide a cost-oriented analytical view of the enterprise to evaluate the cost-effectiveness of the various parts of the enterprise, the pricing structure, etc;
- Optimization and Decision Making Models: deal with issues from operations research and control theory and used by the DSS;
- Rules can be functional, behavioural and structural, constraining some aspects of the business such as processes and resources. They also include the rules enforced by law and regulations.

Figure 5 shows the high level architecture of a GSI Node. A GSI node is a participating entity in the infrastructure. It provides either raw data or geo-processing functions. These two types of elements are described as data-elements or process-elements. List 1 shows the specification of a field data-element. It represents a data type that characterizes parcel data, and it is available at a GSI Node and can be incorporated into a service specification. The description contains other data types used by this data-element like crop, and it also includes attributes, constraints and operations that fully defined the data-element type.

```
data element Field
  components
    crop(ID : int) : Crop → theme(regions)
  attributes
    area : real
    access_tax : |AccessTax|
    rainfall_categ : int
    temp_categ : int
    num_cr : int
    location : geobj(regions)
  operations
    addcrop (ID : int)
  constraints
    crop(id).area +...+ crop(id+num_cr).area ≤ area
end;
```

List 1. Field data-element

Fig * Shows the definition of a process-element. In this case it is called generatingAOI. This description is known as a behaviour block, it presents the action executed to carry out with the process and it also depicts the relationships among those actions (Ferreira Pires L., 1994; Quartel D., 1998).

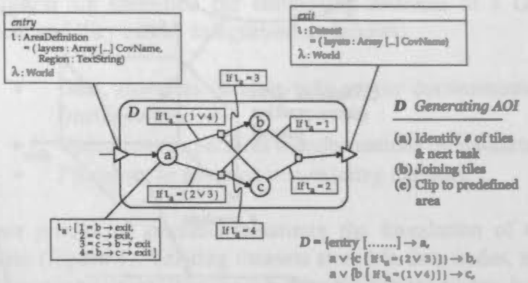


Figure 6. generatingAOI process-element

```
service element TIC-service
  attributes
    …
  data elements
    taxform : (ID : int) : set(TaxForm)
    …
  process elements
    generatingAOI (…) {A}
    cropClassification (…) {B}
    taxfactorsdefinition (…) {C}
    taxcalculation (…) {D}
  behaviour
    {√ → entry
    entry […] → A[if $\iota_{entry}$.Location ∩ Enschede ≠ null],
    $A \vee \{ C [ \text{if} \mid \tau_B - \tau_C \mid > 5 ] \} \rightarrow B$ (λ : all provinces )
      [ $\lambda_B$ = ιentry .Location ],
    $( B \wedge \neg D ) \rightarrow C$ ( λ : all provinces ) [ $\lambda_C = \lambda_B$],
    $( B \wedge \neg C ) \vee ( B \wedge C [ \text{if} \mid \tau_B - \tau_C \mid \leq 5 ) \rightarrow D$
    D → exit […] }
end;
```

List 2. TIC-service definition

A service definition is achieved by combining elements definitions. List 2 shows such a specification. In this case the specification of the TIC-service (Tax identification and calculation service) consists of four parts: attributes, data-elements, process-elements and behaviour. For the first part of the example we do not consider any user defined attributes but all of those required to qualify the service are to be included in this part of the specification. The following part contains all the data elements from which instances of objects are required for use within this service definition, in this example we defined a set of objects of the type TaxForm. In the same way the process element part includes the necessary activities that implement the service. These activities contain internal actions that were defined within every individual process element in their correspondent behaviour block. To simplify the behaviour part of the example we have renamed the activities as letters, *generationAOI* is $\{A\}$, *cropClassification* is $\{B\}$, etc. The last part of the TIC-service specification defines all the possible interactions between the participating elements as a behaviour definition. This definition starts with the process-element *generatingAOI* that determines the areas assigned to a *TaxForm* that require tax calculation; next, if the stated constraints are satisfied, the process-element *cropClassification* is perform to identify existing crops within the defined areas; The remaining steps are concerned with the tax calculations of the various fields. An important aspect to notice here is that these process-elements are in principle independent of each other, they may exist within one organization or they may be supply by an external service provider. The importance is that they are specified in a way such that they can be used in the specification in spite of their physical location or specific implementation. Another advantage of the behaviour definitions is that those definitions do not only specify the interactions between the elements that form the service, but they can be used to automate the service by means of a workflow management tool.

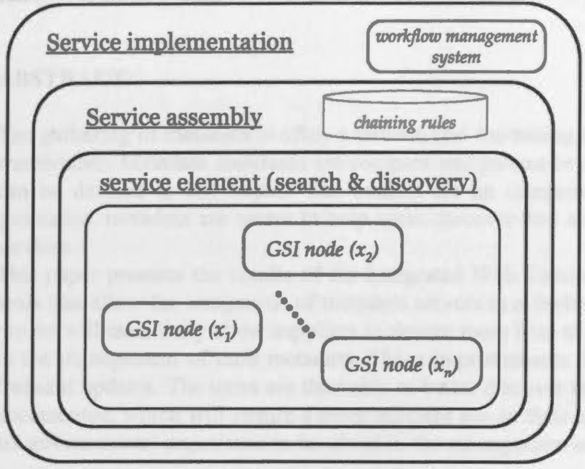


Figure 7. GSI architecture

Figure 7 presents the architecture upon which the GSI lies to provide the desired level of services. The inner layer consists of all the interconnected nodes containing functionality (functions or data) to be share among all participants. These nodes contain metadata descriptions at the system level (List 1, figure 6). The following layer provides functionality for search and discovery of elements to generate service definitions. This search is performed to identify elements that satisfy a required processing step in a service chain definition. The third layer (service assembly) provides functionality to logically connect the various elements into a coherent specification (List 2) that is compliant with an information service request. The outer layer provides functionality to execute process definitions based on the design made in the previous layer.

5. CONCLUSIONS

In this paper we presented the current status of an on-going research that is being carried out in cooperation between the University of Twente, the Netherlands, specifically the Architecture of Distributed Systems group, and the International Institute for Geo-Information Science and Earth Observation (ITC), the Netherlands. Our aim in this research is to define a methodology for the generation of high level generic specifications of geo-information systems components that facilitate their integration for the generation of services that are more compliant with the ever changing requirements of the geo-information users.

The GSI is defined as a set of interconnected companies that are integrated and sharing artifacts (business goals, strategies, data, processes, value-added products, systems, etc.) that are located along their distributed nodes and that have an economic value, to provide specialised services, all this within a framework of reference models and standards that provide the mechanism for specifying open, flexible and compliant systems. During the development of this research project we have recognised the need for a proper mechanism to define, manage and publish metadata at the system level. Since the vision of future systems not only the GSI, is the integration of a broad range of functionality into services to users, it is imperative to be able to express services and functionality in a formal and generic manner. In the field of geographic information large research projects have been executed to define the standards and methods for metadata management at the data level. In this document we have presented our attempt to describe system functionality at the metamodel level. A procedure for the definition and assembly of elements was shown, and concepts for the specification of both behavioural and informational aspects of geo-information systems were addressed, issues that in our application domain has been treated separately and more over no attention had been placed to the behavioural aspect, aspect that becomes very relevant when it comes to share functionality in an open system. We are sure that in the next stages of this work much more detail and valuable information will be develop with respect to these subjects.

Tool support is a crucial element for the success of large scale development projects like the GSI. We do not yet have sufficient automated tools to help in the validation of models, but what is more important even are the tools the manipulation, change and maintenance of the specification models. This is a very relevant issue, because once a particular service functionality specification that is shared in a repository is changed, the effect should be reflected to the services build on top of that specification, and without proper management of this aspect, inconsistencies will rapidly develop and the whole architectural concept will become uncontrollable.

REFERENCES:

D'Souza D.F., and Wills A. C., 1999. *Objects, Components and Frameworks with UML*: The Catalysis Approach. TheAddison-Wesley object technology series. Addison-Wesley, Reading, Massachusetts.

Douglas N.D., 2001. Developing spatial data infrastructures: The SDI cookbook. Technical report, Global Spatial Data Infrastructure Organization.

Ferreira Pires L., 1994. *Architectural Notes: a Framework for Distributed Systems Development*. PhD thesis, University of Twente, Enschede, The Netherlands, Ph.D. Thesis no. 94-01.

Graham I., 2001. *Object-Oriented Methods: Principles and Practice*. The Addison-Wesley object technology series. Addison-Wesley, Reading, Massachusetts, 3rd edition.

Groot R., McLaughlin J., 2000. *Geospatial data infrastructure: Concepts, cases and good practice*. Oxford University Press, New York.

Mayer R.J., Menzel C. P., 1995. Information integration for concurrent engineering (IICE): IDEF3 process description capture method report. Technical report, Knowledge Based Systems Incorporated.

OMG, 1999. Unified modeling language specification. Technical Report Version 1.3, Object Management Group.

Quartel D., 1998. *Action Relations: Basic design concepts for behaviour modelling and refinement.* PhD thesis, University of Twente, Enschede, The Netherlands, Ph.D. Thesis no. 98-18.

U.S. Geological Survey, 2001. The National Map: Topographic mapping for the 21st century. Report, Office of the Associate Director of Geography, Reston, VA, USA.

INTEGRATED WEB CATALOGUE SERVICE (IWCATS)

Serge Kéna-Cohen, Edric Keighan, Guy Rochon

Intelec Geomatics, 696, Ste-Croix, Office 200, Montreal Canada, H4L 3Y2
kena@intelec.ca

Working Group IV/4

KEY WORDS: Metadata, Spatial Warehouse, OGC

ABSTRACT:

The gathering of metadata is often a tedious and frustrating exercise that has to be performed on top of data management in spatial warehouses. Metadata standards are complex and cannot be readily applied to data managed in spatial warehouses where a data set can be defined as any object. The context for an Integrated Web Catalogue service (IWCat) is based on the postulate that in geomatics, metadata are meant to help users discover and assess geospatial information and to help providers publish their data or services.

This paper presents the results of the Integrated Web Catalogue Service (IWCatS) project. The project objectives were to develop tools that allow the integration of metadata servers in a Web environment using protocols similar to OpenGIS standards. The project results will essentially allow suppliers to devote more time to the improvement of their data thanks to the economies of time obtained in the management of their metadata. These improvements could take the form of gathering of new data, better precision or more frequent updates. The users are then able to better discover the data of interest, those will be of better quality and they will be better documented, which will ensure a more efficient use in their specific fields (forestry, mines, etc.). The IWCat service was tested with six governmental organisations involved in the management of geospatial data.

1. INTRODUCTION

Stakeholders involved in distributing and accessing geospatial data agree that the first stage in achieving this goal is the documentation of data sets or the entry of metadata. Organisations handling standards (FGDC, ISO/TC211, OpenGIS) have identified metadata as a priority in their work and geospatial data infrastructures such as the Canadian Geospatial Data Infrastructure (CGDI) often start with the implementation of services related to metadata (discovery, evaluation, access).

However, the gathering of metadata is often a tedious and frustrating exercise that has to be performed on top of data management in spatial warehouses. Metadata standards are complex and cannot be readily applied to data managed in spatial warehouses where a data set can be defined as any object.

Intelec Geomatics has developed, under a GeoInnovations project a multistandard, multilingual metadata cataloguing (M^3Cat) tool (see www.intelec.ca/technologies). CubeWerx and SoftMap provide warehouse products to manage geospatial data. The three companies had an interest in improving metadata and data management in their products to facilitate their clients presence in geospatial data infrastructures.

This paper presents the results of the Integrated Web Catalogue Service (IWCatS) project realised under the GeoInnovations program.

2. CONTEXT

The project objectives were to develop tools that allow the integration of metadata servers in a Web environment using protocols similar to OpenGIS standards.

Figure 1 presents the context for an Integrated Web Catalogue service (IWCat). The context is based on the postulate that in geomatics, metadata are meant to help users discover and assess geospatial information and to help providers publish their data or services.

The model is composed of two objects and three actors[1]. The objects are metadata catalogues accessible through a Catalogue Server service and Spatial Warehouses accessible through a Feature Server service, while the actors are providers, librarians and consumers who have access to diverse services over the Web using standard protocols.

The consumer actor supports graphics and text based interactions by humans and automated agents. Consumers are users of both information found in metadata catalogues and spatial warehouses. Consumers browse and query catalogues in order to access or reference information of interest in spatial warehouses and browse metadata catalogues for more detail metadata information.

The metadata catalogue object includes information (who, what, where, etc.) and search engines to let users identify data of interest. Catalogues describe and reference content found in spatial warehouses, other catalogues or other data collections.

---

[1] In some organisations the Librarian and the Provider may be the same person.

The spatial warehouse object contain data holdings comprised of feature collections, features and feature metadata. Spatial warehouses are offered by data providers.

The librarian actor manages catalogues. Librarians ingest, organize and extract information from spatial warehouses and create and augment catalogues.

The provider actor creates, updates and publishes spatial warehouses.

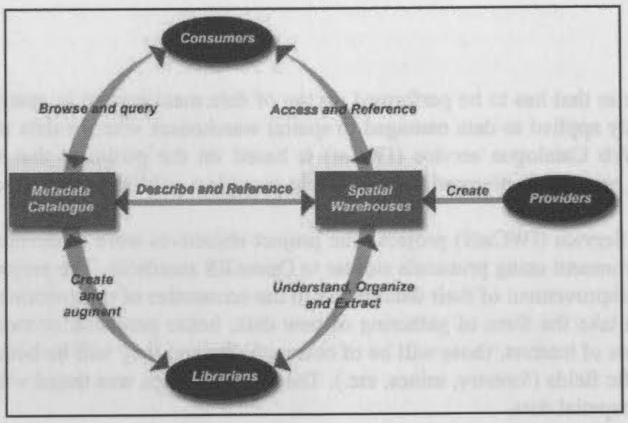


Figure 1. Context for the IWCat service

3. ISSUES AND ORIENTATIONS

To reach the project objectives, the project team had to resolve a number of issues, in particular:

- The level of integration between the metadata and spatial warehouses tools;
- The definition of granularity levels (data set versus object);
- The identification of common identifiers;
- The aggregation of metadata;
- The management of historical aspects;
- The development of protocols.

3.1 Level of Integration

There are two ways to integrate the management of metadata found in catalogues and spatial warehouses:

- to enable each service to interface with each other using open protocols;
- to integrate the management of metadata for discovery services in data warehouses tools.

Both approaches translate into a workable and operational environment. In theory, the second alternative may be preferable since the only reason that the two tools are not integrated is that for historical and performance reasons they have been developed as specific software.

However, the integration of the two (2) tools will not allow discovery services to work with data set collections not in spatial warehouses, e.g. files, paper products. And the integration of both tools will not correspond to the real world where each type of tool exists.

Hence, it was recommended to integrate the management of catalogue and spatial warehouse data by considering the two services as separate, but with the capability to access one service from another in a B2B fashion. Each service has the capability to be invoked by users of the other service. This will lead to global interoperability between the two services beyond proprietary services available from specific vendors.

This recommendation raised issues of security and availability (what happens when one of the services is down, how do we handle access privileges in both services) but these issues are common to any services.

3.2 Granularity

Metadata can describe geospatial data at different levels such as, geometric primitives, features, layers, maps, collections, data base, etc.

Some of these metadata fields are specific to each data set level while others are produced by performing operations on the lower levels metadata (geographic extent, content, dates, etc.).

In order to maintain an integrated environment for Web Catalogue Server (WCatS) services and Web Feature Server (WFS) services each service must define their respective metadata using a common encoding and format based on open standards such as FGDC or ISO 19115. Both the catalogue and spatial warehouse services must maintain a metadata model where the schema and the relations can be discovered over the Web using simple Web Catalogue Server interfaces. However with current metadata implementations (catalogue servers and spatial warehouses) the levels of granularity are often not identified uniquely in both services and the metadata elements they commonly manage do not have the same semantics.

Hence it was recommended that additional functionality be provided from an Integrated Web Catalogue Server environment perspective for maintaining and harvesting a metadata catalogue or data holding based on mapping different metadata models. The data holding is user configurable and contains for each level of granularity the metadata mapping between the catalogue and spatial warehouse environments :

- **The correspondence between the levels of granularity (e.g. level data set in catalogue correspond to level layer in spatial warehouse or feature level in spatial warehouse correspond to sub-data set in catalogue) ;**
- **The correspondence between metadata elements for each metadata standard maintained by the catalogue and spatial warehouse environments (e.g. ArchiveFormats in spatial warehouse corresponds to 6.4.2.2.2.3 Recording Format in a FGDC catalogue);**
- **How do these metadata elements react to one another (direct correlation, format conversion, etc.).**

The detailed structure of this data holding is presented in the data model (see Section 4).

3.3 Common Identifiers

For metadata management to be synchronized between the two environments, a method must be developed to uniquely identify and correlate, at each level of granularity, geospatial data objects for which metadata management must be integrated.

Ideally, the Spatial Warehouse unique identifier should be used in the Catalogue and the relations between the levels of granularity (primitive within a feature, feature within a layer, etc.) maintained by the warehouse.

However, relations between objects in each service are not always one to one:

- A catalogue can describe data in more than one spatial warehouse;
- Multiple catalogue entries could describe one spatial warehouse object (a feature or a pixel with different sources).

It was recommended that warehouse identifiers be used in the Catalogue. To take into account exceptions, a N to N table was also made available in the IWCat environment so that relations between each data occurrence is maintained when required. In all cases the relations between the geospatial data levels of granularity are maintained in the Spatial Warehouse.

3.4 Metadata Aggregation

Two different issues were considered:

- aggregation of metadata
- and metadata about aggregation.

An example of the first case is when one has the same metadata that are used at different levels of granularity, the metadata of the coarser level being derived from the metadata at the finer level (ex. date of last update for a data set being derived from the dates of the updates of object occurrences). When this is of interest to the user, one must check the integrity between the different levels of granularity.

It was recommended that procedures to do this automatically be implemented into the tools involved in the project as well as the corresponding updating rules.

An example of the second case is when one is interested to describe how global data have been derived from detailed data (ex. in multidimensional datacubes of warehouses, the total area of industrial land uses of a county is derived from the sum of the areas of the land uses of individual properties), presently no standard exist in geomatics to describe the aggregation algorithms and **this was not included in the scope of the project. However, it was recommended that the correspondence between metadata in spatial warehouses and in catalogues be considered a one to many relationship.**

3.5 Historical Aspects

We had to consider, if desired, a way to keep a trace of the evolution of the metadata (ex. for legal reasons). This can be done using the different versioning approaches familiar to spatio-temporal database design, however one must keep the synchronization between the different states of a data value and its metadata in the warehouse or the catalogue.

Services to define the requirements for the updating were not developed in the project (ex. frequencies or thresholds in each direction, format of the update metadata, synchronization rules).

3.6 Protocol

It is critical that geospatial data and metadata be maintained current and in synchronization and that the protocol used be flexible enough to support not only the access to metadata but perform all required transactions for the maintenance of metadata between the Catalogue and Spatial Warehouse services.

The on-going draft OGC Web Feature Server specification is both an encoding/ transport mechanism and a transaction mechanism for simple geographic features in XML/GML. The WFS interfaces supports INSERT, UPDATE, DELETE, QUERY and DISCOVERY operations on OGC Simple Feature datastores using HTTP as the distributed computing platform. Although the focus of the OGC WFS specification was a geometric feature and its attribution information there are no restrictions for the applicability of this protocol, or similar protocol, for metadata.

It was therefore recommended that the OGC Web Feature Server (WFS) interface specification be adapted for its applicability for the management of metadata between the Spatial Warehouse and the Catalogue services.

4. DATA

The data model for the Integrated Web Catalogue (IWCat) data store is presented on Figure 2. This data store has to be maintained in both the catalogue and spatial warehouse environments.

The project used XML for encoding metadata information and GML Profile 2 for the encoding of geographic information in XML.

5. OUTCOMES

The project outcomes will generate benefits for suppliers and users within spatial data infrastructures. Suppliers will have the capability to:

- Update their metadata catalogues from their data warehouses.
- Propagate metadata updates from a data set to all the objects it contains.
- Improve the connection of their metadata to the CGDI.

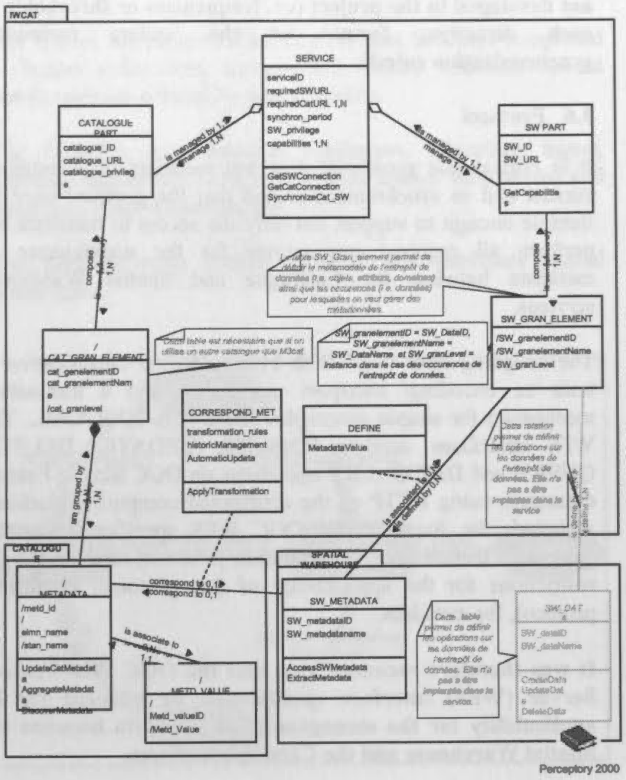

IWCAT
SERVICE
serviceID
requiredSWURL
requiredCatURL 1,N
synchron_period
SW_privilege
capabilities 1,N
GetSWConnection
GetCatConnection
SynchronizeCat_SW
CATALOGUE PART
catalogue_ID
catalogue_URL
catalogue_privilege
SW PART
SW_ID
SW_URL
GetCapabilities
is managed by 1,1
manage 1,N
is managed by 1,1
manage 1,1
compose
1,N
La table SW_Gran_element permet de définir le metamodele de l'entrepôt de données (i.e. objets, attributs, domaines) ainsi que les occurences (i.e. données) pour lesquelles on veut gérer des métadonnées.
SW_GRAN_ELEMENT
/SW_granelementID
/SW_granelementName
SW_granLevel
/ CAT_GRAN_ELEMENT
cat_genelementID
cat_granelementName
/cat_granlevel
Cette table est nécessaire que si on utilise un autre catalogue que M3cat.
SW_granelementID = SW_DataID, SW_granelementName = SW_DataName et SW_granLevel = instance dans le cas des occurences de l'entrepôt de données.
CORRESPOND_MET
transformation_rules
historicManagement
AutomaticUpdate
ApplyTransformation
DEFINE
MetadataValue
Cette relation permet de définir les opérations sur les données de l'entrepôt de données. Elle n'a pas a être implantée dans le service
are grouped by
is associated to 0,N
is defined by 0,N
CATALOGUE
/ METADATA
/metd_id
elmn_name
/stan_name
UpdateCatMetadata
AggregateMetadata
DiscoverMetadata
correspond to 0,1
correspond to 0,1
is associate to
1,1
/ METD_VALUE
Metd_valueID
/Metd_Value
SPATIAL WAREHOUSE
SW_METADATA
SW_metadataID
SW_metadataname
AccessSWMetadata
ExtractMetadata
Cette table permet de définir les opérations sur les données de l'entrepôt de données. Elle n'a pas a être implantée dans le service.
SW_DATA
SW_dataID
SW_dataName
CreateData
UpdateData
DeleteData
Perceptory 2000

Figure 2: Data Model

Users will have the capability to:

- Query additional information on the data from data warehouses.
- Identify, when querying spatial data infrastructures, available data sets as well as specific data elements (concept of evolution from a data set to the information).

The project results will essentially allow suppliers to devote more time to the improvement of their data thanks to the economies of time obtained in the management of their metadata. These improvements could take the form of gathering of new data, better precision or more frequent updates.

The users are then able to better discover the data of interest, those will be of better quality and they will be better documented, which will ensure a more efficient use in their specific fields (forestry, mines, etc.).

6. CONCLUSION

The IWCat service was developed by Intelec Geomatics (Intelec) and its partners CuberWerx, SoftMap Technologies and the Centre for Research in Geomatics (CRG) of Laval University. It has been tested with six governmental organisations involved in the management of geospatial data: the Department of National Defence, Environment Canada, the Canadian Centre for Remote Sensing, the Canadian Forest Service, the Earth Sciences Sector of Natural Resources Canada and the Quebec Ministry of Natural Resources.

The project was a good example of partnership between the Federal Government, Provincial Governments, the Private Sector and Universities.

INCREMENTAL DATA EVOLUTION - COMPLEMENTARY APPROACH TO SCHEMA EVOLUTION OF ORGANISATION'S LARGE OPERATIONAL GIS DATABASES

L. Salo-Merta

Institute of Business Information Management, Tampere University of Technology, P.O.Box 541, FIN-33101 Tampere, Finland – leena.salo-merta@tut.fi

Commission IV, WG IV/4

KEY WORDS: Application, Database, Design, Development, GIS, Management, Software, Spatial

ABSTRACT:

The paper addresses database development and schema evolution from the geographic information provider organisation's and system integrator's point of view. The applicability of incremental approach to database schema evolution is considered, and incremental data evolution (IDE) is introduced and proposed as a complementary method for schema evolution and data restructuring of large geospatial datasets in operational use. The hypothesis is that IDE combined with full data conversion strategy would provide a flexible approach in the schema evolution of large operational databases. The applicability of the IDE approach is considered with a real-world case.

1. INTRODUCTION

1.1 Development pressures in GI provider's evolving digital environment

According to European Union, geographic information (GI) is by far the largest type of public sector information in Europe, comprising 38,5 per cent of all public sector information (European Union, 2001). The computer penetration into the field of GI follows a general pattern: First it is adopted in the GI provider organisation for speeding up the production process. At this stage there are no significant changes in the end products. In the second phase the potential of computerized environment is exploited more widely and the market for GI in digital form starts to develop. In the third phase completely new products and activities arise from the digital culture.

The pioneer organisations that started to digitise GI early, have faced all the stages one-by-one. However, the stages are concurrent and ongoing, and are shown as layers of activities in Fig. 1. In the time being the 'geo–e-business' will be absorbed into the 'normal' service production layer, but at the moment it is feasible to examine it as a layer of its own. The project portfolios of organisations tend to grow into new topics, and the most appealing questions from the manager's or technologically oriented researcher's point-of-view may lie in the expansion of technology and new applications, rather than re-engineering the existing parts.

However, databases (DB) and data management are the true cornerstones of the whole and they take effect on each activity layer, directly or at least indirectly. Commercial solutions for geographic data management (GDM) have developed rapidly during the past few years, showing the evolution of architecture from rather closed monolithic systems towards client-server architectures with open, standardized interfaces. GI provider organisations have made great investments to establish their existing large datasets in digital form. Their concern for their asset is three-fold: Firstly, how to maximize the exploitation of the digital content. This is the driving factor for expanding into new dissemination technologies, like the WWW and mobile, and new types of products and services. These actions are very apparent and promoted. Secondly, the methods and techniques in the data production layer need improving for more economic and efficient data updating. To address these questions, incremental updating of databases has recently become an issue of research and development. Last but not least, user organisations are concerned as to how to keep up with the evolving data management technology and ensure the sustainable platforms for their information asset. Because GI datasets are typically *data-intense* both in size and complex structure, the solutions and new arrangements in data management have significant impacts on all actions. For user organisations of GIS, database re-design and schema evolution are challenging tasks in the development of their next-generation systems. They are more 'back-end' –processes, partly invisible and less promoted, but nevertheless significant and anything but trivial.

1.2 Databases in next-generation system projects

Next-generation GIS projects are current in many organisations. By next-generation projects we mean substantial development projects in an environment that already comprises a previous-generation GIS and digital datasets. Next-generation GIS projects usually include a database project, where data is translated from the existing data storage into a new database management system (DBMS). DBMS is typically a Commercial-Off-The-Shelf (COTS) component in an integrated system. A DBMS with spatial capabilities has been strongly associated to certain vendor specific client solutions and product families, but the situation is changing rapidly. Currently one has alternative clients, components and development tools to choose from. It is also possible to implement the 'business intelligence', or should we say here: the 'geospatial intelligence' in different layers of the software architecture: the database server, the middleware or the client.

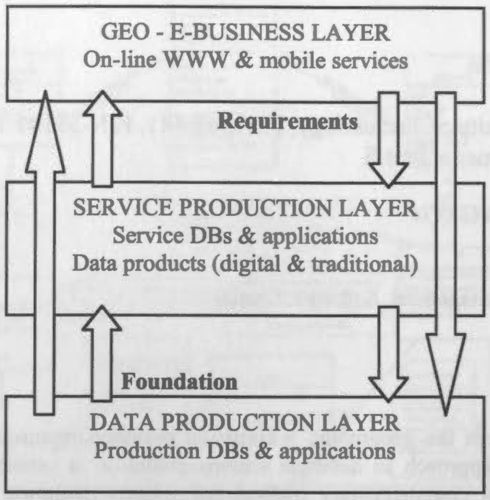

GEO - E-BUSINESS LAYER
On-line WWW & mobile services
Requirements
SERVICE PRODUCTION LAYER
Service DBs & applications
Data products (digital & traditional)
Foundation
DATA PRODUCTION LAYER
Production DBs & applications

Figure 1. Activity layers of a GI provider organisation

In a next-generation GIS project, the user organisation needs a practical solution for database management, and it needs it now. Although the latest technology always seems to be three months ahead, one just has to try to make the best out of what is available here and now (Kucera et al., 1999). According to our experience a practical solution can be defined with three keywords: Economy, Maintainability and Technical requirements. *Economy* stands for both investment cost and operational costs. *Maintainability* stands for the wish to choose a provider and platform with a good probability to survive and be competitive and technologically advanced in the future. *Technical requirements* include general and specific database capabilities, and issues concerning architecture, interfaces, scalability, extendibility and programming facilities. All of these need to be examined in the context of the user organisation, to determine the 'customer need'. As the DBMS is a central component of the system, major development actions invoke needs to develop the other subsystems as well. In the most dramatic case it means the replacement of existing clients with new ones.

1.3 Content of this study

The focus of this paper is on primary schema evolution, as it is seen from the system integrator's point of view. By primary schema evolution we mean structural changes to user's data tables. The fundamental questions are: Is primary schema evolution possible with existing large geospatial datasets? Could it be carried out without significant disturbance for production? Can we allow pluralism in schema definitions and storage structures as a permanent, or at least long-lasting state, and how do we support pluralism?

First we introduce taxonomy of DB development activities and consider the factors that influence these activities in organisations. Then we consider the applicability of incremental approach to DB schema evolution. A method called *incremental data evolution* (IDE) is introduced. The hypothesis is that IDE combined with full data conversion strategy would be a flexible approach in the schema evolution of large production databases.

The IDE approach is considered in a case study, where the possibilities of interrupting the production use of the DB for data conversions or any other reason are limited. The principle of the needed DB interface is introduced.

In conclusions the advantages and disadvantages of IDE approach are considered. Finally the question is raised whether we actually need primary schema evolution and for what reason? And if we don't, for how long are we able to survive?

2. TAXONOMY OF DATABASE DEVELOPMENT ACTIVITIES

Based on experience, we find it feasible to study the development of a system's DBMS environment as three basic types of activities: *Database product change, Version migration* and *Schema evolution* (Fig. 2).

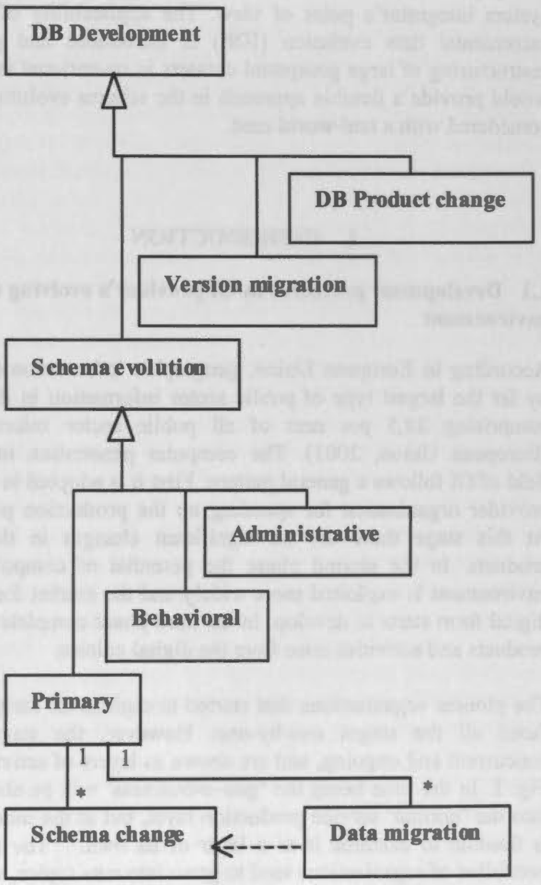


Figure 2. Taxonomy of database development activities

The database product change projects include:

- Design and implementation of the new DB schema
- Preparation of data translation
- Data translation
- Data validation
- Further data re-structuring (if included at this stage)

It is noteworthy that because of the history of proprietary DBMS solutions for geospatial data, considerable effort may be required for other software activities in database product change projects. It may be necessary to change the client applications as well or to re-write the interfaces to the existing ones. For this reason the activities concerning the database itself are probably only a part of a large software project, and the pressure to minimize the size of the project by altering the data as little as possible at this stage may be substantial.

Version migration means upgrading the system to a higher version of the same DBMS product. The changes are product driven: It may include internal changes in data storage, indexing, partitioning etc. On the logical level there may be changes in system tables, e.g., metadata table structures. With new DBMS functionality and capabilities version migration may inspire the customer to further development, but as is, it includes no customer-driven changes in the logical structure of the database. DBMS vendors usually provide migration tools with their products to facilitate the shift into the latest technology and guarantee the evolution in this sense.

Schema evolution means altering the structure of the database, here especially the customer-driven changes. Basically it means changes in table definitions, but as other types of objects (e.g., views, triggers, DB functions and procedures) are stored in the DB as well, it needs to be concerned in a wider sense. However, the consequences of committing a behavioural change (e.g., inserting a new function into a package of DB functions) are quite different from the attempts to commit structural changes in stored data. Purely administrative tasks dealing with physical storage and performance (e.g. table space definitions and indexes) can be considered as another group of tasks. Therefore three subtypes of schema evolution activities are identified: *Administrative changes*, *Behavioural changes* and *Primary changes*. Primary changes concern structural changes to user's data tables. If the data tables are populated, primary changes mean two things: *Schema change* and *Data migration*. We use primary database evolution in a broad sense to cover all actions that are needed to change the schema and propagate the changes to the data, regardless of the existence and capabilities of DBMS's schema evolution facilities. If the DBMS's schema evolution facilities do not support the actions needed, they have to be carried out by other means, e.g. by creating new or temporary tables and implementing ad-hoc migration modules.

3. FACTORS OF DATABASE DEVELOPMENT ACTIVITIES

The factors that influence database development activities are discussed below. They are: Data conversion cost, Evolution strategy, Developing DBs in operational use and Pluralism in DBs.

3.1 Data conversion cost

Converting the existing data into the new system as quickly and painlessly as possible may be a key question in an organisation's DBMS implementation project. Therefore, commitment to the old design may become a critical factor in the design of the application data model. (Salo-Merta et al., 2001). For a very large database with objects with complex structure, which geographic objects appear to be, each extra step included in the conversion process may have a dramatic effect in the throughput time, if it causes a need for operator intervention. Data quality may become the bottleneck in the conversion process, if the target system's validation rules are tighter than the source system's. Relatively small numbers of exceptions that need to be handled manually may turn into vast number of working hours, if we consider millions of objects.

3.2 Step-by-step evolution

All the user requirements can seldom be satisfied at once, and they keep on changing. Yet the data acquisition for digital geo-spatial datasets is expensive and time consuming, there is a strong motivation to preserve the existing data and pass it on to the next system generation. We claim that the lifetime of a dataset can exceed the lifetime of the DBMS that is used to maintain it (Salo-Merta et al., 2001). Therefore the development of the data system's database environment gets a step-by-step evolutionary nature.

In a step-by-step evolution strategy one proceeds with a series of development steps to achieve the desired state. For example, in the Next-Generation System Project of the Topographic Data System, The National Land Survey of Finland used this approach to restructure the data. The primary data loading into the new database included basic validation of attributes and topological conditions, and polygon reconstruction. Further data restructuring was carried out as a later stage. In this case it contained the merging of objects that had previously been fragmented by map sheet's edges. (Salo-Merta et al., 2001).

3.3 Developing databases in operational use

Perhaps the migration to a new DBMS is made with minimum changes to the structure of the data, with the intention to evolve the schema later in the future. However, the same principal problem, - how and when to propagate the structural changes to the data, - arises again if we need database schema evolution. For very-large production databases this may be a real life constraint in the system development and design, because the operational use of the database should not stop during a lengthy conversion phase. It may not be possible to stop the production for a full conversion. Instead, other alternatives have to be considered.

3.4 Pluralism in databases

Kucera (1999) discussed pluralism in geospatial databases and suggested the acceptance of data heterogeneity in contrast to a monolithic database, which reconciles any pluralism as part of the update process. In a pluralistic system a geographic feature may have several different representations stored 'as-is', as they existed in provided datasets, without attempts to integrate schemas, generalise, resolve spatial discontinuity, or otherwise encourage consistency in representation. Pluralistic data management and representation require specific techniques: e.g. versioning, metacontent description, feature linking with same-as links, on-line schema mapping etc.

4. INCREMENTAL DATA EVOLUTION APPROACH

4.1 Origin of incremental approach

The incremental approach originates from the software industry, where it has been applied for incremental development and compiling of software modules. In the context of GIS, it has previously been proposed for generalisation (Kilpeläinen & Sarjakoski, 1995).

Incremental updating of databases has recently become an issue of research and development. The International Cartographic

Association established a working group on Incremental Updating and Versioning in 1999. The working group's main interest relies on the management of updates between the base dataset supplier and the value-added dataset provider. Research issues include: bi-directional, multi-level, historical and temporal updating, planning for future changes, database maintenance, feature identifiers, modularity (dimension, context, layer, theme and size), inconsistent updating and simultaneous updating by field teams (Cooper & Peled, 2001). The implementation of new features to support incremental updating and versioning are likely to put pressure on product change or schema evolution on existing systems as well.

4.2 Production discontinuity problem

Traditionally database schema evolution is carried out by *full data conversion strategy* using the following sequence:

- Stop production
- Change the schema / Create a new schema
- Migrate all data to the new structure defined in the changed schema
- Continue production with the new schema

Here we assume that schema evolution includes such structural changes to the data, that a specific software module is required for migration (Fig. 3). This migration module may be a SQL-script, a database program, an application program or a translation process with external transfer files. We do not consider the implementation technology here. The characteristic of the migration process is the discontinuity that it causes to normal production. Normal production is carried out with software A using DB interface a. Data that conforms the new schema, has to be accessed with an updated interface a'. Therefore two versions of the DB interface are needed.

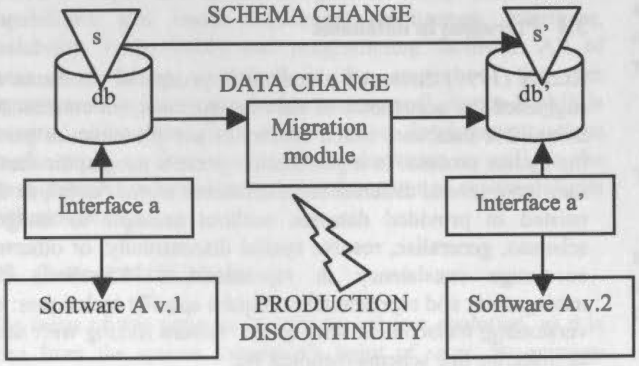


Figure 3. Migration module propagates schema changes to data. Migration causes discontinuity in production, because application program's DB interface needs updating to conform the new schema.

This approach can be described as one very-long transaction. From the production applications' point of view, the database is in inconsistent state during the migration. Since the migration starts, the whole database is not accessible for client A 1 anymore, but on the other hand, only the migrated objects are accessible to client A 2.

If the basic problem is the duration of data change and we assume that it cannot be solved reasonably, we should consider ways of managing the production and the migration simultaneously. Could it be possible to use versions 1 and 2 of client A and run the migration process in combination? In theory this could be possible in certain conditions. We need to be able to partition the data. That means recognizing and isolating meaningful and appropriate subsets of data for processing. In some data systems feasible partitioning criteria can be found, e.g. in working area based map production, but that is not always the case.

4.3 Principle of incremental data evolution

It may be a characteristic to a data system that the updates are made object-based and on-request. The requests scatter all around the dataset and no partitioning criteria for isolation can be found. In that case we suggest the acceptance of pluralism in schema definitions of object classes and consider means to manage it. We introduce an approach called *incremental data evolution* (IDE) with the following definition: Incremental data evolution is an approach where only those objects that are touched as part of the normal production process are migrated to the new structure defined in the changed schema.

We accept the fact that a feature may have alternative schema definitions as a permanent or at least long lasting state of the DB. To manage this kind of pluralism, we suggest certain changes in the DB's read/write –interface and embedding the migration module in the updating client (Fig. 4).

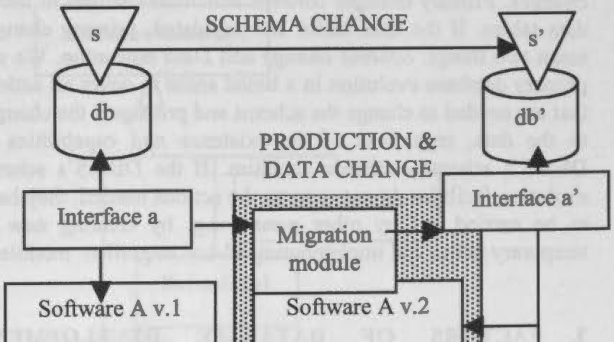


Figure 4. Migration module embedded in the production software module.

We call this approach incremental object evolution, because the schema evolution is propagated to the data incrementally, as the objects are touched in the normal production. It is not forced to the whole dataset, but only to those objects that are most actual.

4.4 Implementation aspects

We try to encapsulate the read/write –interface between the DB and applications to hide the pluralism and to encourage the data conversion from the old structure to the new. The database interface needs to be modified so that 'read' operations retrieve objects primarily from the new structure, and secondarily, for the objects that were not found, from the old structure (Fig. 5). Depending on the application case, the result sets for read operation may be separate or merged. 'Write' operations always save the objects only into the new structure, freeing the old structure. Therefore every processed object will be converted from the old structure into the new structure. If there are any problems in the merging, conversion or validation, they will be solved by the operator during the normal production work.

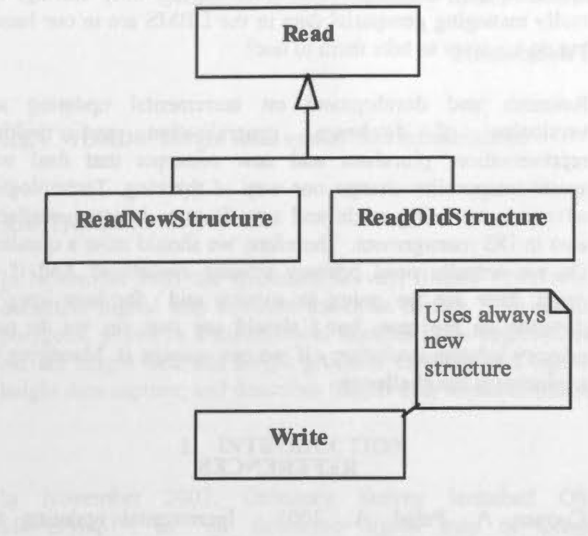


Figure 5. Principle of DB read/write interface implementation

5. CASE STUDY

The suitability of the incremental object evolution approach is considered on the Finnish Land Parcel Identification System (FLPIS). FLPIS is a national database for controlling the farmers' parcel-area-based subsidies that come from the European Commission. FLPIS is a subsystem of the Ministry of Agriculture and Forestry's data system for managing the farmer's subsidies. All activities on FLPIS have been contracted, currently to Genimap Ltd. The contractor is responsible for both the production work and the production environment, including system integration.

5.1 General description

Annually 20 000 farmers' data are updated by the contractor, according to requests from farmers and the controlling authorities. The production year is tightly time-scheduled with severe economic delay sanctions.

The system comprises the farmer information, orthophotos of the country and the parcel information including boundaries in vector form. The main functions of the system are to digitise the parcel boundaries, identify the parcels and calculate the parcel areas. The system preserves history of each project year's final states, which can be queried, viewed and compared to previous and later states.

The database is updated both based on the announcements from the farmers, and the requests from the controlling authorities. Both the farmers and the controlling authorities are served with massive document output, including maps.

The database can be divided into two parts: the core with the actual parcel data, and the extension for managing the production. There are client applications for:

- Data exchange (with Ministry's system)
- Document registration and management
- Digitising
- Validation
- Quality control (internal and external)
- Document output
- Reporting the production
- Extracting data on request

The database environment is Oracle Spatial 8.1.5. The client applications are implemented with a variety of tools including MapInfo Professional, MapBasic, Visual Basic, PL/SQL and Oracle Reports.

5.2 Motivation for change

The evolution of Ministry's data system for managing the farmer's subsidies follows the general pattern shown in Fig. 1. First the data production layer was established, with limited services: First paper documents only, then file extracts on request.

In 2000 the Ministry started the development of an intranet/extranet service pilot for the controlling authorities. The service is based on a service database. In the beginning it was a read-only -service, but new features for managing digital update requests with graphics from the authorities are developed and tested. The ultimate goal is to offer an up-to-date on-line service to the farmers as well, and reduce the need for paper documents.

The role of the production database has changed from a rather isolated system with only a few off-line interactions a year with the Ministry's systems, to a core dataset that needs to be replicated to the service database continuously. However, only those changes that have passed the two-phased quality control, are allowed to show. This generates new requirements to long transaction and version management. The major schema evolution challenges concern the modernisation and enhancement of the spatio-temporal capabilities of the system. More intelligence concerning the management of geographic objects would be implemented in database instead of the client application.

5.3 Applicability of incremental data evolution

The IDE approach is appealing to FLPIS for the following reasons:

- The project year is tightly scheduled. The scheduled breaks last only a couple of weeks, and causing longer breaks in production for conversions is not possible.
- The amount of historical data in the system is high, since the production started in 1996. The storage and management of historical data could remain as-is.
- IDE in combination with full conversion of subsets selected on area-basis would make an ideal processing method for FLPIS, because of it's flexibility.

On the implementation level, a new digitising client with embedded migration module and read/write DB interface would be needed. The data exchange module would need updating, but with streight-forward changes. The changes for document output client would be manageable.

6. CONCLUSIONS

This paper considers the applicability of incremental approach to database schema evolution, and proposes a complementary method for schema evolution and data re-structuring of large GI databases. It is proposed and considered as a resolution to a practical problem of schema evolution of a production database, in a situation where the possibilities of interrupting the production are limited. The options are either to arrange the data conversion concurrently with the ongoing production, or to forget about the schema evolution. The applicability of IDE approach is considered in a case study, but so far there is no implementation and therefore no proof of concept.

The advantage with incremental object evolution is that one does not have to run down the production line for the migration process. The schema evolution can be propagated to the data 'on-line' during normal production. The combination of IDE and full-conversion of selected subsets offers a flexible processing method.

There are risks in the IDE approach. If we consider a schema evolution step with data conversion as a very-long transaction, there is no reasonable 'rollback', if the conversion takes place embedded in the normal production. On the other hand, do we ever have a real rollback option in system change projects?

The database schema becomes more complicated as pluralism is accepted. But the pluralism does not have to be forever – the old structures may be deleted when there is no need to access them anymore. The pluralism within object definitions may be hidden by using DB views and encapsulation in the read/write interfaces, e.g. by database programming.

User requirements change in time, but so does the platform for database management. The advances in DBMS technology create a internal pressure for further development, as the changing user requirements create an external pressure. One has to migrate to new versions of DB products to keep one's system sustainable and watch out not to become obsolete. But do we actually need primary schema evolution that generates from technical advances? 'Why fix it if it ain't broken?'

The relational model is simple and provides a fairly low level of abstraction with tables, rows and columns. For geospatial data, a higher level of abstraction would be preferred, to make the investment cost of a new application reasonable, and to facilitate standards. Higher level data models for geometry and topology, that have been provided by proprietary GIS's, are coming into mainstream general DBMSs as well. Oracle has been developing the spatial concept step-by-step since the first release of Spatial Data Object (SDO) in 1994, including new features in each product version. However, the specific geospatial features provided with Oracle 9 Spatial are still far behind from some GIS vendor's DBMSs, e.g., there is no support for storing topology.

Let's assume an organisation that has implemented it's GIS on Oracle's BLOBs (Binary Large Objects) or first versions of spatial with the idea of getting a standard mainstream solution. However, Oracle's geospatial solution has not been adequate yet, and therefore plenty of proprietary features have been implemented in the database and the application programs by the system integrator, e.g. the management of topology and temporal dimension. As the DBMS's geospatial capabilities improve in the new product versions, they remain unexploited in the application, because the issues are handled already – with a proprietary, non-standard and probably complicated way. Under these circumstances, we may need to fix, although it wasn't actually broken, to survive and to gain the benefits of openness and standards. The tools for not only storing, but really managing geospatial data in the DBMS are in our hands, but do we want to take them to use?

Research and development on incremental updating and versioning of databases, generalisation and multiple representation, pluralism and new concepts that deal with spatio-temporality change our way of thinking. Technological advances are being made and new features become available, also in DB management. Therefore, we should raise a question: Do we actually need primary schema evolution? And if we don't, how are we going to survive and for how long? It depends on the case, but I should say that yes we do need primary schema evolution - if we can manage it. Managing the evolution is the challenge.

REFERENCES

Cooper, A., Peled, A., 2001. Incremental updating and versioning. In: *20th International Cartographic Conference, Conference Proceedings*, Beijing, China, Vol. 4, pp. 2804-2809.

European Union
http://www.cordis.lu/euroabstracts/en/june01/feature02.htm (accessed 2 Feb. 2002)

Finnish Land Parcel Identification System FLPIS. Genimap Ltd., Ministry of Agriculture and Forestry in Finland. (Unpublished system documentation)

Kilpeläinen, T., Sarjakoski T., 1995. Incremental Generalization for Multiple Representations of Geographic Objects. In: Müller, J-C, Langrane J-P, Weibel R., (eds) *GIS and generalization*, Gisdata 1, ESF, Masser, I., Salgé, F. (series eds.), Taylor & Francis, pp. 209-218.

Kucera, G., 1999. Pluralism in spatial information systems. Paper on the IV International Conference on GeoComputation, USA, on 25-28 July 1999. http://www.geovista.psu.edu/sites/geocomp99/Gc99/060/gc_060.htm (accessed 19 Feb. 2002)

Kucera, H., Lalonde, B., Lafond, P. 1999. GEO-2001:Designing and Building a Spatial Information Warehouse. *19th International Cartographic Conference, Ottawa, Canada. (unpublished workshop material)*

Salo-Merta, L., Tella, A., Vanhamaa, I. 2001. Database Design in Migration from Traditional to Object-Oriented GIS – the Evolution Story of the Topographic Database of Finland. In: *20th International Cartographic Conference, Conference Proceedings*, Beijing, China, Vol. 2, pp. 1393-1400.

ACKNOWLEDGEMENTS

The author wishes to acknowledge Mr. Matti Vilander from the Ministry of Agriculture and Forestry and Mr. Juha Haapamäki from Genimap Ltd. for their support and co-operation, and Prof. Tuija Helokunnas from Tampere University of Technology for support and comments concerning this paper.

DEVELOPING A NATIONAL HEIGHT DATABASE

David Holland
Ordnance Survey, Romsey Road, Southampton, SO16 4GU UK - dholland@ordsvy.gov.uk

Commission IV Working Group IV/4

KEY WORDS: Height data, spatial data infrastructure

ABSTRACT:

In November 2001 the Ordnance Survey, Britain's National Mapping Agency, launched the first release of OS MasterMap^{TM}, the definitive digital map database for Great Britain. The initial data within OS MasterMap consisted of topographic points, lines and polygons, stored as 2-dimensional features in an object-oriented database. This paper outlines the structure of OS MasterMap; the current height data and height products captured and supplied by Ordnance Survey; the research undertaken into new methods of height data capture; and describes the 3D data which Ordnance Survey may incorporate into the data in future.

1. INTRODUCTION

In November 2001, Ordnance Survey launched OS MasterMap™, as "the definitive digital map of Great Britain". The first release of OS MasterMap consists of large-scale topographic data, cleaned and processed into a topologically-stuctured point, line and polygon model. The OS MasterMap brand is designed to be continually updated and expanded to include many layers of spatial information, including imagery; transport networks; hydrological networks and height data. Height data, and its relationship to OS MasterMap, is the subject of this paper.

2. OS MasterMap

2.1.1 OS MasterMap data

OS MasterMap is based on the concept of a definitive core of spatial reference data, captured, maintained and supplied to a well-defined specification – forming a Digital National Framework for spatial data. The name OS MasterMap is the brand name of the product which conforms to this Digital National Framework..

OS MasterMap currently contains over 400 million spatial features, constructed from Ordnance Survey's large scale topographic data. Before the advent of OS MasterMap, the data consisted of unstructured points and lines, stored in almost 230,000 separate "tiles". During 2000-2001 Ordnance Survey undertook a major re-engineering programme, which transformed this "unintelligent" data into a single, seamless, topologically-structured point, line and polygon object database. This restructuring, and the services developed to deliver the data, has many benefits to the user, which are discussed below.

OS MasterMap contains features representing buildings, roads and paths, administrative boundaries, railways, water features, and many other topographic features. These are collected into themes (such as "buildings" and "roads, tracks and paths"), to allow users to extract data within a particular theme, or group of themes. Eventually OS MasterMap will encompass many other types of spatial information, which may include smaller scale topographic data (ideally derived from the large scale data), imagery, points of interest, integrated transport networks, address information, pre-build data (e.g. data from building plans before construction begins) and building heights and terrain models.

Each feature in OS MasterMap is assigned a unique identifier (a Topographic Identifier, or "TOID"), allowing users to make a definitive reference to any building, road segment, land parcel or any other feature in the database. This makes it much easier for users to associate their own information to the feature; to refer unambiguously to a particular feature and therefore to share spatial information with other users.

One of the great strengths of OS MasterMap is that the different spatial data layers will conform to a common framework. All features will be registered to the same reference system, enabling users to link information in one layer with data in another layer. This integration of data will help users to perform more complex analysis, and to build up applications which address their specific needs. For example, a mobile phone service provider may take the road network and the height data to determine which transport routes are likely to receive signals of sufficient strength from a proposed antenna site.

2.1.2 OS MasterMap Services

The OS MasterMap service allows users to make their orders and receive quotations over the Internet. The novel features of this service are that users select their own area of interest (unconstrained by arbitrary map-tile boundaries) and may choose a subset of the available themes. The data is supplied on CD or DVD, or may be downloaded via an ftp server.

One feature of the service which has been high on the requirements list of many users for a long time is "change only update". Previously, when a customer requested an update to their data, the data was supplied as whole "tiles", which included every feature in the area covered by the tile, whether or not it had been updated. The change-only-update service changes all this. After the initial supply of OS MasterMap data, users may make online queries of the database, to determine any changes within their area of interest since a given date. They can then request an update to their data, and will be supplied with only those features which have changed. Since a large proportion of OS

MasterMap customers require regular updates to their initial data holding, this will greatly reduce the amount of data which passes between the OS MasterMap database and the user.

2.1.3 Standards

When developing the Digital National Framework, it was clear from consultation with our customers and partners that the data should be made available to users in a standard format. The Geography Markup Language (GML), developed by the Open GIS Consortium (OGC), was suggested as a possible data format. There was wide support for this from our partners and software vendors, since it is an open, non-proprietary language, designed to work in an Internet environment across multiple platforms and applications. It is also a language based on XML, acknowledged as the universal format for structured data on the Web (Quin, 2001). As such, the tools to manipulate the data are widely available. Since almost all the major GIS vendors are members of the OGC, there is also strong support from within the Geographic Information industry for GML.

3. ORDNANCE SURVEY HEIGHT PRODUCTS

3.1 Land-Form PANORAMA

The current (2002) height products available from Ordnance Survey are sold under the "LandForm" product name. These include digital contour products derived from 1:50 000 scale and 1:10 000 scale source maps; and ground surface Digital Terrain Models (DTMs) derived from these contours augmented by spot heights captured by both photogrammetric and field survey methods. The Land-Form PANORAMA product, derived from 1:50 000 scale maps, is used extensively for such diverse applications as telecommunications network planning, visual impact assessment, and drainage network modelling. As a DTM, the data is structured as a grid, with a point-spacing of 50m and a vertical accuracy of ±5m (RMSE). Data for Land-Form PANORAMA was captured in the 1980's and has not been updated since – the product thus provides a snapshot in time of the general terrain structure of Great Britain.

3.2 Land-Form PROFILE

LandForm PROFILE is Ordnance Survey's detailed national height dataset. This product is based on contours digitised from the reprographic material used as the basis for 1:10 000 scale topographic maps. These contours were originally captured photogrammetrically, using analogue and analytical stereoplotters. The Land-Form PROFILE DTMs derived from these contours take the form of a grid, with a point-spacing of 10m and a vertical accuracy (RMSE) of between 1m and 1.8m (depending on the original source data).

Although the Land-Form PROFILE products have been used successfully in many application areas, they still retain some artefacts of their cartographic origins. For example, the contours on the maps were broken at embankments and cuttings, so these features are not included in the contour data. In the DTM creation process (based on a Delaunay triangulation of the contour and spot height data), the software automatically fills in the gaps between contours, but this can give a false picture of the terrain. This has become more of an issue as computer systems have improved; software capabilities have increased, leading to a demand for higher quality, more accurate height data. This trend is set to continue for several reasons. The fears over the rise in sea levels due to global warming, augmented by the increase in flooding in Britain in recent years, have ensured that accurate flood forecasting and effective flood defence mechanisms are both high on the government's agenda. Also, the telecommunications industry is looking for more accurate signal-propagation models, to allow them to optimise the siting of their antennae for the next generation of mobile communications networks. Both of these markets would benefit from detailed height data, especially when linked with other data such as property information (allowing insurance companies to determine the financial risk of flooding) or land use data (allowing telecommunications companies to differentiate between agricultural land and forest, and to adjust their predictions accordingly). A height model linked into a national spatial infrastructure, as provided by OS MasterMap, would be of great benefit to users in these, and other, markets.

Over the past few years, new techniques and new sensors have revolutionized the way in which height data is captured. During this time Ordnance Survey has been involved in several research projects which investigated the collection and management of height data. Several of these projects are outlined below.

3.3 High Resolution Satellite Imagery

In 1996, Ordnance Survey led a BNSC-supported project to investigate the potential of high-resolution satellite imagery for the creation of digital surface models. This project, detailed in Ridley *et al.* 1997, involved the analysis of simulated 1m-satellite data, derived from aerial photography. At the time, it was concluded that 1m satellite imagery would not provide an effective means of updating a national high-resolution DTM. To test this conclusion on real satellite data, Ordnance Survey planned to create DTMs from stereo IKONOS imagery, as part of a recent OEEPE project (Topographic Mapping from High Resolution Space Sensors – OEEPE report not yet published). Unfortunately it proved impossible to acquire suitable stereo imagery during the lifetime of the project, so this aspect of the research was dropped.

If possible, Ordnance Survey would still like to evaluate stereo IKONOS and/or QuickBird imagery, to determine whether they could produce elevation models of the quality required and at a price comparable with conventional data sources.

3.4 Lidar Data

An investigation into the use of lidar (airborne laser scanning) data for the capture of heights of features such as buildings, vegetation and embankments was undertaken in 1997/1998. Data for an area of NE England was kindly provided to Ordnance Survey by the Environment Agency, who have been collecting lidar data for flood-modelling purposes since 1997. The results of this research indicated that lidar could provide a very high resolution digital surface model (DSM) which could be used as a product in its own right. However, when it came to integrating lidar with existing height data (i.e. with the Land-Form PROFILE product), this proved to be more problematic. The very detailed lidar DSM includes all the surface features (buildings, vegetation, vehicles, etc) which must be filtered out to produce a ground-surface DTM compatible with Land-Form PROFILE. Algorithms to perform such filtering were in the early stages of development at the time, and a manual process proved to be very time-consuming and not particularly effective.

Lidar has now acquired a reputation as an effective way of collecting very-high resolution height information for a variety of applications (Flood, 2001, Plaster, 2001). Lidar has proved to be particularly effective in flood-plains and coastal areas where small differences in height are important in the planning of flood defences and in the determination of flood-risk; and where conventional techniques are often unsatisfactory.

3.4.1 Synthetic Aperture Radar (SAR)

Interferometric Synthetic Aperture Radar (inSAR) has been proposed as a tool for topographic mapping since the mid 1980's (Zebker & Goldstein, 1986). Both satellite and airborne SAR systems have been used for the rapid capture of DSMs over large areas. Recently, improvements in technology (such as the incorporation of integrated GPS and Inertial Measurement Units to determine the exact position and orientation of the sensor) have led to an increase in commercial airborne inSAR systems.

At the Ordnance Survey, height data produced using SAR Interferometry has been investigated as a possible replacement for the DTMs created from contours. University College London (UCL) carried out an evaluation of "Tandem" SAR data (from the European ERS-1 and ERS-2 satellites) as part of the aforementioned BNSC project (Ridley *et al* 1997). This research showed that the ERS tandem data could be used to produce DEMs with a height accuracy of between 5.7m and 20m RMSE. Elevation models of this accuracy are valuable, but they do not meet the required specification for a national height dataset in countries, such as Great Britain, which are relatively rich in geospatial information.

The potential of airborne SAR systems has also been investigated. In a joint project with QinetiQ, Ordnance Survey recently used SAR data for an area around Portsmouth in Hampshire. The SAR DEM was merged with Land-Form PROFILE data in an attempt to fill in the gaps between embankments and cuttings and within a disused quarry. Preliminary results from this research suggest that DEMs from airborne inSAR may be sufficient to update Land-Form PROFILE in rural areas, but they are not suitable for producing height models in urban areas. This technical research project did not investigate the cost-effectiveness of inSAR data, but this would obviously be a major factor in determining the potential of inSAR as a viable replacement for traditional photogrammetric height data capture.

3.5 Digital Photogrammetry

The conversion of Ordnance Survey's photogrammetric operations, from hybrid "analytical" systems to fully-digital workstations, took place in 2000. Stereo photography is now routinely scanned, passed through a digital aerial triangulation process, and analysed on digital photogrammetric workstations. The main purpose of these workstations is to update the topographic data for OS MasterMap, but they are also used to update the data used to produce Land-Form PROFILE. At the present time, the maintenance of Land-Form PROFILE occurs mainly in areas of major road development, and sites where old quarries and gravel pits have been re-landscaped. Methods of automatic DTM creation have been investigated and are beginning to be used within the production environment. As the software steadily improves, the process of updating Land-Form PROFILE data should become easier and more effective.

4. HEIGHT DATA WITHIN OS MASTERMAP

The initial development phase of OS MasterMap has concentrated on upgrading the 2D topographic data. However, it has always been envisaged that height will play a significant part of future development, and subsequent product releases. Work is underway to determine how best to capture height data in future, and how to integrate the height information with the other layers in OS MasterMap. Several of the technologies described above may be utilised in the development of the height layer, and it is quite possible that different technologies may be used for different areas of the country. For example, lidar is an excellent data capture tool in low-lying coastal areas, but there is little demand for lidar's level of accuracy in mountains or moorland, so it is not cost effective in such areas.

5. CONCLUSIONS

This paper has shown how Ordnance Survey has developed a new national spatial data framework and its associated products and services. OS MasterMap is designed to underpin the spatial data infrastructure of Great Britain. It will be continuously updated and improved, as new types of data and new services are introduced. Height data will play a role in this development, forming a layer within the OS MasterMap structure. Various different techniques for the capture of this data have been investigated, and it is expected that a hybrid solution will be adopted, to collect data which may be used in many varied applications and is fit-for-purpose in all the different terrain-types within Great Britain.

6. REFERENCES

Flood, M, 2001, Laser Altimetry: From Science to Commercial Lidar Mapping, *Photogrammetric Engineering & Remote Sensing*, **67**, no. 11, 1209-1217.

Plaster, R.L., 2001, Lidar: Transitioning Technologies from Lab to Applications Around the Globe, *Photogrammetric Engineering & Remote Sensing*, **67**, no. 11, 1219-1223

Quin, L. World Wide Web Consortium, 2001. The eXtensible Markup Language, http://www.w3.org/XML (accessed April 2002)

Ridley, H.M., Atkinson, P.M., Aplin, P., Muller, J-P. and Dowman, I., 1997, Evaluating the potential of the forthcoming commercial U.S. high-resolution satellite sensor imagery at the Ordnance Survey, *Photogrammetric Engineering and Remote Sensing*, **63**, 997-1005.

Zebker, H., and R. Goldstein, 1986, Topographic mapping from interferometric synthetic aperture radar observations, *J. Geophys. Res.*, **91**, 4993-5001.

MODELING MULTIPLE REPRESENTATIONS INTO SPATIAL DATA WAREHOUSES: A UML-BASED APPROACH

Bédard Yvan, Ph.D, Marie-Josée Proulx, M.Sc, Suzie Larrivée B.Sc., Eveline Bernier, M.Sc.

Centre for Research in Geomatics Laval University, Quebec City Canada G1K 7P4 {yvan.bedard, marie-josee.proulx, suzie.larrivee, eveline.bernier}@scg.ulaval.ca

KEY WORDS: GIS, multiple representation, spatial database modelling, spatial data warehouse, spatially-extended UML, Perceptory.

ISPRS Commission IV, Working Group IV/4

ABSTRACT: Creating database schemas that include complex geometries and that take into account map generalization as well as multiple geometric representations is quite a challenge. The created schemas are either too complex to become useful, too large to be usable, too detailed to be checked with the overall integrity in mind, or even too labor-intensive to even be created or edited. This paper presents a method and a tool that has been used with success in modeling generalization and multiple representation for a spatial data warehouse.

RÉSUMÉ: La création de schémas de bases de données qui incluent des géométries complexes et qui tiennent compte des processus de généralisation aussi bien que des représentations géométriques multiples représente tout un défi. Les schémas ainsi créés sont soit trop complexes pour être utiles, trop volumineux pour être utilisables, trop détaillés pour être vérifiables dans une optique d'intégrité générale du schéma, ou même trop exigeants à produire ou éditer. Cet article présente une méthode et un outil qui ont été utilisés avec succès pour la modélisation de la généralisation cartographique et la représentation multiple dans le cadre de création d'un entrepôt de données spatiales.

1. INTRODUCTION

Organizations involved in map production regularly provide several products at different map scales or for different purposes. Ideally, they maintain only one cartographic data warehouse and derive the different products from this warehouse. Since these products meet different goals, they often need to represent the same geographical features in manners that differ in their level of detail, their precision, their symbology, their aggregation and generalization states, etc. The diversity of products stemming from this unique dataset calls for efficient data processing if one wants to derive automatically every product. In fact, in many cases and more specifically when using semi-automated procedures, it becomes more efficient to keep in the database the results of the derivation, leading to several geometric representations for a same feature, i.e. to multiple representations. In some cases, it may even be impossible to derive the desired geometry of an object and additional data acquisition becomes required, also leading to the storing of multiple representations in the database.

From a data modelling point of view, this represents an interesting challenge since the level of complexity can become so high that the resulting database schema rapidly becomes unusable. In order to solve this problem, we propose a solution based on the UML object class model (Unified Modelling Language) that has been extended to better accommodate spatial data. The proposed solution uses UML stereotypes to represent in a very simple manner all the possible 0D, 1D and 2D geometries that an object may have (simple, complex, alternate, multiple). This solution also supports derived geometries as well as the derivation of attributes, object classes or associations. It has been implemented in a free visual modeling tool called Perceptory and the mapping to ISO/TC211 as well as OGC geometric and topological primitives has been done. The complete solution has been applied to the design of the provincial topographic spatial data warehouse of the Quebec Department of Energy and Resources. Thus, the present paper will start with an overview of some fundamental concepts of multiple representations and spatial data warehousing. We will also introduce a spatial database modelling solution we have developed: Perceptory. Then, we will describe the approach we have used with Perceptory to facilitate the modelling of multiple representations and generalization; in particular, we will describe how, using object-oriented concepts and examples extracted from our project with the Quebec Topographic Data Warehouse, we have combined multiple representations with automatic derivation in the database schema. Finally, we will discuss about the presented solution and conclude on this work.

2. FUNDAMENTAL CONCEPTS OF MULTIPLE REPRESENTATIONS AND SPATIAL DATA WAREHOUSING

Multiple representations result from seeing the world from different abstraction levels as well as different points of view. In fact, each user of a spatial database has its own needs regarding the representation of the objects. These representations may vary depending on several factors such as the user's needs or the level of details desired. Storing these multiple representations in order to better fulfill the needs of spatial databases users has been a challenge since many years.

Insofar, research in multiple representations has been focussing mostly on developing data structures. In most cases, these structures allow an object to have different geometries that only vary in their scale. We term them "multiscale structures".

Some of the most recent research projects aim at combining on-the-fly generalization with multiple representation structures [Bernier *et al.*, 2002; Cecconi 2001, Cecconi 2002]. Doing so results in a more flexible solution that can offer products adapted to the user's requirements and at any scale (given the on-the-fly generalization process). These solutions are especially useful for applications such as web-based on-demand mapping and spatial on-line analytical processing (SOLAP).

Of particular interest are the research projects using spatial data warehousing architectures to combine the above-mentioned solutions. A typical solution is a two-tiers data warehousing architecture where a multi-representation data warehouse is fed by independent data production systems and, where possible, by cartographic generalization processes. These acquired/derived data produce multiple representations of a same object (ex. a small town represented by a detailed polygon at 1:20000, a simplified polygon at 1:100000, a demographic centroid at 1:500000) that is then queried to provide only the appropriate geometry for the map being delivered on-line or being printed. In order to facilitate and accelerate the process for the end-users of the warehouse data, three-tiers infrastructures using datamarts can be built around pivot scales (ex. 1:50000, 1:100000, 1:250000) and pivot-themes (ex. transport) that provide "almost-ready" ranges of maps that need only minimal on-the-fly processing to satisfy the users (Bédard et al 2001a). These datamarts are used to quickly derive topic-oriented maps at scales compatible with the pivot scale of the datamart, mostly (but not uniquely) using automatic generalization capabilities (ex. producing a road map at 1:125000 from the transportation 1:100000 datamart). This differs from the data warehouse that relies primarily, but not uniquely, on multiple representations to support the pivot datamarts and to provide object-referencing capabilities across representations. In such three-tiers architecture, automatic derivation is used wherever possible to reduce data updating problems across representations and data acquisition costs; otherwise multiple representations are used.

Finally, few research projects have tried to find a flexible manner to conceptually describe multiple representations and generalization (ex. Proulx et al 2002; Spaccapietra, Parent and Vangenot 2002). Based on a case tool called Perceptory, we have developed a new solution for the modelling of spatial databases that support multiple representations and generalization processes.

3. PERCEPTORY: A SPATIAL DATA MODELLING SOLUTION

Perceptory is a visual modelling tool for designing conceptual object class models based on the UML object-oriented formalism and which includes extensions that facilitate the modelling of spatial databases. Perceptory includes several UML basic components (package, class, attribute, operation, association, multiplicity, aggregation/composition, generalization/ specialization, comments, etc.) as well as formal extension components (called stereotypes) that allow the modelling of spatial characteristics of cartographic objects in a simple manner (Bédard, 1999a) These extensions can be used within the three sections of any instantiated, abstract and associative class, i.e.:

1. at the class name level, to depict the geometry of the objects of a class;
2. at the attribute level, to describe the geometry of the descriptive characteristics within objects of the class; and
3. at the operation level, to describe the initial or the resulting geometry of a process occurring on the geometry of the class.

These stereotypes are fully illustrated by pictogram embedded in a font called PVL (Plug-in for Visual Language). This font can be used by any case tool, including Perceptory. However, the latter uses it in a much more simple and elegant manner. Perceptory proposes a schema and a data dictionary that includes UML components and some ISO/TC211 (19103, 19107, 19110, 19115) standard components. The UML schema extended with the PVL illustrates the content of the future database while the dictionary describes in more details the schema components with a natural (eg. english or french), formal (eg. Z, OCL) or coded language (eg. datatypes, nulls). The spatial PVL was especially built to visually describe, in a spatial database conceptual schema, the geometric dimension (0D, 1D, 2D) of an object to appear on a map, either following digitizing or derivation. The spatial PVL is also used to visually describe the geometric dimension or spatial extent of an attribute value that is heterogeneous within the boundaries of an object (ex. the attribute "number of lanes" of the class "road" may change within an instance of road and be located with a linear referencing system using "from" and "to" distances from an anchor point).

The three basic geometries of the spatial PVL can be used alone to describe a simple geometry or can be combined to describe an alternate, a complex or a multiple geometry. Each semantic occurrence of a spatial object class can be represented either by:

1. a simple geometry, i.e. composed uniquely by one occurrence of a geometric object;
2. a unique geometric aggregate, i.e. a simple aggregate geometry composed by geometric objects of the same dimension or a complex aggregate geometry composed by geometric objects of different dimensions;
3. an alternate geometry if the geometry can be represented by one dimension or another. The proposed geometries can be simple, simple aggregated or complex;
4. several independent geometries, where usually only one is used at a time. Thus, we say that the geometry of this object class is multiple. These geometries can be simple, simple aggregated, complex or alternatives.

The following figure (figure 1) shows the hierarchies between the different types of geometry used by Perceptory and gives an example for each one.

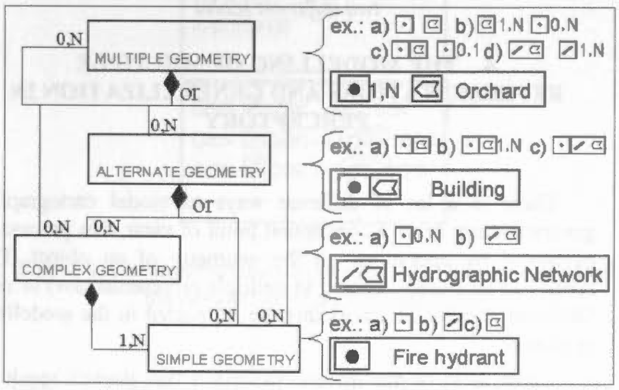


Figure 1. *Examples of possible geometry types (OR is and inclusive OR).*

The geometry can be described in much more details in the dictionary of Perceptory. We precisely define each geometry of an object, i.e. the data acquisition specifications (the "what" and not the "how"), we also include additional information depending on the geometry type (point, line, polygon) as well as the minimal geometric dimensions required for digitalizing the object (ex. point if area < 1000, polygon if area > 1000).

Just like the class attributes, the geometry can also be derived. The pictograms are then in *italic*, which visually remind the slash (/) of UML normally used in front of derived attributes. To do so, the derivation process must be entirely automatic, i.e. without any manual processes. In the case of multiple or alternative geometries, it is possible to have only one derived geometry.

Although an object can have different geometries for the same scale (for example, a house can be mapped according to its roof or its foundations) the kind of multiple representations that we are interested in are the multiscale representations. The following figure is based on a spatial database that can produce paper maps at the scales of 1:1000 and 1:20 000. At the first scale, all the buildings are represented by a polygonal geometry while at the second scale (the smaller one), only the public ones are represented by a derived polygonal geometry. Besides, roads have a polygonal geometry at the scale of 1:1000 and a derived linear geometry at the scale of 1:20 000.

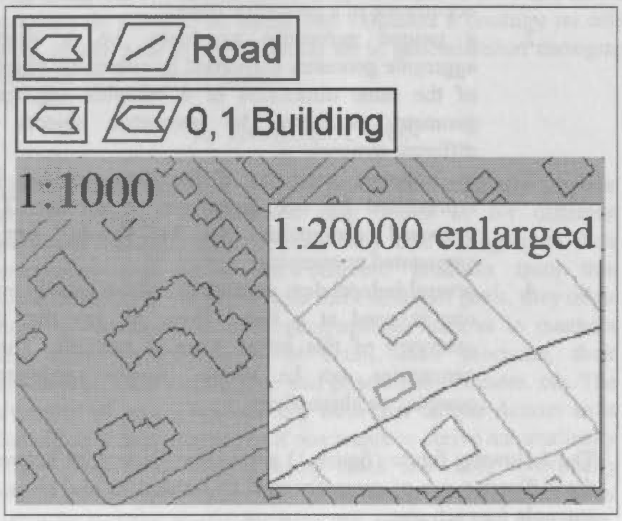


Figure 2. *The modelling of an object class that is represented at two different scales*

4. THE MODELLING OF MULTIPLE REPRESENTATIONS AND GENERALIZATION IN PERCEPTORY

There is a lot of different ways to model cartographic generalization. From a conceptual point of view, this process is expressed by operations on the geometry of an object. The result can be stored (leading to multiple representations) or not. Different possibilities must then be supported in the modelling approach:

- automatic on-the-fly generalization that doesn't result in the storage of multiple representations;
- automatic generalization that leads to stored representations for performance reasons;
- semi-automatic generalization that also leads to storing the results (to avoir repeating the operation);
- manual generalization that leads to storing the results;
- several data acquisitions for the same object in order to obtain the representations needed for every scale.

4.1 General modelling rules

According to our approach, the generalization operations are modelled the same way as any other operation in an object-oriented model. In addition, given the typical rules that govern the modelling process, we only model the data and the operations that must be programmed in the database, i.e. we don't represent the operations used to populate the database. Thus, in theory, we only model on-the-fly generalization (i.e. automatic generalization where the result is not stored in the database) without modelling the generalization operations that leads to multiple representations in the database. Nevertheless, because automatic generalization is very context-sensitive and it is difficult to distinguish in advance what will require human assistance from what can be done automatically (ex.with small programs triggered by the objects), it is possible to include in the first iteration of the model all generalization operations and, after some testing, keep in the last iteration of the model only those operations that can be done fully automatic. By applying these modelling rules, we obtain a database schema corresponding to the implemented data warehouse.

4.2 Specific modelling rules

The modelling rules that express generalization and multiple representations according to our approach are hereunder:

Rule #1: all the generalization operations are defined by an operation in the object class (in the operation section) whether these processes are manual, semi-automatic or entirely automatic (keeping only the automatic ones for the last iteration);

Rule #2: all the multiple representations are defined by the PVL using a multiple geometry pictogram in the object class;

Rule #3 : if the geometry resulting from a generalization process (manual, semi-automatic or automatic) has to be stored in the system, this geometry will be added in the object class with the PVL multiple geometry pictogram.

As previously mentioned, when building the schema, it can be useful to proceed by iteration and to retain only the operations that will be implemented to generalize automatically the object classes of the database.

4.3 The graphic notation

Following is the graphic notation used to model the above rules.

Generalization operations:

- if the generalization process relies on a unique operator, the name of this operator is used as the operation identifier (eg. collapse, simplification, displacement). If the generalization is done by applying a set of generalization operators, we use a general operation called "GEN" and we describe the operators sequence in the data dictionary.

- we indicate, in the operation, the initial and the resulting geometries with PVL pictograms. The pictogram that represents the resulting geometry is in italic if the process can be partly or completely automated (i.e. derivable) (eg. GEN ⧄ -> *⧄*). If generalization is entirely manual, then the pictogram uses a regular style (ex. GEN ⧄ -> ⧄).

- we also identify the scale or the abstraction level for each geometry (ex. GEN ⧄ 20K -> *⧄* 100K).

Multiple representations

- all the geometries that can be used to represent an object are illustrated by a PVL multiple geometry pictogram in the object class (upper left corner);

- the geometries are ordered from the detailed level to the more general level (from the largest scale to the smallest one);

- the geometries that result from a generalization process and that are stored in the database are defined by a derived multiple geometry pictogram (see rule #3).

4.4 Application examples

We have applied this modelling method for the design of a spatial data warehouse for the Quebec Ministry of Natural Resources (Bédard & *al.* 2001a).

To better explain the different concepts presented so far, we will use two examples from this application. These examples are based on cartographic objects used in different scales: 20 000, 100 000, 250 000, 1 000 000 and 8 000 000 (i.e. 20K, 100K, 250K, 1M et 8M). The first modelling example shows an object class (waterstream sections) for which all geometries (except the most detailed one) can be obtained from generalization, but some of them are stored in the database because their generalization process is quite complex. The second example presents an object class (airport) for which the generalization process give rise to an object replacement in addition to a geometric change.

4.4.1 Example 1: The modelling of multiple representations and generalization for a waterstream section

Let's take a waterstream section that is acquired for the scale of 20k according to certain geometric rules (eg. is a polygon if its area > $2000m^2$ and its minimal width is 20m, is a line if its area < $2000m^2$ and its minimal length is 150m). So, in the object class, we will have an alternate linear or polygonal geometry (⧄▣) (figure 3). The geometric rules governing the way each geometry is acquired are kept in the data dictionary.

The geometries at the scale of 20K are then generalized to obtain the geometries used at the scale of 100k. Given that a waterstream sections can have two geometries at the scale of 20k, we thus have two different generalization operations. The first one starts with the linear geometry and results also with a linear geometry. The second one starts with the polygonal geometry and results in either a linear or a polygonal geometry, depending on certain rules. The term "GEN" shows that it is a complex operation that includes more than one generalization operators. The complete set of operators as well as their order are indicated in the data dictionary. In the same way, the polygonal geometry at the scale of 20k is also used to obtain the geometry at the scale of 250k, leading also to a linear or a polygonal geometry according to specific rules. Finally, the conceptual object class depicts two geometries acquired alternatively at the scale of 20k and three generalization operations that produce the geometries for the 100k and 250k scales.

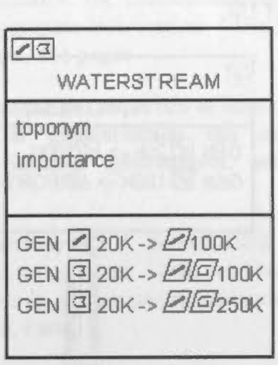


Figure 3. *The modelling of the object class " Waterstream ".*

During the modelling phase, it is possible to find some operations that turned out to be too complex to automate so the resulting geometries should be stored in the database. For instance, let say that it is the case for the generalization process that produces the waterstream section geometry for the scale of 250k. So we just have to add these geometries on the object class (by a derived multiple geometry pictogram, figure 4) to immediately see the multiple representations of this class.

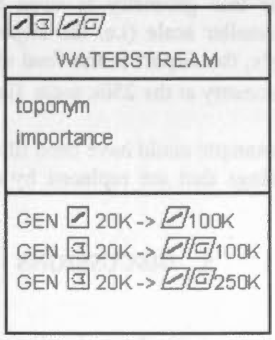


Figure 4. *The waterstream object class after having evaluated the complexity of a given operation*

4.4.2 Example 2: The modelling of a generalization that gives rise to an object replacement

In some cases, an object can have different semantics depending on the scale. For example, at a large scale an airport is represented by its landing strips and its main buildings while at a smaller scale, the same airport is represented by a simple point. At the beginning, there must exist in the database schema a semantic relationship of composition between the landing strips class and the building class with the airport aggregate class. So during the generalization process, the two first objects will be replaced by the third one, i.e. the airport. Given our graphic notation, it is possible to express such a replacement by including the name of the future object directly in the generalization operation of the object class that will be replaced (ex. AIRPORT ⊡ 250K, figure 5).

First of all, the landing strips are acquired at the scale of 20k. These lines are then generalized for the 100k and replaced by a point airport at 250k. This point airport is then generalized for the scale of 1M.

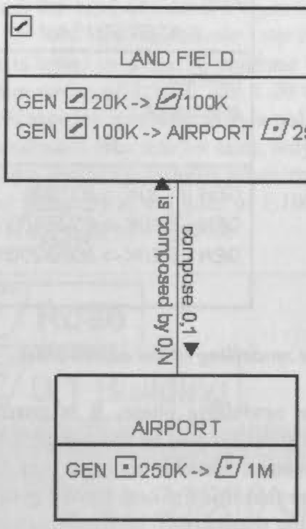


Figure 5. *Example of generalization for airport*

However, in the final database schema, the analyst will probably find more appropriate to store the point geometry of the airport at the scale of 250k instead of systematically derive it. Especially if this geometry is used to produce another geometry at a smaller scale (i.e. the airport geometry for the 1M). Accordingly, the airport in the final schema would have a derived point geometry at the 250k scale (i.e.⊡).

The same example could have been illustrated by point and polygonal buildings that are replaced by an urban zone at a smaller scale.

5. DISCUSSIONS

This new approach has been developed and tested for the modelling of the spatial data warehouse of the Quebec Ministry of Natural Resources (Bédard & *al.* 2001a). Using these simple modelling rules, it was possible to adequately represent the geometries of object classes for the different levels of the data warehouse, that is to say five cartographic scales (20k, 100K, 250K, 1M et 8M). These scales supports, after having been integrated, the objects imported from 10 different products or legacy systems.

Using a traditional modelling approach (based on simple geometric representations and without generalization), we had to model 108 object classes in order to depict the 5 levels of detail of the data warehouse. Obviously, there were several object classes that differed only by their geometries (quarry 20K, quarry 250K, etc.). Although the generalization relations between these classes showed the existence of these types of relation, they lack the richness needed to detail the generalization operations. Also, the schema was difficult to read given the high number of elements (class and relation) when done with a traditional approach (figure 6).

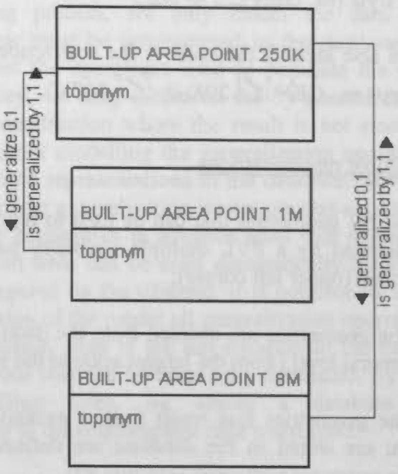


Figure 6. *Example of a traditional approach to model the object class BUILT-UP AREA (250K, 1M et 8M) using relationships.*

Another traditional approach regroups all the different geometries of an object in the form of attributes, which reduces the number of object classes present in the schema. In fact, by using this solution for the modelling of multiple geometric representations and generalization, we reduced the number of object classes in our schema by 40% (i.e. we went from 108 to 66 object classes). This solution is also much more easy to understand since all the representations of a same object are recorded at the same place (i.e. one geographical phenomenon results in only one object class). Furthermore, it would be better to remove the relations between the object classes and describe them as operations in the class (figure 7).

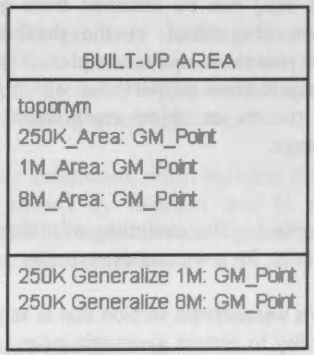


Figure 7. *Example of a traditional approach to model the object class BUILT-UP AREA (250K, 1M et 8M) using attributes and operations.*

Nevertheless, without spatial PVL, the model remains heavy and difficult to read, especially for those who are not familiar with ISO or OCG primitives. The use of PVL pictograms facilitates the reading of the schema while increasing its power of expression. It reduces its size significantly, especially when dealing with alternate geometries. This results in less work to draw the schema, less work to edit the schema, easier reading, easier checking and easier discussions with the clients (fig. 8).

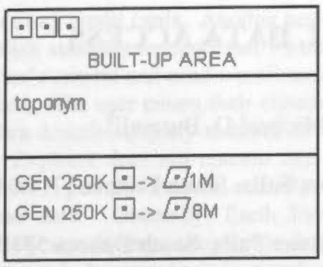


Figure 8. *Example of the used approach to model the object class BUILT-UP AREA (250K, 1M et 8M) with PVL pictograms in Perceptory.*

6. CONCLUSION

We showed in this paper that it is possible to use the spatial PVL in Perceptory to efficiently model geometric multiple representations and cartographic generalization. Given its richness of expression, Perceptory helps analysts in the modeling of multiscale databases. Although we present just a subset of our solution (i.e. geometric multiple representations and cartographic generalization), our approach supports also semantic and temporal multiple representations and generalization. Other works are in progress in related application domains and will help us to improve our solution and make it more flexible and powerful.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support of the GEOIDE Network of Centers of Excellence via the DEC#9 project: *Development of automated techniques to extract, generalize, and access geospatial information from hyperspatial remotely sensed data.* The authors also acknowledge the financial support of the Natural Sciences and Engineering Research Council and of the Quebec Ministry of Natural Resources (DGIG).

REFERENCES

Bédard, Y., 1999a «Visual Modeling of Spatial Databases Towards Spatial Extensions and UML», Geomatica, Vol 53, No2, pp. 169-186, 1999.

Bédard, Y., Proulx, M.-J. Larrivée, S., Nadeau, M., Frenette, P., Rivest S., 2001a « Mise en place d'une infrastructure d'accès aux données de la DGIG- Modélisation des données de l'entrepôt, Avis sur les orientations d'architecture technologique pouvant influencer la modélisation. » Version préliminaire du 21 mai, Rapport de recherche produit pour le Ministère des ressources naturelles du Québec, Centre de recherche en géomatique, Université Laval, 65 pages.

Bernier, E., Bédard, Y, Cecconi, A., Weibel, R., 2002, « La généralisation à la volée », Book chapter submitted to Hermès Edition, Paris.

Cecconi A., Weibel R. «Map Generalization for On-demand Mapping», GIM International, 15(12), p. 12-15, 2001.

Cecconi A., Weibel R., Barrault M. «Improving Automated Generalization for On-demand Web Mapping by Multiscale Databases», 10th International Symposium on Spatial Data

International Organization for Standardization, ISO/TC 211 19103, 2001a « Geographic Information- Conceptual schema language », Draft for International Standard, 75 pages

International Organization for Standardization ISO/TC 211 19107, 2001b « Geographic Information- Spatial Schema », Draft for International Standard, 185 pages

International Organization for Standardization, ISO/TC 211 19110, 2001c « Geographic Information- Methodology for feature cataloguing », Draft for International Standard, 60 p.

International Organization for Standardization, ISO/TC 211 19115, 2001d « Geographic Information- Metadata », Draft for International Standard, 151 pages

Martel, C, 1997. Analyse des capacités et limites de MGE/Map Generalizer pour la généralisation des cartes routières. Département des Sciences géomatiques, mémoire de fin d'étude.

Proulx, M.-J., S. Larrivée and Y. Bédard 2002. Représentation multiple et généralisation avec UML et l'outil Perceptory. Chapitre de livre sur la généralisation cartographique, soumis aux Éditions Hermes, Paris.

Poole, J., Chang, D., Tolbert, D., Mellor, D., 2002. Common Warehouse Metamodel, An Introduction to the Standard for Data Warehouse Integration. Willey & son, 269 pages.

Spaccapietra S., Parent C., Vangenot C. « GIS Databases: From Multiscale to MultiRepresentation», Abstraction, Reformulation, and Approximation, B.Y.Choueiry and T.Walsh (Eds.), LNAI 1864, Springer, 200 (Proceedings 4th InternationalSymposium, SARA-2000, Horseshoe Bay, Texas, USA, July 26-29, 2000).

U.S. GEOLOGICAL SURVEY SPATIAL DATA ACCESS

John L. Faundeen [a], *, Ronald L. Kanengieter [b], Micheal D. Buswell[b]

[a] U.S. Geological Survey, EROS Data Center, 47914 252nd Street, Sioux Falls, South Dakota 57198-0001, USA - faundeen@usgs.gov

[b] Raytheon STX Corporation, EROS Data Center, 47914 252nd Street, Sioux Falls, South Dakota 57198-000, USA - kanengieter@usgs.gov; buswell@usgs.gov

Commission VI, WG IV/4

KEY WORDS: *The National Map*, Earth Explorer, Seamless Data Distribution System, Web Mapping Portal, spatial data, land remote sensing, Web data delivery

ABSTRACT:

The U.S. Geological Survey (USGS) has done a progress review on improving access to its spatial data holdings over the Web. The USGS EROS Data Center has created three major Web-based interfaces to deliver spatial data to the general public; they are Earth Explorer, the Seamless Data Distribution System (SDDS), and the USGS Web Mapping Portal. Lessons were learned in developing these systems, and various resources were needed for their implementation.

The USGS serves as a fact-finding agency in the U.S. Government that collects, monitors, analyzes, and provides scientific information about natural resource conditions and issues. To carry out its mission, the USGS has created and managed spatial data since its inception. Originally relying on paper maps, the USGS now uses advanced technology to produce digital representations of the Earth's features. The spatial products of the USGS include both source and derivative data. Derivative datasets include Digital Orthophoto Quadrangles (DOQ), Digital Elevation Models, Digital Line Graphs, land-cover Digital Raster Graphics, and the seamless National Elevation Dataset. These products, created with automated processes, use aerial photographs, satellite images, or other cartographic information such as scanned paper maps as source data.

With Earth Explorer, users can search multiple inventories through metadata queries and can browse satellite and DOQ imagery. They can place orders and make payment through secure credit card transactions. Some USGS spatial data can be accessed with SDDS. The SDDS uses an ArcIMS map service interface to identify the user's areas of interest and determine the output format; it allows the user to either download the actual spatial data directly for small areas or place orders for larger areas to be delivered on media. The USGS Web Mapping Portal provides views of national and international datasets through an ArcIMS map service interface. In addition, the map portal posts news about new map services available from the USGS, many simultaneously published on the Environmental Systems Research Institute Geography Network.

These three information systems use new software tools and expanded hardware to meet the requirements of the users. The systems are designed to handle the required workload and are relatively easy to enhance and maintain. The software tools give users a high level of functionality and help the system conform to industry standards. The hardware and software architecture is designed to handle the large amounts of spatial data and Internet traffic required by the information systems. Last, customer support was needed to answer questions, monitor e-mail, and report customer problems.

1. INTRODUCTION

Access to the large spatial data holdings of the U.S. Geological Survey (USGS) is required by U.S. public law in some cases. The law states, "It is in the best interest of the United States to maintain a permanent, comprehensive Government archive of global Landsat and other land remote sensing data ….shall provide for long-term storage, maintenance, and upgrading of a basic, global, land remote sensing data set and timely access for parties requesting data." [2] The mission of the USGS's National Mapping Discipline is to meet the Nation's need for basic spatial data, ensuring access to and advancing the application of these data and other related earth science information for users worldwide.[3]

Online access to USGS data began in 1991 with the release of a system called the Global Land Information System (GLIS). Users could connect to the service with Telnet or dial-up modems. Queries using spatial, temporal, and dataset-unique criteria were allowed, resulting in a list of the files, photographs, or images that matched the user's criteria. Graphics were employed to illustrate the spatial extent the data represented on the Earth. Some data were subsampled and provided as browse images as a means to preview the actual products. This capability was used primarily to assist in the determination of cloud cover and the quality of satellite imagery.

The GLIS was replaced in 2000 by Earth Explorer, a system with considerably more capabilities. The system still used metadata queries, but users could search multiple inventories from a single query, save the metadata in the Federal Geographic Data Committee (FGDC) Geo profile, and place orders securely using credit cards. Another new feature was the ability to establish standing requests that would query for data based on the user's criteria and send e-mail messages whenever a match occurred. The user enters their criteria once and then receives metadata describing query matches entirely outside the system. Earth Explorer does not contain any data within the interface itself, but it portrays datasets and allows users to place orders for those data. Generally, Earth Explorer lists and describes remotely sensed and cartographic data. Examples of the remotely sensed datasets include Landsat Multispectral Scanner (MSS), Corona, and aerial photographs. Cartographic holdings include Digital Line Graphics and Digital Raster Graphics (DRG). Table 1 is a complete list of the datasets represented and the number of records for each.

---

* Corresponding author

Earth Explorer is available at URL http://earthexplorer.usgs.gov.

Dataset	Number of Records
CORONA	853,000
Landsat MSS	631,178
Landsat TM	480,222
Landsat ETM+	194,684
AVHRR	257,675
SPOT	720,000
Digital Orthophoto Quadrangles	221,337
National Aerial Photography Program	1,629,838
National High-Altitude Photography	511,844
Digital Line Graphs	195,356
Digital Elevation Models	90,231
Digital Raster Graphics	57,285
Paper Maps	56,000

Table 1. Datasets and their total number of records represented in Earth Explorer.

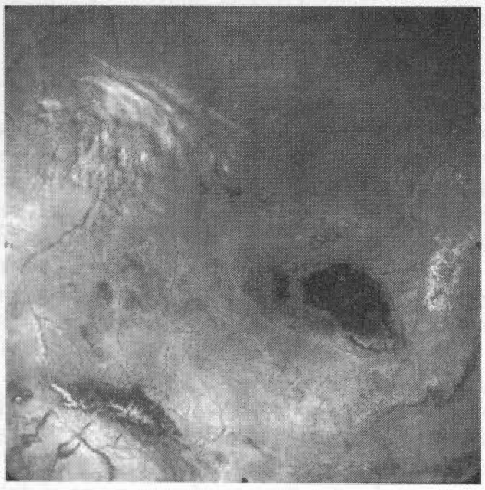


Figure 1. CORONA image over Black Hills, South Dakota, August 21, 1964. (Image DS09066A063MC065)

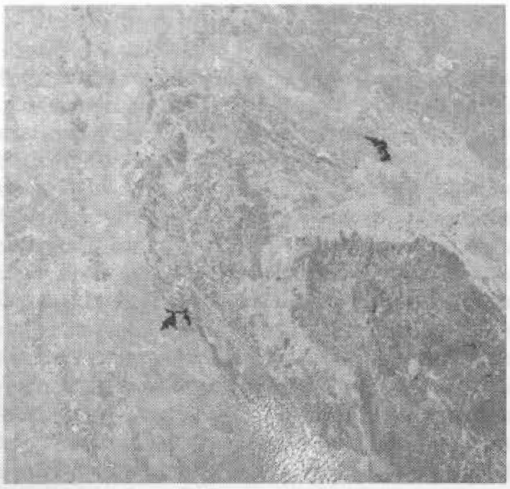


Figure 2. Landsat ETM+ image over Black Hills, South Dakota, August 2, 1999. (Image 7034029009921450)

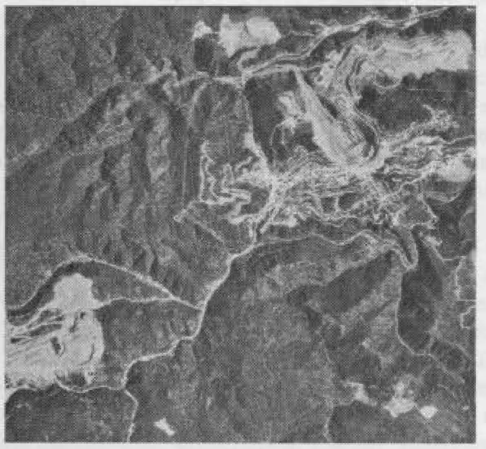


Figure 3. Digital Orthophoto Quadrangle image in South Dakota, June 16, 1995. (Image DI00000000042790)

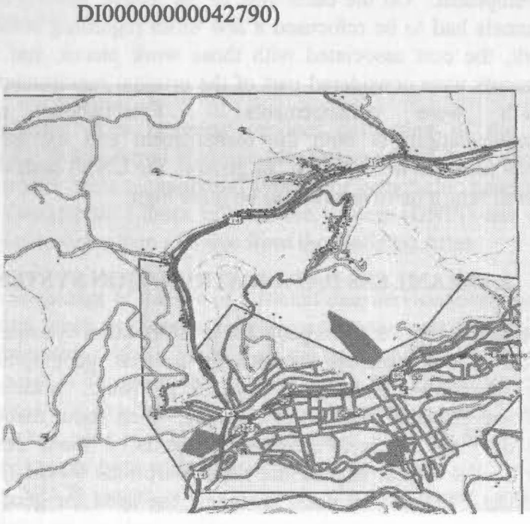


Figure 4. Digital line graph image in South Dakota. (Image DL025210HY144244)

Figures 1 through 4 illustrate examples of the CORONA, Landsat ETM+, Digital Orthophoto Quadrangle, and Digital Line Graph datasets (available through Earth Explorer).

2. EARTH EXPLORER

The resources to design, build, and operate Earth Explorer are different from those the USGS had used previously. After compiling the requirements, USGS staff did internal analysis to determine whether the USGS possessed the skills and experience to design and build an interface that would meet the requirements, which included a firm delivery schedule. The analysis indicated that the USGS possessed the skill to do the work and some of the experience, but the effort could not be accomplished within the schedule required. The requirements were advertised through an RFP (request for proposal) and bids were solicited from commercial vendors. A company was selected and contractual arrangements were made regarding the design and building of this next-generation system. USGS staff representing database management, information systems, and software engineering held numerous meetings and teleconferences with the commercial vendor to ensure that the requirements were understood. A schedule was established that conveyed the periodic management and technical meetings, the initial deliverables, and an acceptance test period. After the first interface code was officially accepted, mechanisms were established to accommodate trouble reports and enhancements not covered within the requirements. As tracking the delivered code became more complex, a configuration management system was used to ensure proper code migration from a development, to a testing, and finally to a production environment.

The USGS learned many valuable lessons through its first use of a commercial vendor arrangement in a public information system. The commercial vendor was able to deliver new and patched code much faster than the USGS could. The USGS could extend personnel resources in its software engineering by leveraging the commercial vendor for much of the work rather than controlling all aspects of an information system development project from within the organization. The commercial vendor was able to keep the USGS abreast of developing international and ad hoc standards that benefited development. On the other side of the ledger, communication channels had to be refocused a few times regarding authorized work, the cost associated with those work pieces, and which elements were considered part of the original requirements and which were enhancements. Establishing regular communications at both the management and the technical levels resolved most issues. In general, the USGS considers its overall return on investment to be quite high.

3. SEAMLESS DATA DISTRIBUTION SYSTEM

One of the key trends in geographic information systems (GIS) is the rise of seamless datasets. In the past, geographic data were partitioned because of physical, hardware, software, and political considerations. For example, when paper maps were the dominant form of geographic media, it made sense to partition a dataset so that the paper partitions would fit on a desktop. Today GIS users dominate the need for geographic data. Their computer processing, storage, and network capabilities continue to increase, making GIS more powerful and able to process larger datasets. Many researchers are using GIS to study features or trends that are global or national in scope. By removing arbitrary boundaries, seamless datasets provide these researchers a more natural and powerful mechanism to analyze and study large areas of interest. This is useful in emergencies or disasters, when data need to be available for immediate use. Furthermore, during emergencies there is not time to search for the best data available over a region or to edge match and process partitions of a dataset to a desired quality and consistency. The data need to be GIS ready, the best available, of consistent quality, and accessible. All these requirements have given rise to seamless datasets and their distribution on the Internet.

To meet these requirements, the USGS released the Seamless Data Distribution System (SDDS) in January 2001. The SDDS allows Internet users to determine areas of interest, established output parameters, and directly download actual spatial data. Currently, 30-meter resolution National Elevation Data (NED) and National Land Cover Data (NLCD) seamless datasets are available from SDDS. As the USGS continues to develop seamless datasets, more will be available to the public. Table 2 provides more information on the seamless datasets available for ordering and datasets used for reference layers. The SDDS is available at URL http://gisdata.usgs.gov/seamless.

Dataset	Coverage	Data File Size
National Elevation Data (NED) *	Conterminous U.S., Alaska, Hawaii, Puerto Rico, and Virgin Islands	60 GB - Demgrids 20 GB -Shaded Relief 1 GB – Metadata
National Land Cover Data (NLCD) *	Conterminous U.S.	16.25 GB Integer Data .001 GB – Metadata
National Hydrography Dataset (NHD)	Conterminous U.S., Alaska, Hawaii, Puerto Rico, and Virgin Islands	22 GB – Data 8.6 GB – Reference Layer
Geographic Names Information System (GNIS) - Condensed Version	Conterminous U.S., Alaska, Hawaii, Puerto Rico, and Virgin Islands	.015 GB – Reference Layer
Bureau of Transportation and Statistics (BTS) Transportation	Conterminous U.S.	4 GB – Reference Layer
National Atlas Reference Layers: U.S. State and County Boundaries	Conterminous U.S., Alaska, and Hawaii	.32 GB – Reference layer

** Datasets available from SDDS.*

Table 2. Seamless datasets provided by the USGS and map services used as reference layers.

Greater resources are required to design, build, and operate a seamless information system than for a metadata-based information system. Seamless data required changes in the way data are stored, maintained, and accessed. The data and metadata are stored online, so a large amount of disk space is required. Special emphasis is placed on the maintenance of seamless data to ensure that they remain the most current, best available, and of consistent quality. The distribution of seamless data over the Internet requires a large communications infrastructure with adequate server resources.

Because the seamless data concept is new, a different user interface was envisaged. Rather than focusing on the metadata, the seamless interface focused on the data. The user interface is much more graphical; an example is shown in Figure 5. The view of the data rendered in the interface is a map service. Each new display layer added or deleted, and each zoom, pan, or select command chosen by the user create a new map that is delivered live over the Internet. Map content and geographic coverage defined by the USGS can contain one or more data layers, depending on the purpose of the map. The SDDS uses NED and NLCD data layers to preview orderable data. Reference data layers (e.g., state and county boundaries, city names, streams, and roads) help define the area of interest. The map services displayed in the SDDS interface uses map pyramids and scale-dependent rendering. This provides faster data transfer rates to the user and allows browsing data at smaller scales without sacrificing detail at larger scales.

The latest GIS technology was needed to create, store, display, and distribute the seamless data. The key commercial off-the-shelf tools used in the creation of SDDS include Oracle database software and Environmental Systems Research Institute's (ESRI) Spatial Data Engine (SDE) and Internet Map Service (IMS) software tools. The SDDS uses the latest versions of SDE and IMS to manipulate the large seamless raster datasets.

The ESRI technology was obtained by the USGS team under a Cooperative Research and Development Agreement (CRADA). The CRADA enables the USGS and ESRI to collaborate on developing seamless data technology, data analysis capabilities, and resource requirements. The Seamless Development Team leveraged the CRADA experience to create the production data delivery system.

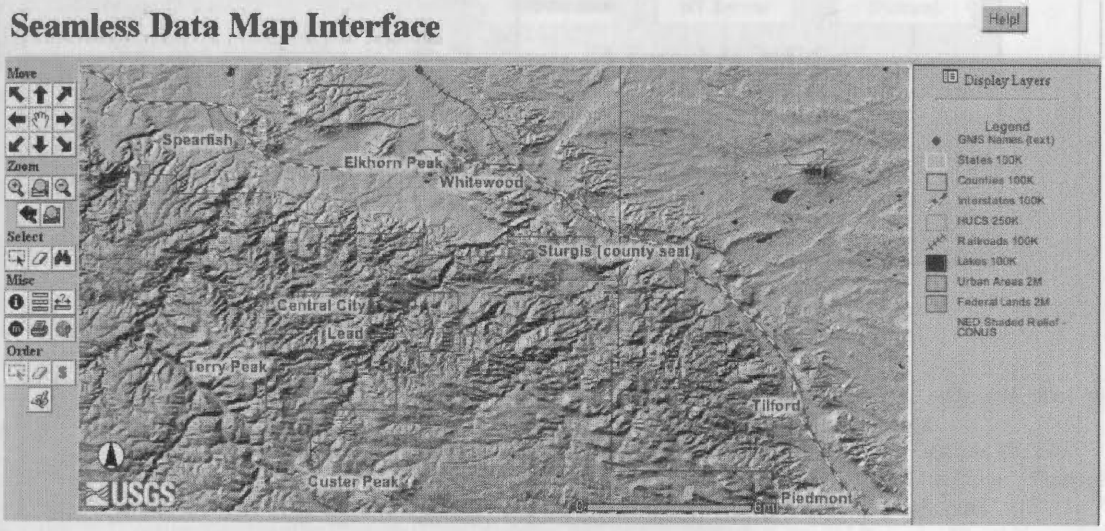


Figure 5. Screen capture of the Seamless Data Distribution System

4. WEB MAPPING PORTAL

Experiences gained were also used to publish GIS-ready information over the Internet, significantly cutting the cycle time from design to a production release. Concurrently with the portal development, ArcIMS map services were created for seamless data layers, updating the metadata, creating Web Mapping Server Specification (WMS) connectors, publishing datasets on the Geography Network, and establishing a mirror site maintained by ESRI. The USGS is working to extend the Internet interfaces for map and data delivery and build the USGS infrastructure for future data access.

The USGS EROS Data Center (EDC) provides access to national GIS-ready information through map services published on the Web Mapping Portal. The portal offers multiparticipant, distributed Web mapping services consisting of the NED, NLCD, and National Hydrography Dataset (NHD). In addition, the portal hosts a condensed version of geographic names from the Geographic Names Information System (GNIS) and vector reference-layer map services from the National Atlas.

The advantage of the use of different map services over the Web is that these services can be brought together in a mapping environment without loading all the data locally. For example, the map in Figure 6 was created using USGS-published map services for shaded relief, roads, geographic names, streams, and lakes, plus NLCD data downloaded from the SDDS. The map services were added directly from the Geography Network. The NLCD layer was rendered semitransparent and was overlaid on the shaded relief to give it a sense of depth.

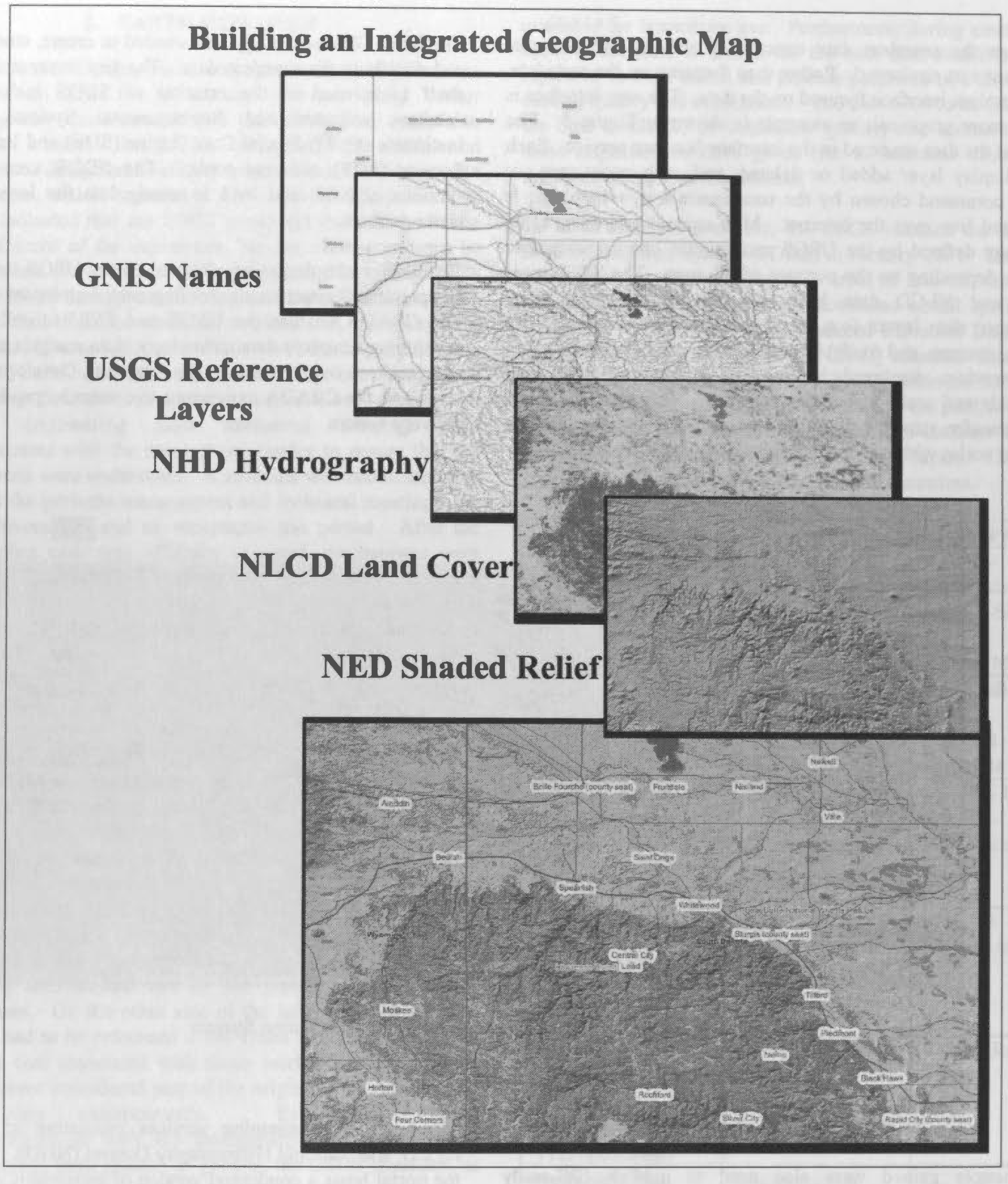

Building an Integrated Geographic Map
GNIS Names
USGS Reference Layers
NHD Hydrography
NLCD Land Cover
NED Shaded Relief
Newell
Belle Fourche (county seat)
Fruitdale
Nisland
Aladdin
Vale
Beulah
Saint Onge
Spearfish
Whitewood
Wyoming
Sturgis (county seat)
South Dakota
Central City
Lead
Tilford
Moskee
Piedmont
Nemo
Black Hawk
Rochford
Four Corners
Silver City
Rapid City (county seat)

Figure 6. Building an integrated map using data services and datasets
(Map built using ArcMap. GNIS, USGS administrative and reference layers, NHD, and NED shaded-relief layers downloaded directly from the Geography Network. NLCD ordered from the Seamless Data Distribution System and downloaded by means of FTP.)

The offerings consist of map services in ArcIMS, WMS format, and FGDC metadata. These services are accessible by using the GeoExplorer search facilities of the Geography Network site at http://www.geographynetwork.com or from the USGS site http://gisdata.usgs.net.

Also on the Web Mapping Portal are three map viewers that highlight the International Hydro 1K map services for Asia and Africa and *The National Map* Rapid Prototype datasets:

(http://gisdata.usgs.net/HYDRO1K/ASIA/viewer.htm)
(http://gisdata.usgs.net/HYDRO1K/AFRICA/viewer.htm)
http://edcw2ks18.cr.usgs.gov/website/National_Map/viewer.htm

The Web mapping services have been designed to be scalable in a multiuser environment that centers on the use of geodatabases that store vector and raster datasets, the interrelationships, and the metadata. Figure 7 shows a schematic diagram of the portal's public architecture.

When a user enters the site, the network router will direct the http Web traffic to one of the internal Web servers or to one located at an Internet Service Provider (ISP) provided by ESRI. If the USGS servers are down, requests will be directed to the ESRI mirror site. The network switch will direct the ESRI service requests to one of the ArcIMS that sends the client's request to the map servers for processing. The map servers can access and bundle maps and data into appropriate formats before returning the information to the user client either as JPEG, PNG, or GIF images. The foundation for the Web Mapping Portal is the SDE server. This is where all the data, georeferencing information, and metadata reside.

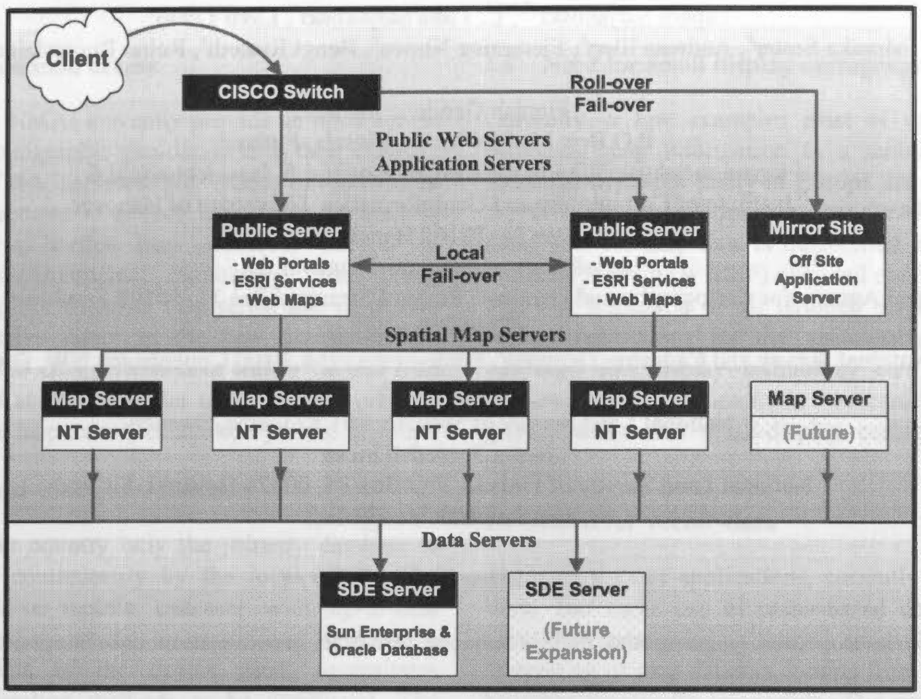

Client
CISCO Switch
Roll-over
Fail-over
Public Web Servers
Application Servers
Public Server
- Web Portals
- ESRI Services
- Web Maps
Local
Fail-over
Public Server
- Web Portals
- ESRI Services
- Web Maps
Mirror Site
Off Site
Application
Server
Spatial Map Servers
Map Server
NT Server
Map Server
NT Server
Map Server
NT Server
Map Server
NT Server
Map Server
(Future)
Data Servers
SDE Server
Sun Enterprise &
Oracle Database
SDE Server
(Future
Expansion)

Figure 7. Web Mapping Portal public system architecture

5. SUMMARY

Access to the Nation's spatial data holdings is the foundational requirement on which the three Web-based interfaces described in this paper were built. Adding to this challenge is the fact that the USGS spatial data holdings continue to grow and evolve. The USGS archive at EDC manages more than 750 terabytes of satellite, aerial, and cartographic data. The Landsat 7 program alone acquires approximately 300 new scenes per day and coordinates activity across a network of 12 international ground stations.[i]

The USGS has a long history of providing spatial data access to the general public. In the last two decades, the task of customer order and inquiry has evolved from mail, telephone, and fax to Web-based interfaces. Furthermore, product delivery has evolved from physical media such as paper, computer tape, or CD to customers taking delivery of products through the Internet. The three Web-based interfaces are important steps forward for the USGS. By developing these interfaces using advanced Web technology, the USGS is now able to serve its customers better, has increased staff experience and skills with these tools, and has prepared the USGS for *The National Map* requirements. *The National Map* is envisioned to be next step for the USGS in providing Web based spatial data access.

6. END NOTES

1 Raytheon STX Corporation, under Government contract, No. 1434-CR-97-CN-40274. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

i Land Remote Sensing Policy Act, October 28, 1992

ii USGS, National Mapping Discipline Strategic Plan, http://mapping.usgs.gov/misc/strategic.html

iii EROS Data Center Annual Report – Fiscal Year 2000

GEOSPATIAL INFO-MOBILITY SERVICES – A CHALLENGE FOR NATIONAL MAPPING AGENCIES

Tapani Sarjakoski[a*], L. Tiina Sarjakoski[a], Lassi Lehto[a]
Monika Sester[b], Andreas Illert[c], Flemming Nissen[d], Bengt Rystedt[e], Reino Ruotsalainen[f]

[a] Finnish Geodetic Institute
P.O.Box 15, FIN-02431 Masala, Finland
tapani.sarjakoski@fgi.fi, tiina.sarjakoski@fgi.fi, lassi.lehto@fgi.fi
[b] Institute of Cartography and Geoinformatics, University of Hanover
Appelstrasse 9a, 30167 Hanover, Germany
monika.sester@ikg.uni-hannover.de
[c] Federal Agency for Cartography and Geodesy, Richard-Strauss-Allee 11, 60598 Frankfurt, Germany
Illert@ifag.de
[d] National Survey and Cadastre- Denmark, Rentemestervej 8, 2400 Copenhagen NW, Denmark
fln@kms.dk
[e] National Land Survey of Sweden, 801 82 Gävle, Sweden
Bengt.Rystedt@lm.se
[f] National Land Survey of Finland, P.O.Box 84, 00521 Helsinki, Finland
reino.ruotsalainen@nls.fi

Commission IV, WG IV/4

KEY WORDS: interoperability, geospatial, mobile services, real-time, generalisation, data integration

ABSTRACT:

The aim of this paper is to describe the background and the objectives of an European Commission funded project GiMoDig (Geospatial info-mobility service by real-time data-integration and generalisation), which was recently started. The background of the project lies in the goal to improve the accessibility of the national primary topographic databases, especially for mobile users. A prototype system providing geo-spatial service will be implemented and used as a test-bed in the frame of the project. The common access interface is based on the emerging standards like XML and Open GIS Consortium's specifications. Besides methods for transferring vector-formatted data based on standards like GML and SVG, the work comprises several important issues like development of methods for real-time generalisation and data integration, definition of global schema, thematic harmonisation of geodata, development of methods for data transformation to a EUREF co-ordinate system, a study on small-display cartography and identification of the most important use cases for topographic data in mobile environment.

1. INTRODUCTION

1.1 Background

Geospatial information is vital for many mobile applications, like personal navigation, tracking of vehicles, shipment, and search of services of varying nature. Although many different spatial datasets might be used in services involving a mobile user, the national topographic map plays a key role in applications requiring detailed geospatial information with an extensive coverage. In Europe the National Mapping Agencies (NMAs) are responsible for maintaining the primary topographic databases, covering the whole country. Several smaller scale map products are derived from these databases by applying cartographic generalisation, e.g. (Jakobsson and Salo-Merta, 2001; Just, 2001; Illert, 2001; Kilpeläinen, 1997). Therefore efficient methods for generalisation are essential for versatile production of highly varying set of map products from geo-databases, and for presenting results obtained in geospatial applications on a mobile platform.

1.2 Need to harmonise spatial data

National primary geo-databases are designed mainly national tasks in mind, without considering international co-operation and demands set by the mobile use scenarios. The semantic content of these databases varies from country to country. To create maps covering several countries, harmonisation of these datasets is often carried out as a separate, time consuming conversion task. For a mobile user, real-time harmonisation of geospatial data to a known semantic frame is indispensable, because up-to-date geospatial data is needed especially in an unknown, foreign environment. For the traditional map production data integration can also be seen as a possibility to update generalised spatial data (Illert, 2001). Harmonising national primary databases lays foundation to the possibility to integrate cross-border data – a necessity for a mobile user.

1.3 Need for common access

A few European NMAs currently provide an open access to the primary topographic geo-database to their citizens. The existing NMA services are based on individual technical architectures – a fact effectively precluding international cross-border map services, needed in activities like environmental monitoring, from being introduced. Pre-prepared raster map series also constitute a serious restrictive factor in the new demand-driven approach to spatial data provision in which the end user or a value-added service provider is seen as the primary actor defining the map extent, contents, style and layout.

1.4 Need for real-time generalisation

In each European country only the primary database is usually updated continuously by the local NMA. The need to provide a mobile end-user with up-to-date information, often in a drastically generalised form due to display constraints of the device used, necessitates efficient generalisation methods to be developed. The generalisation methods related to the production of *ad hoc*, personalised maps in a real-time environment have not yet been extensively investigated (Reichenbacher, 2001; Kraak and Brown, 2001).

However, the future user of geospatial data services will expect means for defining, by himself, the contents, coverage, scale, and visual appearance of the product when data are accessed via a network by a mobile device. It can also be assumed that network-based approaches for organising the maintenance and distribution of data will be further developed and also become everyday practice in different areas of life. These trends necessitate the development of more advanced methods for generalisation of geodata, such as real-time generalisation (Lehto and Kilpeläinen, 2000; 2001; van Oosterom and Schenkelaars, 1995; Jones et al., 2000). When using real-time generalisation, the datasets are not stored in the database but rather computed during the data request – response dialogue.

1.5 Need for common coordinate system

Modern, satellite based positioning systems like Global Positioning Systems (GPS) make use of the internationally defined geodetic reference frame WGS'84. As a consequence, many European countries have started to modernise their geodetic coordinate systems to be based on EUREF that can be called as the European implementation of WGS'84. However, most of the national primary geo-databases are still maintained in national geodetic co-ordinate systems and nationally defined map projections. To be used in positioning and navigation applications for a mobile user, it is important to transform the data of national primary spatial databases to a common EUREF-based coordinate system in real time.

1.6 Need for small display cartography

Currently, a few examples exist of operative services providing map information to a mobile terminal. The solutions available today in Europe are mostly based on the Wireless Application Protocol (WAP)-technology. In these services map data is transmitted to the terminal as Wireless Bitmap (WBMP)-encoded one-bit raster images. The low resolution of the terminal display naturally sets serious restrictions for the visualisation of map data. Although the display technology can be expected to improve rapidly, intensive research on real-time generalisation and small-display cartography is clearly needed.

1.7 Need for vector-data

There are lots of applications, currently available on the Web, that make use of raster-based delivery of spatial data. The approach has several serious limitations. The integration of map datasets coming from different sources becomes practically impossible. The provision of highly customised screen maps is difficult, because map layout and content are usually fixed in the initial configuration of the service. The zooming of the map on the terminal can be accomplished by retrieving a more detailed raster map on each zoom level from the server. Vector-based map information, once downloaded to the mobile terminal, could be locally zoomed without any loss in the graphical quality. The display style of the map could be locally modified according to personal preferences of the user, and additional information could be easily requested from another service and displayed, overlaid on top of the previously downloaded vector map. The use of vector-formatted spatial information on the Web, specifically by applying Extensible Markup Language (XML)-based encodings, is currently under intensive research and development.

2. THE SCOPE OF GIMODIG

2.1 Project resources

The paper introduces an European Commission-funded project, GiMoDig, Geospatial info-mobility service by real-time data-integration and generalisation, (GiMoDig, 2002). This development and research project is a part of the European Union's Information Society Technologies (IST) – 5th framework programme, the theme "Ubiquitous and intelligent info-mobility and geo-information systems". The Finnish Geodetic Institute acts as the co-ordinator for the project and the other partners are: the University of Hanover (Institute of Cartography and

Geoinformatics); Federal Agency for Cartography and Geodesy (Germany); National Survey and Cadastre of Denmark; National Land Survey of Sweden, and National Land Survey of Finland. The project started in November 2001 and its duration is 3 years.

2.2 Main goals

The purpose of the GiMoDig project is to develop methods for delivering geospatial data to a mobile user by means of real-time data-integration and generalisation. The project aims to a seamless data service infrastructure providing access, through a common interface, to topographic geo-databases maintained by the National Mapping Agencies in various countries. A special emphasis will be put on providing appropriately generalised map data to the user depending on a mobile terminal with limited display capabilities.

2.3 Sub-objectives

Sub-objectives of the project are listed as follows:

- Development of methods and usage practices for generalising the graphic representation of geospatial data in real-time, to be suited for display of the data in varying scales on small, mobile devices with different display resolutions.
- Investigating the problems between national primary geospatial databases, often mutually heterogeneous in thematic definitions, and developing means for real-time harmonisation of data.
- Analysis of mobile use cases to adapt real-time generalisation and harmonisation of geospatial data to the requirements of a user in varying usage situations.
- Development of methods for real-time transformation of spatial data from different national geo-databases to a common, EUREF-based co-ordinate system.
- Investigating and developing methods for transferring vector-formatted spatial data to a mobile user using emerging standards, like Geography Markup Language (GML) (GML, 2002) and Scalable Vector Graphics (SVG) (World Wide Web Consortium, 2002) and testing the applicability of the standards for Web-based spatial services in a international pilot project involving national primary geodata sets.
- Development and implementation of a prototype system that can be used as a test-bed for the developed methods.

2.4 Scientific goals

The scientific aims of the project lay on the development of methods for real-time generalisation of spatial data – a field of science so far not extensively researched, but becoming increasingly critical in the modern networked society where up-to-date databases are to be directly accessed by consumers using a widely varying set of terminal devices. The spatial datasets in the GiMoDig project will be served in an XML-based vector format to facilitate flexible data integration and processing by and also support the traditional Web browsing. The pilot will enable a seamless access to the NMAs' data resources from a set of different mobile devices, and from a traditional PC as well. Real-time generalisation of the map data will be performed either on the data providers' server, on a middleware service platform or on the client device. In the GiMoDig system architecture spatial data is maintained in the NMA's server and provided to the network through a standardised access interface. Various value-added services can then be built as layered middleware solutions. The communication between actors in various levels of the system hierarchy are to be based on the emerging spatial Web-standards, especially those developed by the Open GIS Consortium (OGC), (OGC, 2002). The encoding of spatial data being transferred across the network from database level to the middle-layer service providers is going to be based on the XML-technology. Therefore, a specific goal in the research on generalisation is to employ XML-based computing technologies in the process.

3. SYSTEM ARCHITECTURE

In Figure 1 a schematic model of the GiMoDig system architecture for the aimed testbed prototype service is shown. The design architecture should facilitate the system to serve a wide array of mobile end user devices with varying characteristics.

The main points to be taken into consideration in system development are:

- the need to integrate spatial data from heterogeneous sources into one seamless service
- the need to perform real-time generalisation of data
- the wide variation in display characteristics of the end-user devices on which the final data visualisation will be carried out
- the aim to satisfy the highly individual needs of a mobile user by providing ad hoc, personalised maps

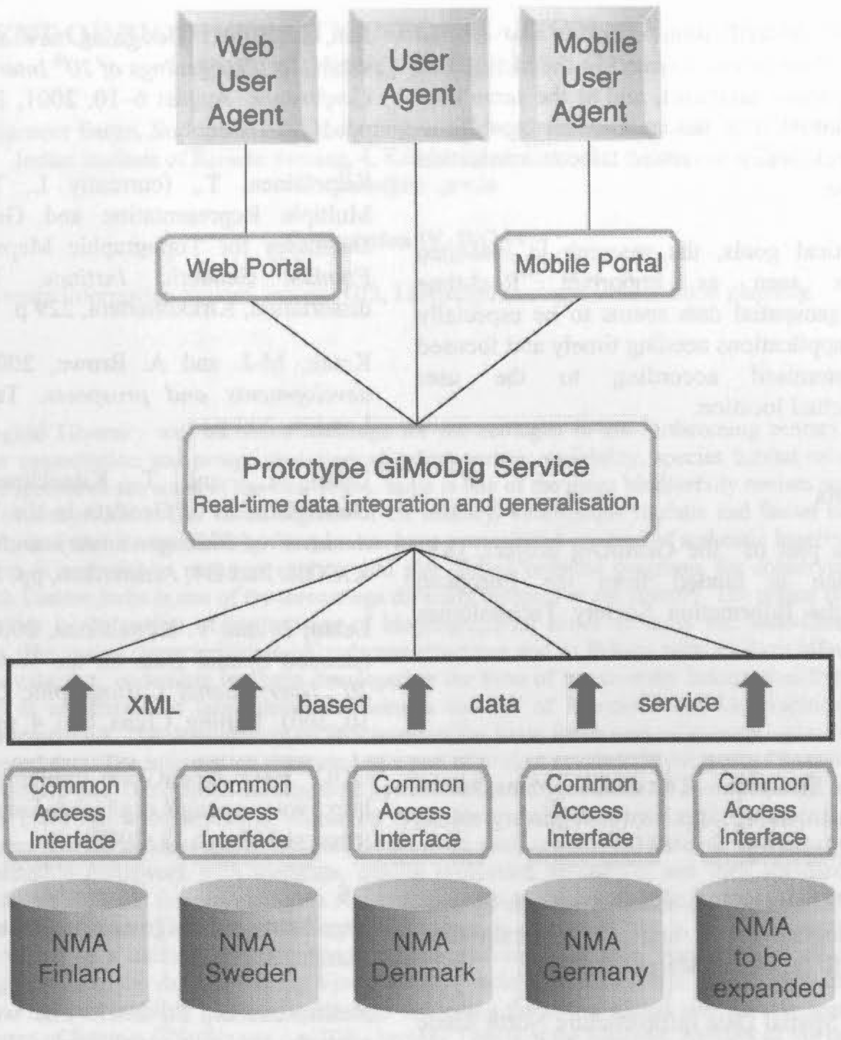

Web User Agent
User Agent
Mobile User Agent
Web Portal
Mobile Portal
Prototype GiMoDig Service
Real-time data integration and generalisation
XML based data service
Common Access Interface
Common Access Interface
Common Access Interface
Common Access Interface
Common Access Interface
NMA Finland
NMA Sweden
NMA Denmark
NMA Germany
NMA to be expanded

Figure 1. The system architecture of the GiMoDig test bed.

The system should be scalable to support a potentially extensive user community, given the importance of the spatial datasets involved. The architecture should be modular enough to facilitate a reasonable amount of the harmonising work to be done locally by each NMA participating in the project. The system architecture should also support internationally developed interface standards to achieve a maximum possible benefit on the global mobile geospatial applications market place.

4. DISCUSSION AND CONCLUSIONS

GiMoDig is a research and development project with the goal to improve accessibility to and interoperability between national topographic databases. On European level there are other projects, like the already completed "European Territorial Management Information Infrastructure"-project (ETeMII, 2002) and the just recently initiated "Infrastructure for Spatial Information in Europe"-project (INSPIRE, 2002) having a parallel goal but with a broader and more policy oriented approach.

On national and regional level many mapping organisations have already taken steps towards for establishing web-based access to their databases, e.g. the Ordnance Survey's MasterMap (OS MasterMap, 2002) and Spatial Data Infrastructure North Rhine Westphalia (GDI NRW, 2002). The NMAs in the GiMoDig Consortium have also initiated similar projects. One of the motivations for these projects is the NMAs' need to deliver the most updated data to their customers. A clear tendency is that these services are built on standard technology, especially using XML and GML.

The main practical goal of the GiMoDig project is to design and implement a prototype of a cross-border spatial data service providing access through a common interface, conformant with international spatial Web standards, into the primary national topographic databases. The provision of a standards-compliant service will promote creation of third-party, value-added mobile information services needed in the position-dependent applications, like traffic guidance, rescue operation management and personal navigation. The GiMoDig project thus aims at facilitating efficient use of the extensive amount of resources invested in the creation of nation-wide topographic databases, and at the same time advances the exploitation of the multimedia-capabilities available in the technically advanced telecommunication networks in Europe.

Besides the practical goals, the research in real-time generalisation is seen as important. Real-time generalisation of geospatial data seems to be especially critical in mobile applications needing timely and focused information, customised according to the user preferencies and actual location.

Acknowledgements

This research is a part of the GiMoDig project, IST-2000-30090, which is funded from the European Commission via the Information Society Technologies (IST) programme.

References:

ETeMII, 2002. European Territorial Management Information Infrastructure, http://www.ec-gis.org/etemii (accessed March 22, 2002).

GiMoDig, 2002. Geospatial info-mobility service by real-time data-integration and generalisation, http://gimodig.fgi.fi/ (accessed March 22, 2002).

GDI NRW, 2002. Spatial Data Infrastructure North Rhine Westphalia GDI NRW, http://gdi-nrw.uni-muenster.de/index_e.html (accessed March 22, 2002).

GML, 2002. Geography Markup Language (GML) 2.0 OpenGIS® Implementation Specification, 20 February 2001, http://www.opengis.net/gml/01-029/GML2.html (accessed March 22, 2002).

Illert, A., 2001. Multi-Scale GIS: From Generalisation to Data Integration. In: *Proceedings of 20th International Cartographic Conference,* August 6–10, 2001, Beijing, China, Vol. 3, pp. 1690-1699.

Jakobsson, A. and L. Salo-Merta, 2001. Definition of a Basic Topographic framework for National GI Polocy – One Database for All Basic Topographic Data. In: *Proceedings of 20th International Cartographic Conference,* August 6–10, 2001, Beijing, China, Vol. 4, pp. 2197-2205.

Jones, C. B., A. I. Abdelmoty, M. E. Lonergan, P. van der Poorten, and S. Zhou, 2000. Multi-Scale Spatial Database Design for Online Generalisation. In: *Proceedings of the 9th Spatial Data Handling Symposium,* Beijing, pp. 7b.34-7b.44.

Just, L., 2001. Designing new maps adapted to user needs. In: *Proceedings of 20th International Cartographic Conference,* August 6–10, 2001, Beijing, China, Vol. 2, pp. 1056-1062.

Kilpeläinen, T., (currently L. T. Sarjakoski), 1997. Multiple Representation and Generalization of Geo-Databases for Topographic Maps. *Publications of the Finnish Geodetic Institute,* No. 124, Doctoral dissertation, Kirkkonummi, 229 p.

Kraak, M-J. and A. Brown, 2001. *Web Cartography, developments and prospects.* Taylor & Francis Inc, London, 209 p.

Lehto, L., and T. Kilpeläinen, 2000. Real-Time Generalization of Geodata in the Web. In: *International Archives of Photogrammetry and Remote Sensing,* Vol. XXXIII, Part B4, Amsterdam, pp. 559-566.

Lehto, L. and T. Kilpeläinen, 2001. Generalizing XML-encoded Spatial Data on the Web. In: *Proceedings of 20th International Cartographic Conference,* August 6–10, 2001, Beijing, China, Vol. 4, pp. 2390–2396.

OGC, 2002. OpenGIS® Implementation Specifications, http://www.opengis.org/techno/implementation.htm (accessed March 13, 2002).

OS MasterMap, 2002. http://www.ordsvy.gov.uk/os_mastermap/ (accessed March 22, 2002).

Reichenbacher, T., 2001. The world in your pocket – towards a mobile cartography. In: *Proceedings of 20th International Cartographic Conference,* August 6–10, 2001, Beijing, China, Vol. 4, pp. 2514-2521.

van Oosterom, P., and V. Schenkelaars, 1995. The Development of an Interactive Multi-Scale GIS. In: *International Journal of Geographical Information Systems,* Vol. 9, No. 5, pp. 489-507.

World Wide Web Consortium, 2002. Scalable Vector Graphics (SVG), http://www.w3.org/Graphics/SVG/ (accessed March 8, 2002).

DEVELOPMENT OF BIODIVERSITY INFORMATION SYSTEM FOR NORTH EAST INDIA USING INTERNET GIS

P. S. Roy, Sameer Saran, Suddhasheel Ghosh, Nupoor Prasad, Harish Karnatak and Gautam Talukdar
Indian Institute of Remote Sensing, 4, Kalidas Road, Dehradun (Uttranchal), INDIA
dean@iirs.gov.in

Commission IV, WG IV/4

KEYWORDS: Biodiversity Information System, Internet GIS, Landscape Ecology, Conservation planning

ABSTRACT

Conservation of Biological Diversity will be major challenge for the ecologist in the forthcoming century. In-situ conservation, biotechnology tools for conservation and prospecting, understanding genetic variability, species habitat relationship and allowing evolutionary process of speciation are some of the challenges. India is one of the mega biodiversity centers and is also known for its traditional knowledge of conservation. The varied regions of the country, with unique floristic and faunal richness, their vastness, endemism, heterogeneity and also inaccessibility of large areas have necessitated creation of authentic baseline data on biodiversity. This information system is essential to monitor, analyze and plan action oriented programs for conserving and preserving our biological wealth. North Eastern India is one of the three mega diversity hotspots in the country. The region is referred as a cradle of flowering plants as it lies in the region of conjunction of biogeographical zones of India viz. Indo-China, Indo-Malayan and Gondwana land masses. The spatial characterization of landscape structures and its linkage with attribute information on the floristic composition, economic valuation, endemism has been developed in the form of Biodiversity Information System (BIS) on sharable environment. The BIS is integration of large databases using a concept of Internet based Geographical Information System commonly known as Internet GIS. The development of BIS involves the basic framework of concept, selection, and aggregation of fundamental and processed data. The information generated as a part of project entitled "Biodiversity Characterization at Landscape Level using Remote Sensing and Geographic Information System" of the Department of Biotechnology and the Department of Space, Government of India has information on vegetation type/ land use, disturbance regimes, biological richness, ancillary information from different sources and non spatial attribute database on plant species (field sample and flora). The database has been organized under the sharable framework with metadata, quality evaluation procedures and their standardization. The different database structure created on different formats/platforms comes as an input to the information system for standardization that is they are converted into the same format. The non-spatial species database is converted into a relational database structure and the geospatial data is converted to a uniform spatial projection. The conversion of data from the various available formats also constituted reverse engineering of the databases using Visual Basic, Oracle and the database format software and Open Data Base Connectivity (ODBC). The web enabling part comes through uploading the entire spatial and the non-spatial data at a common platform using the concept of Internet GIS. We use ArcSDE, ArcInfo, Oracle at the backend, ArcIMS as the Internet Map Server and Microsoft Internet Information Server 4.0 as the web server.

1. INTRODUCTION

Since the late nineties the world has been experiencing a great boom in information dissemination and communication technologies. This recent boom in the technological capabilities has enabled humankind to bring education and information right at its doorstep. In a scenario where ecological imbalance is at its heights and forest covers are depleting at an exponential rate day by day, important studies to conserve such resources should not remain restricted only to paper maps and higher authorities. In order to implement suggested strategies under a study it is very important to use the current information and communication technologies to its best. It is therefore felt that the results of such studies should be projected onto the Internet wherein monitoring of such resources and respective conservation issues be handled.

Acquisition of images of earth from space has opened new frontiers in mapping. The multispectral satellite images provide definitions of vegetation patches, which are related to phonological types, gregarious formations and communities occurring in unique environmental setup. The temporal images help in monitoring all back processes a landscape has experienced (Delcourt and Delcourt, 1988). The images also provide digital mosaic of spatial landcover and vegetation types amenable to computer processing (Coulson *et al.* 1990 and Chuvieco, 1999). Biophysical spectral modeling techniques allow to stratify vegetation types based on the canopy closure (Roy *et al.* 1996). Such an approach allows mapping and monitoring the forest condition and degradation processes.

2. WHY A BIODIVERSITY INFORMATION SYSTEM?

Biological diversity is a part of our daily life and livelihood for it constitutes resources upon which humanity depends. From time immemorial man and animal have been using this diversity in harmony for their sustenance, but this harmony is getting disturbed owing to natural as well as human action of late. This leads to loss or extinction of plants and animals. This again may be either due to the inability of the species to evolve and adapt to the newer conditions or falling prey to the human greed. Sustainable management of natural resources has become a key issue for survival of planet earth. This is an urgent need to conserve gene pool *in situ* before it is lost forever because a large percentage of biological wealth and its importance is still unknown to us.

The north east India – a bowl of plant diversity is a part of the biodiversity *hot spot* in Eastern Himalayas. The latter is the meeting point of temperate east Himalayan flora, palaeo – arctic of Tibetan high land and wet evergreen flora of south-east Asia and Yunan. The region is well known for its unique flora and fauna, high endemism and rare plants. At the same time, the region is experiencing alarming rate of deforestation and degradation, causing threat to the world's well known *hot spot*. Conservation of bioresources has become essential in view of the loss of species, both flora and fauna. There has been loss biodiversity due to shrinkage and/or loss of habitats and fragmentation of forests.

The goals and scales of inventorying and monitoring programmes may change with time. Hence the baseline data at landscape level should be sufficiently robust to accommodate changes. It should be based on robust samples enabling calibration for future Rapid Biodiversity Assessment. Landscapes contain all levels of the biological hierarchy, from ecosystem to species and genes that are targeted from diversity inventories and conservation. The project titled 'Biodiversity Characterization at Landscape Level using satellite remote sensing and GIS' first of its kind at national level. Though attempts have been to identify the gaps in biodiversity management, this is the first attempt to quantify species richness, biodiversity valuation and phytosociological analysis and use these themes to characterize biodiversity in conjunction with other ecological parameters. In this project methodology has been developed to analyze and prioritize the different ecosystems for conservation by using the existing knowledge and assessing the disturbance impacts/regimes by way of investigating and inventorying biological richness of the areas. Once biologically rich sites/areas are identifies then detailed study can be taken up for possible potential areas for bioprospecting. The present effort to characterize vegetation cover, fragmentation, disturbance and biological richness across the landscape is organized in the form of Biodiversity Information System. The field samples of key ecological characters have been used for geospatial extrapolation. The species data base has been linked with above spatial details. The Biodiversity Information System allows the user to identify gap areas, species/habitat relationship and helps in biodiversity conservation planning by setting priority areas. Such database coupled with detailed site specific field inventories helps in identifying areas for bioprospecting.

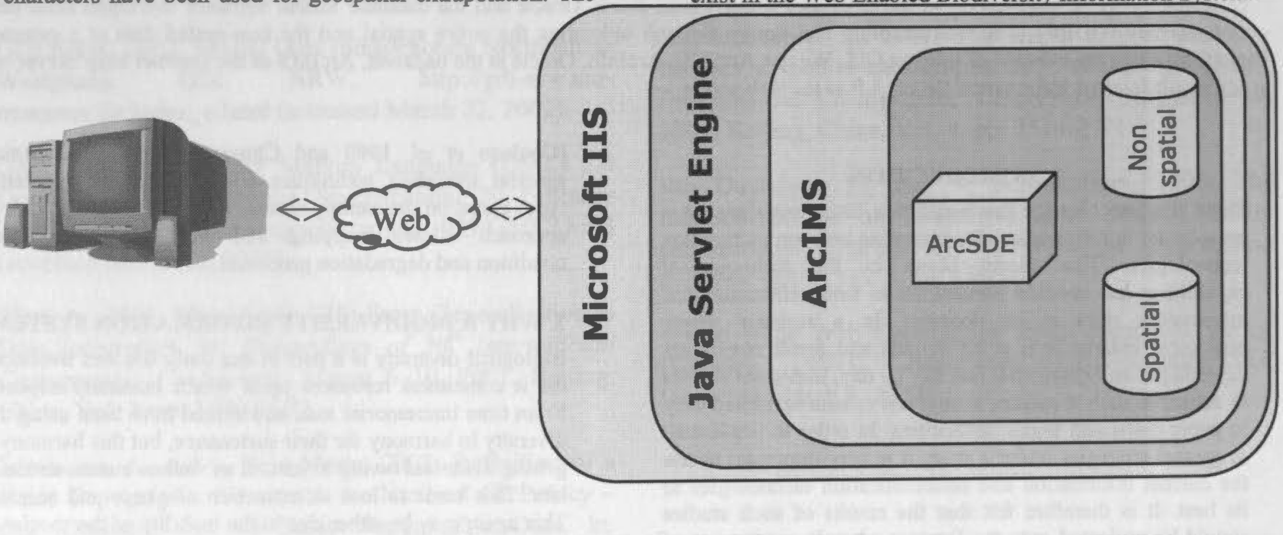


Fig 1: Architecture of the Biodiversity Information System

The Department of Biotechnology (DBT) and Department of Space (DOS) have joined hands to undertake study on biodiversity characterization. The main objectives of this project are to identify the biorich areas, to evaluate the forest types for their value, to provide locational information of the economically important species for bioprospecting. The outcome of this project will assist the forest managers, decision makers and the national lead institutes involved in the genetic diversity and bioprospecting.

This multi institutional project is being participated by various lead organizations like Botanical Survey of India, Forest Survey of India, Wildlife Institute of India, G. B. Pant Institute of Himalayan Environment and Development, French Institute, Pondicherry and various specialized remote sensing centers in Department of Space.

In this paper we present an approach to organize spatial and non spatial data into a Biodiversity Information System (BIS) for the north east India in webenabled environment. The spatial information on biodiversity character of landscape have been developed using multicriteria analysis in GIS and facilitates the following:

- Rapid Assessment of monitoring biodiversity loss and/or gain;
- Assess nature of habitat and disturbance regimes;
- Evolve species—habitat relationship;
- Mapping biological richness and gap analysis; and
- Prioritizing conservation and bioprospecting sites

3. ARCHITECTURE OF THE WEB ENABLED BIODIVERSITY INFORMATION SYSTEM

As given in the diagram, the following software components exist in the Web Enabled Biodiversity Information System.

3.1 Operating System

Operating system is the basic requirement for any system to function. In our case we chose for the Microsoft Windows NT Server 4.0 as a standard server.

3.2 Web Server

In order to serve pages to the net a basic requirement is a web server. As we opted for a Microsoft designed operating system, we also chose to use Microsoft's Internet Information Server 4.0

3.3 Java Servlet Engine

We planned to display dynamic and interactive maps through the software ArcIMS. It is a prerequisite for ArcIMS that a Java Servlet Engine be installed on the system in addition to a web server. As per ESRI's recommendations, we installed New Atlanta ServeletExec ISAPI 4.1 as the Java Servlet Engine.

3.4 Internet GIS Software

In order to display interactive maps through the net, we require a map server. We chose to use ESRI's Arc Internet Map Server (ArcIMS). As mentioned earlier, ArcIMS runs on a Java Servlet Engine and the web server.

3.5 GIS Software

ArcGIS 8.1 was installed on the server so as to import the entire spatial database generated.

3.6 Spatial Database Engine

All the imported maps were now to be directed to ArcIMS for which the Arc Spatial Database Engine 8.2 was required. The Arc Spatial Database Engine (ArcSDE) support various RDBMS platforms like DB2, SQL Server, Oracle 8, Oracle 9i etc. It is to be noted that the previous version of ArcSDE that is ArcSDE 8.1 did not have support for Oracle 9i and ESRI released the new version with support for Oracle 9i on May 9, 2002.

3.7 Non spatial database engine

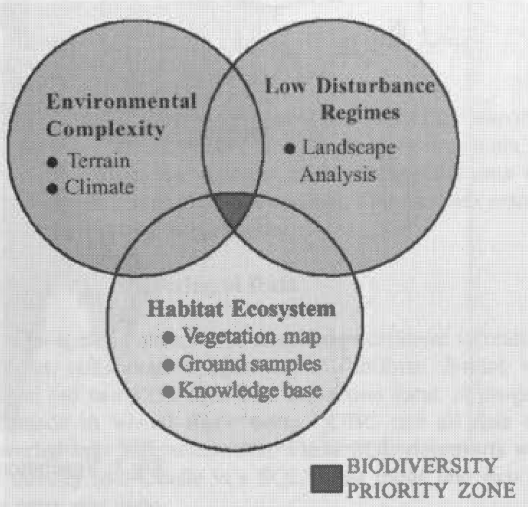
The entire non-spatial database which consisted of data of sample points, species database, tables regarding the economic, medicinal and ecological importance. In addition to these, tables containing ecological parameters like Shannon Weiner indices, Similarity indices etc were also added.

The web server hardware taken was a high end server. In addition to this a 256 kbps leased line Internet connectivity was taken from Videsh Sanchar Nigam Limited (VSNL). Leased line connectivity is required to handle the high Internet traffic and to quickly serve pages and results to the net.

4. Data

The data which comes as an input to this information system is the result of all the work which had been carried out at Phase I of the DOS-DBT project (Anonymous, 2002). The mapping task has been done by Indian Institute of Remote Sensing (Arunachal, Assam and Meghalaya), Regional Remote Sensing Service Centre, Dehradun (Mizoram and Tripura) and Regional Remote Sensing Service Centre, Kharagpur (Manipur, Nagaland and Sikkim).

Thematic spatial data are generated using satellite remote sensing data through digital classification. In addition to that spatial and non-spatial data from other ancillary data sources is collected. Database is created/generated in a GIS domain and analysed Finally biological richness has been assessed using ground based data on ecosystem uniqueness, species richness, biodiversity value, terrain complexity, diversity index etc. (Roy and Tomar, 2000)



4.1 Field Data Collection

Stratified random sampling method has been used to do the sampling in all the vegetation types. For trees 20m x 20m / 32 m x 32 m, for shrubs 5m x 5m or one 10m x 10m and for herbs 1m x 1m sample plots were laid. In all more than 1750 sample plots have been laid in various types of vegetation. For shrubs and herbs nested sampling approach was followed. Geographical position of these points has been recorded. All the plant species and life forms occurring in the plot were enumerated. For test areas classification accuracy has been tested and over all it is more than 85%.

4.2 Vegetation Map

The North Eastern region is very rich and varied and fosters vegetation from tropical forests in Assam to Alpine in Arunachal Pradesh and Sikkim. About 45 land cover and land use classes have been delineated using satellite data. And for the sake of uniformity Level II vegetation classification was attempted. The task of classification has been accomplishes by doing reconnaissance survey and collecting ground truth using False Color Composite (FCC) on 1:250,000 in combination with 1:50,000 scale where ever required.

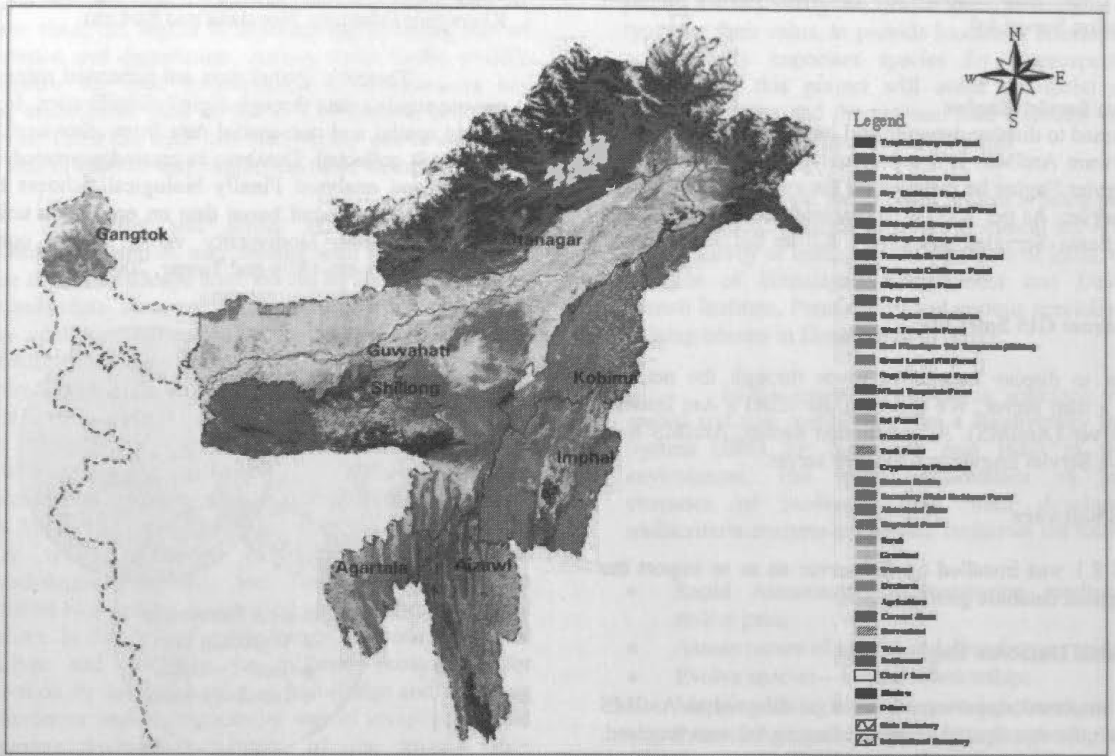


Fig 2: Vegetation Type Map of North East India

4.3 Fragmentation Map

Fragmentation is the number of patches of forest and non – forest type per unit area. The forest type map was reclassified into two patches, viz. forest and non-forest resulting in a new spatial data layer. This layer was convolved with a grid cell of n x n (e.g. n = 500m) with a criterion of deriving the number of forest patches within the grid cell. An output layer is derived. To this a look up table is generated which keeps the normalized data of the patches per cell in the range of 0 to 10.

4.4 Disturbance Regimes

In nature various factors affect the richness of biodiversity. Topographic variations, micro-climatic conditions, biotic factors and land itself control the distribution of plant communities. Thus coexistence of plants or animals is the result of interactions. Studies have revealed that there exists a relationship among various communities forming patches and the occurrence of species. Discontinuities in the vegetation as a whole and forests in particular are the result of such disturbances and can be explained through patch characterization. Various landscape ecological parameters like fragmentation, porosity (number of patches in primary vegetation types per unit area), patchiness (number of patches of all vegetation types per unit area), interspersion (intermixing of the vegetation types) and juxtaposition (adjacency of vegetation types) and the disturbances from the road and settlements etc have been used for characterizing the disturbance. Since biotic pressure is most significant and influences the population and diversity, disturbance regimes have been established for the entire north eastern region.

The Disturbance Index Map is computed as a linear combination of the defined parameters on the basis of probabilistic weightages The final spatial data is rescaled to a range of (0-100) for final preparation.

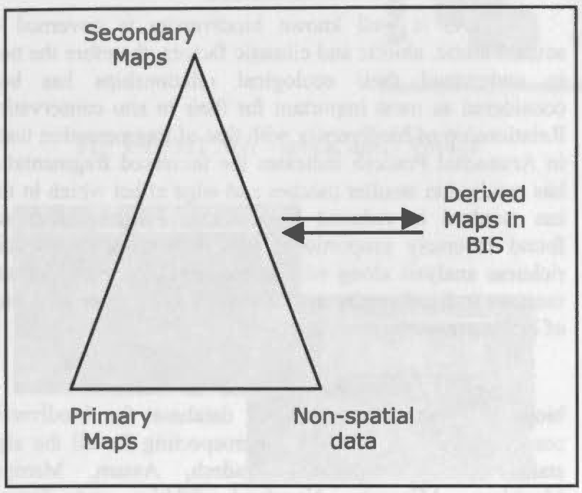
4.5 Biological Richness Estimation

North eastern region is endowed with high floristic richness and has been rightly called as the 'cradle of flowering plants'. The region has 4 micro – endemic centres out of the 26 such centres in India. More than 5000 plant species have been reported from this region. The region is also famous for orchids and ferns. About 500 species of orchids occurring in wide variety of conditions are reported from here. The region has about 132 wild relatives of crop plants out of the 686 reported from Indian sub continent. This is important for genetic engineering and bioprospecting. This information has been incorporated while characterizing the biodiversity from the field observations. More than 1750 sample plots were laid and analysed to determine the total number of species, endemic species, economic species (more than 2000) has been reported from Arunachal Pradesh. Occurrence, frequency and abundance of the species is governed by the climatic and biotic factors.

The biological richness at landscape level is determined as function of ecosystem uniqueness, species diversity, biodiversity value, and terrain complexity and disturbance index. The main parameters like ecosystem uniqueness, species richness, Biological values *etc.* come from ground based observation. Various vegetation types available in the terrain are evaluated. It is also dependent on localized information. The final integration of landscape parameters, disturbance regimes, habitat & their attributes and terrain complexity was made. The output BR spatial layer is scaled to (0 - 100) for final presentation.

The additional data sets which were used directly or indirectly were the following:

- Digital elevation models
- Digital Chart of world (DCW) for village location, road network, drainage information
- Survey of India topo-sheets on 1:250,000 and 1:50,000 scales
- Biogeographical digital maps available from Wildlife Institute of India, Dehradun (Rodgers and Panwar, 1988)
- Climatic maps of NATMO on 1:1 million and agroclimatic maps for entire country available with National Bureau of Soil and Landuse Survey, Nagpur, India
- Socio economic data of national census 1991
- Protected Area Network map available with the Wildlife Institute of India, Dehradun, India
- Forest Vegetation Maps 1:250,000 scale available with Forest survey of India
- Administrative boundaries available from Survey of India, Dehradun, India



In addition to these databases, the result of the extensive field sampling has resulted in the generation of a huge species database, which can be used as a tool to measure and monitor the biodiversity of the area.

Another species database from the Bioprospecting and Molecular Taxonomy Programme of the Department of Biotechnology has also been used. Under the Bioprospecting and Molecular Taxonomy Programme (BPMT), some environmentally / medicinally / economically important plant species are being studied and a good amount of data has been generated. It was also planned to integrate this database with the species database of the North East India in order to correlate forest types, ecological parameters and molecular level studies.

5. METHODOLOGY

5.1 Uniformity of Data

Since the project is a collaborative one, the various results received from different institutions were obtained in different formats. When it was considered that the whole data be projected in a single platform, it was decided that the entire data should be recast into a uniform format. We chose the projection given by Survey of India, Dehradun, India as a standard. We then registered all maps as per this norm.

The non-spatial data consists of a huge species database, which contains information regarding species, importance values, photographs, ecological parameters etc. A common format of the data was worked out using the field sampling form used in this project using the RDBMS concept.

The non-spatial data from the BPMT programme was organized in a format agreed upon jointly between the personnel of Department of Biotechnology and the Department of Space

5.2 Data Organization Plan

The spatial data and the non spatial data were then imported into ArcSDE, ArcIMS and Oracle 9i using various tools. As mentioned earlier, the structure of the non-spatial data was made according to an RDBMS concept. The data was entered into Oracle 9i using the methodo

5.3 Reverse Engineering of Data

The non-spatial data too was supplied in different formats by different collaborative agencies. A uniform format was worked out using the field data collection form. A program was made in Visual Basic using ODBC and all data was converted into SQL statements. These SQL statements were run directly into Oracle 9i's SQL* Plus mode and thus the data entry was done.

5.4 Metadata

Metadata information has been keyed using the ArcGIS 8.1 package. This information provides details regarding all the layers used in the Information System. The format for this data has been formulated using directives of National Spatial Data Infrastructures (NSDI)

5.5 Scripting Techniques

It was planned that access to the Biodiversity Information System should be through a user name and password. Further the users would have different levels of access to the maps. In order to handle such requirement coding through Active Server Pages (ASP) was done at server side and Javascript at client Side.

6. RESULTS

This is the first attempt in the north east region to generate baseline information on vegetation types, floristic composition of forest types and various other parameters like their value and uniqueness. In all 21 vegetation cover types have been identified and mapped using satellite data. A total of 45 land cover and land use classes have been mapped at 1:250,000 scale. Further information on the endemic, economic and ecologically important species in each of the forest types has been compiled.

Arunachal Pradesh still has original or virgin forests. Many areas like Namdapha Biosphere Reserve and adjacent areas in Chhanglang and Lohit districts, Mehao Wildlife Sanctuary, upper reaches of Debang Valley district beyond Anini on the west etc. still are botanist's paradise. Well known tropical evergreen forests of Dipterocarpus macrocarpus have been mapped in Chhanglang and Lohit districts of Arunachal Pradesh and eastern districts of Assam (Tinsukhia, North Cachar Hill etc.). Similarly North District of Sikkim specially the Kanchanjungha Biosphere Reserve, Zemu and Lhonak valleys, areas around Djongri are repository of the excellent flora. Temperate forests of Sikkim are moist and humid and support unique plants of Sino-Tibetan are well known for the Rhododendrons, Primulas, Asters etc. Alpine vegetation of Sikkim is moist and humid and support unique plants of Sino-Tebetan region. In Arunachal Pradesh two prominent subtropical evergreen vegetation types with different species composition have been mapped. More than 1750 sample plots have been laid in all the vegetation types and field sampling has been conducted. Forest cover vegetation occupies about 37.04% in the region. It is estimated that about 7.65% of the landscape is under abandoned shifting cultivation. Open bamboo brakes 3.62%. The secondary forest developed as degraded forests formations with scattered trees occupy 5.79%. Results indicate that temperate broad leaf forest occupy the highest area which is followed by subtropical evergreen forest, moist deciduous, tropical semi-evergreen and tropical evergreen forests. Tropical evergreen forests of Barak valley dominated by *Dipterocarpus turbinatus - Palaquium polyanthum* are unique in the region and indicator of high rainfall. Large areas have Pine Forests, like *Pinus insularis* in Meghalaya and Nagaland, *Pinus merkusii* in Lohit District of Arunachal Pradesh and other species Pines in Sikkim and western Arunachal Pradesh.

Assam has largest area of man-made landscape and followed by Tripura. The land usually after one or two years of cropping and the process in repeated for another new forested site. The practice of jhumming is rampant in Karbiang long and the North Cachar hill of Assam. In addition to jhum large chunk of Dipterocarpus dominated area are being converted into open degraded forest resulting in the loss of rich biodiversity. It is observer that only 22.07% of the total forested area is intact and it indicates that only 7.39% of the forest is under high to medium richness. As discussed earlier the disturbance is caused by several factors. Mizoram has largest area under disturbance because of large scale shifting cultivation spread throughout the state and re-growth of bamboo. Meghalaya is one among most developed states in the region and its plateau been subjected to intensive shifting cultivation and higher settlements and road network resulting fragmentation of habitats. Khasi and Jaintea hills are highly disturbed. Nagaland also has very high disturbance, but for some areas towards Myanmar border and Assam border remaining areas fall under very high and high categories of disturbance. Tripura has large areas under disturbance. Arunachal Pradesh and Sikkim have less area under disturbance zone. Tirap district of Arunachal Pradesh has shifting cultivation and degraded forest represent low biodiversity.

Arunachal Pradesh tops the list among the high biorich area. The state still has intact and high biodiversity region from tropical to sub-alpine zone. Tropical evergreen forest of Dipterocarpus are unique ecosystems and grow in Chhanglang and Lohit districts. Sub-tropical forests Kameng, Subansiri, Siang etc. are very rich. The state has less population as compared to all other states in the region and therefore the upper reaches harbor excellent biodiversity. Sikkim has high biodiversity throughout and reason could be the permanent terrace cultivation and protection given by the communities in the form of sacred groves. North district has varied climatic condition and is extremely valuable for bioprospecting as it has several medicinal plants. Garo hills and some areas of Jaintea hills of Meghalaya have high Biodiversity. North Cachar hills district of Assam has good forest cover and has high biorich area. Inner line reserve forest of Barak valley has some pockets of good primary forests of *Dipterocarpus turbinatus – Palaquium polyanthum*. States like Mizoram, Manipur, Nagaland and Tripura have less areas under high biorichness.

About 0.45 million families in this region annually cultivate 10,000 km^2 of forests whereas total area affected by Jhumming is believed to be 44,000 km^2. Anthropogenic factors are leading to the erosion in biodiversity and therefore there is need to study and protect before it is too late. Mapping is the first step for any resource planning and management and remote sensing technology has been providing special details about the vegetation or a landscape and provides answer to several questions asked by the resource management.

As is well known biodiversity is governed by several biotic, abiotic and climatic factors, therefore the need to understand their ecological relationships has been considered as most important for their *in situ* conservation. Relationship of biodiversity with that of fragmentation tested in Arunachal Pradesh indicates the increased fragmentation has resulted in smaller patches and edge effect which in turn has resulted in reduced biodiversity. Fragmentation was found inversely proportional to the biodiversity. Species richness analysis along with disturbance gradients indicates increase in Biodiversity as one moved away from the source of biotic pressures.

The study has resulted in characterization of biodiversity and formation of database for biodiversity conservation as well as for bioprospecting for all the eight states namely Arunachal Pradesh, Assam, Manipur, Meghlaya, Mizoram, Nagaland, Sikkim and Tripura. Database in the form of layers of land cover and land use, fragmentation, disturbance and biorich areas has been generated as a part of Biodiversity Information System for better planning management and decision making. These biodiversity rich or biorich areas have categorized into very high, high, moderate and low richness areas.

The Biodiversity Information System has been developed for faster retrieval and visualization of the information. The following results were obtained as a result of integration of ArcIMS, ArcGIS and ArcSDE. The first picture is the homepage of the Biodiversity Information System which was generated using the ASP and Javascript coding techniques.

The Biodiversity Information System comprises of three sub-information systems:

- Bioprospecting and Molecular Taxonomy Programme (BPMT)
- Species Information System (SIS)
- Forest Resource Information System

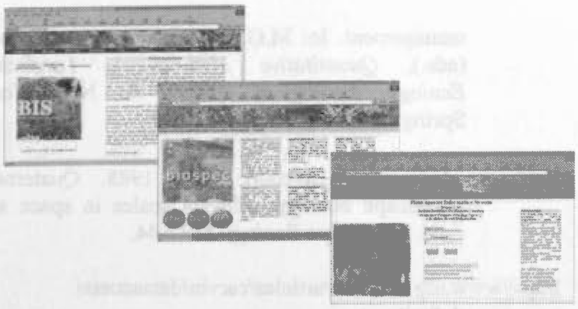


Fig: Front Page Screens of Biodiversity Information System

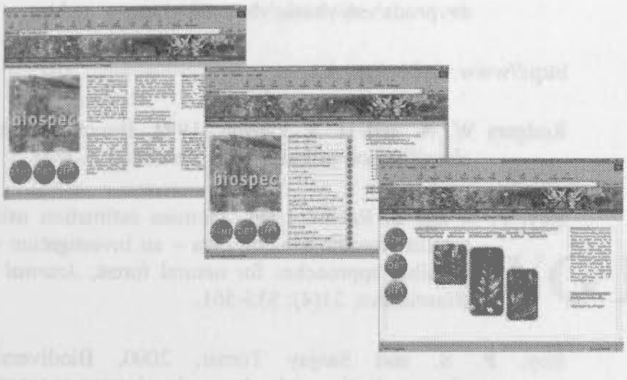


Fig: 3. BPMT Non – spatial query Results

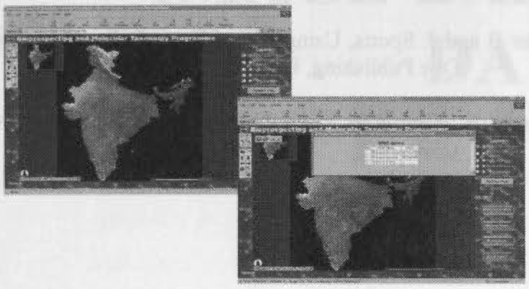


Fig 4. BPMT Spatial Query Shell using Internet

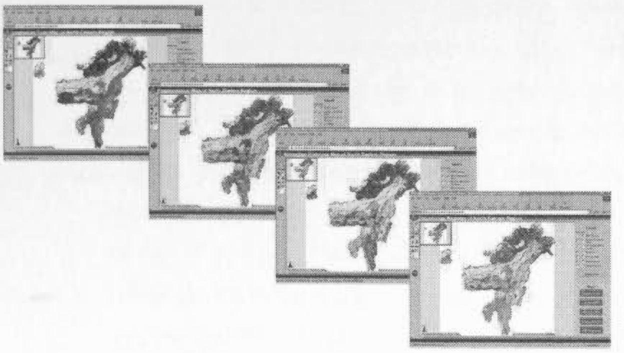


Fig 5. Displays of Vegetation, Fragmentation, Disturbace Regimes and Biological Richness Maps

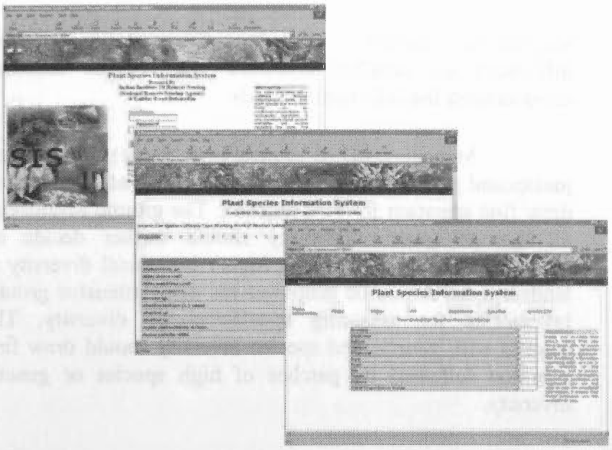

SIS
Plant Species Information System
Plant Species Information System
Plant Species Information System

Fig 6: SIS – Display of species query shell

6.1 Using the Biodiversity Information System

It was decided that the access to the Biodiversity Information System be divided into three categories. The category of users will have access to all information and updation privileges. The second category of users will have right to browse all information on the site. Further this user will be able to see interactive maps. The interactive maps in the information system have various tools like zoom-in, zoom-out, query, info, pan etc. The user can interactively query and view more layers like state boundary, district boundary and taluk boundary. The third category of users will be able to access static maps through the web and will not be able to see or query dynamic maps.

7. DISCUSSION

Biologically Rich areas are those habitats where landscape ecological conditions are favourable for natural speciation and evolutionary process. These areas can be expected to be in equilibrium where species can occur, grow and evolve in natural conditions. Biological Richness Indices (BR) map asserts the areas, which should be treated on priority in decision making and management level for conservation of biodiversity. The 'Gap Analysis' carried out on these maps will guide management and decision making for bioprospecting.

All plant species have a basic requirement of its ecological optima in particular habitat or niche within range of tolerance and requirement. Habitat identification and economic importance of the species can be useful input for bioprospecting and biodiversity conservation. Each species requires a specific ecological niche (minimum/optimum area for its survival, evolution, gene exchange). Analysis of landscape parameters like, habitat fragmentation, patchiness, interspersion and juxtaposition have shown impact on the definition of the limits in different habitats. The greater the variety of types of habitat, the greater the diversity of the species. Diversity also increases with expanding architectural complexity of the physical habitat. Most of the species growing in the natural conditions have some sociological association with the species–environment complex and in general have fairly well defined niches. Similar ecological conditions in different geographical location bear similar biodiversity if not the same. But they will have differences at genetic level. The vegetation cover types, their composition, association, latitude, altitude, fragmentation levels, inferences on possible corridors and species database complements the information needs.

Management of contagious (large), intact and juxtaposed patches of high diversity in any landscape should draw first attention for conservation. The ground inventories on species/ genetic diversity should further decide on priorities. The patches having higher biological diversity at landscape level will be subjected for more intensive ground inventories for assessing species/genetic diversity. The patches with genetic and species diversity should draw first attention followed by patches of high species or genetic diversity.

Bibliography

Abbey M, Corey M and Abramson I, Oracle 9i, A beginner's guide, Oracle Press, USA, 2001

Anonymous, Biodiversity Characterization at Landscape Level in North East India using Satellite Remote sensing and Geographic Information System, Project Report – Phase I, Indian Institute of Remote Sensing, Department of Space, Government of India, 2002

Anonymous, Oracle Magazine, Oracle Press, USA, 2002

Chuvieco, E. 1999. Measuring changes in landscape pattern from satellite images: short-term effects of fire on spatial diversity, *Int. J. of Remote Sensing,* 20(12): 2331-2346.

Coulson, R.N., Lovelady, C.N., Flamm, R.O., Spradling, S.L. and Saunders, M.C. 1990. Intelligent geographic information systems for natural resource management. In: M.G. Turner and R.H. Gardner, (eds.), *Quantitative Methods in Landscape Ecology,* Ecological studies 82 New York: Springler-Verlag. pp 153-172.

Delcourt, H.R. and Delcourt, P.A. 1988. Quaternary Landscape ecology: relevant scales in space and time. *Landscape Ecology* 2: 23-44.

http://www.asp101.com/articles/carvin/dataaccess/default.asp

http://msdn.microsoft.com/library/default.asp?URL=/library/devprods/vs6/vbasic/vbcon98/vbstartpage.htm

http://www.vb-bookmark.com

Rodgers W. A. and H. S. Panwar, 1988, Biogeographical classification of India, WII, Dehradun

Roy, P. S. and S. Ravan, 1996, Biomass estimation using satellite remote sensing data – an investigation on possible approaches for natural forest, *Journal of Biosciences,* 21(4): 535-561.

Roy, P. S. and Sanjay Tomar, 2000, Biodiversity characterization at landscape level using geospatial modeling technique, *Biological Conservation,* 95 (1): 95-109.

Siler B and J. Spotts, Using Visual Basic 6: Special Edition, Que Publishing, USA

Urman S., Oracle 9i PL/SQL Programming, Oracle Press, USA, 2001

WG IV / 5:

IMAGE-BASED GEOSPATIAL DATABASES

ecosystems, habitats, animals, fragmentation, levels, inventory of genetic resources and species diversity complements the information needs.

Management of ecologically fragile, intact and [illegible] of high diversity in any landscape should draw first [illegible] for conservation. The [illegible] of species [illegible] should [illegible] priorities. The [illegible] showing higher biological diversity at landscape level will be [illegible] for [illegible] ground [illegible] for [illegible] species/genetic diversity. The patches with genetic and species diversity should draw first attention followed by patches of high species or genetic diversity.

Bibliography

Ahern M, Green M and Abrahams [illegible], Oracle 9i, A beginner's guide, Oracle Press, USA, 2001

Anonymous, Biodiversity Characterisation at Landscape Level in North East India using Satellite Remote Sensing and Geographic Information System, Project Report – Phase 1, Indian Institute of Remote Sensing, Department of Space, Government of India, 2001

Anonymous, Oracle Manuals, Oracle Press, USA, 2001

Chandra, [illegible] 1993. [illegible] Changes in landscape pattern from satellite imagery: short-term [illegible] species diversity. Int. J. of Remote [illegible] 1385-2040.

Coulson, R.N., Lovelady, C.N., Flamm, R.O., Spradling, S.L. and Saunders, M.C. 1990. Intelligent geographic information systems for natural resource management. In: M.G. Turner and R.H. Gardner, (eds.), Quantitative Methods in Landscape Ecology. Ecological studies 82. New York: Springer-Verlag. pp 153-172.

Delcourt, H.R. and Delcourt, P.A. 1988. Quaternary Landscape ecology: relevant scales in space and time. Landscape Ecology 2: 23-44.

http://www.[illegible]

http://[illegible]

http://www.[illegible]

Rodgers W. A. and H. S. Panwar, 1988, Biogeographical classification of India, WII, Dehradun

Roy, P. S. and S. Ravan, 1996. Biomass estimation using satellite remote sensing data – an investigation on possible approaches for natural forest, Journal of Biosciences, 21 (4): 545-561

Roy, P. S. and Sanjay Tomar, 2000, Biodiversity characterisation at landscape level using geospatial modelling technique, Biological Conservation, 95 [illegible]

[illegible] Spatial, Using Visual Basic 6, Special Edition, Que Publishing, USA

Urman S., Oracle 9i PL/SQL Programming, Oracle Press, USA, 2001

Modeling Middle Urban Population Density With Remote Sensing Imagery

Lu An-min[1,2] Li Cheng-ming[2] Lin Zong-jian[2]

(1. College of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079;

2. Chinese Academy of Surveying and Mapping, Beijing, 100039)

Key words: population density, least square principle, remote sensing imagery

Abstract: A new land use density method is proposed based on least square principle after analyzing usual remote estimation methods. The main idea is: First, habitation types is defined in the study region; and the boundaries of every habitation types are lined out based on remote sensing imagery and the area of every habitation types are calculated; then mathematical models according to the population data of every sub-region are established; the best population density estimation of every habitation types are calculated with the least square principle when sub-regions number (m) is more than the number (n) of habitation types. The population estimation of any new region can be calculated since the population density of every habitation types are known as well.

1 INTRODUCTION

Population data is very important to stipulate for decision-making for governments. Census is carried out every decades. Though the precision is very high, the cycle is long and the cost is high. So every kinds of population estimation method with low cost and short cycle is worthy studied. People estimated population using remote sensing since 1960 and obtained some results. Wang H H (1990) estimated population in Gulou district in Nanjing in 1990 with land using density method . The sampled region is difficult to select in this method. Wang Faceng (1990) pointed out a method for estimation urban population using multispectral imagery. The land using density algorithm is used in the paper as well. Langford et al. (1991) derived population with LandSat TM imagery. The results were encouraging, but failed in globally applying the regionally derived model. Otherwise, Flowerdew and Green (1992) estimated population density using areal interpolation. Sautton P (1998) modeled population density with night-time satellite imagery and GIS. Langford et al (1994) and LIN Z J et al (2001) modeled population density by means of GIS. The land using density method is used in practice when modeling urban population density. The key of land using density method is the accurate represent of sampled region. The error of population estimation will be notable if sampled region is not correct. In fact, selection sampled regions is very difficult. There is still different between sampled regions and real regions. So errors of population estimation is inevitably. This paper tries to estimate population density of every kind of habitation types avoiding selecting sampled regions. So, usual methods of population estimation is overviewed.

2 Traditional methods for estimation population

In recent decades, scientists have done a lot of works in modeling population density in cities with remote sensing imagery. The main study results are habitation element method, land using density algorithm and built region area method, and so on.

2.1 Inhabitation unit method

The formula of inhabitation unit method is

$$P=N_1F_1+N_2F_2+\ldots+N_nF_n \quad (1)$$

P is total population, N is average population of every family, F is the count of family, 1…n are inhabitation types.

The distribution and instruction feature of buildings is analyzed with big scale remote sensing imagery. First, Inhabitation area and other building area is distinguished. Then, the type of inhabitation is distinguished and the inhabitation sum of every type is counted. The average population of every family obtained from sample. This method is fit for big scale remote sensing imagery and the precise of inhabitation count is 99%. The inhabitation is disperse in county. It is easy to count the inhabitation. So this method is fit for county.

2.2 Land use density method

In city, the population density is different from with every kind of land use types. The relation of population and land use type is

$$P = \sum_{i=1}^{n} (A_i D_i) \qquad (2)$$

A_1、$A_2 \ldots A_n$ are the area of every land use type. D_1、$D_2 \ldots D_n$ are the population density of every land use type, P is total population.

The land use type is obtained from the different of population density. The flow of this method is: first, the inhabitation and the non-inhabitation is distinguished with remote sensing imagery. Every kind of inhabitation type is distinguished in inhabitation area. The boundary of every kind of inhabitation is draw out. Then, the area of every kind of inhabitation is measured. The area of every kind of inhabitation type multiplying corresponding population density from sample is the estimation population of every kind of inhabitation type. The total sum of every kind of inhabitation type is the total population in study area.

3 Land use density method based on least square principle

The excellence of the land use density method is it's idea and calculation is simple; the disadvantage of the method is that the selection of sampled region is much more difficult. Is there a method the population density of every habitation type is estimated but the sampled region need not be selected. In some region, there is mathematical relation among the area, the population density and it's total population of every habitation type. The population density of every habitation type can be estimated via the mathematical relations.

Suppose there are j regions which the population is known, j=1,2…m. The population sum of every region is P_j. There are i kinds of habitation types, i=1,2…n. The population density of every habitation type is D_i, then

$$P_j = \sum_{i=1}^{n} (A_{ji} D_i) \qquad (3)$$

the quantity of (3) is

$$P_1 = A_{11}D_1 + A_{12}D_2 + \ldots + A_{1n}D_n$$

$$P_2 = A_{21}D_1 + A_{22}D_2 + \ldots + A_{2n}D_n$$

$$P_m = A_{m1}D_1 + A_{m2}D_2 + \ldots + A_{mn}D_n$$

the matrix of (3) is

$$P = AD \qquad (4)$$

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \ldots \\ P_m \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \ldots & \\ a_{m1} & a_{m2} \ldots a_{mn} \end{bmatrix}$$

$$D = \begin{bmatrix} D_1 \\ D_2 \\ \ldots \\ D_n \end{bmatrix}$$

When m>n, that is, the region count known population is more than habitation type count. The least square principle is used to calculate the best population density estimation of every kind of habitation which the errors of statistical population is least.

$$D = [A^T A]^{-1} A^T P \qquad (5)$$

General speaking, the more that m being bigger than n, the more the quality of population density estimation. The population estimation of any region can be calculated since the population density of every habitation types are known as well.

4 Experiment and analysis

There are 6 region which their population known including fengsheglu, daguanzhuang, xihoujie, beidajie, ronghejie and lingyuanxi, etc. The experiment data include 1:10000 remote sensing imagery(1995).

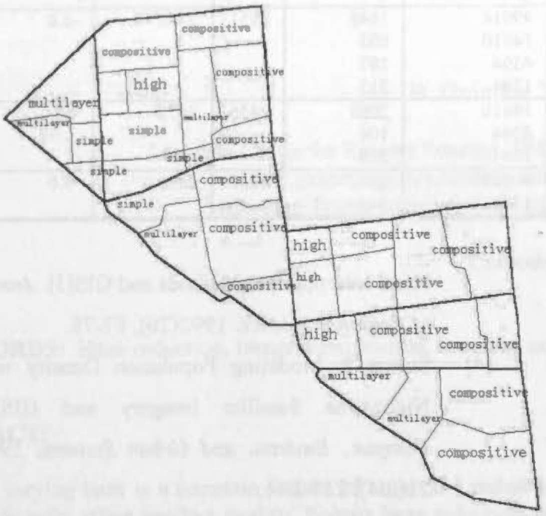


Fig.1 study region and its habitation types

4.1 Habitation types and their area

There are 4 kinds of habitations in the study region: high building, multilayer building, simple building and compositive building. The high building indicates a building with more than 7 layers. The multilayer building indicates a building with from 3 to 6 layers. The simple building indicate a building with less than 3 layers. The compositive building indicate a building which used for business and habitation.

First, habitation region is distinguished from other types. Every kind of habitation is distinguished in habitation region(Fig.1). Then, the boundary of every kind of habitation is drawn out and the area of every kind of habitation is measured(Tab.1)

Tab.1 The area and estimation population density

region	Total area (km^2)	Habitation type	Area (km^2)	Estimation density (person/km^2)	Estimation population	total	Statistical population	Relatively error (%)
Fengsheglu	0.09384	multilayer building simple building	0.06945 0.02439	14810 4394	1028 107	1135	1192	-4.7
Daguan zhuang	0.31318	high building multilayer building simple building compositive building	0.02663 0.04028 0.09269 0.15358	49014 14810 4394 1784	1305 596 407 274	2582	2173	18
xihoujie	0.21977	multilayer building simple building compositive building	0.07363 0.03141 0.11473	14810 4394 1784	1090 138 205	1432	1572	-8.9
beidajie	0.23179	high building compositive building	0.04578 0.18601	49014 1784	2243 331	2574	2621	-1.7
yonghejie	0.15311	high building multilayer building compositive building	0.0329 0.02792 0.09229	49014 14810 1784	1612 413 165	2190	2260	-3
lingyuanxi	0.21567	multilayer building compositive building	0.03282 0.18285	14810 1784	486 326	812	889	-8.6
total	1.22736					10725	10707	-0.1

4.2 Models calculation and error analysis

There are 4 kinds of habitation types and 6 sub-region with known population. 6 equations can be listed from formula (3)

$$P_1=A_{11}D_1+A_{12}D_2+A_{13}D_3+A_{14}D_4$$

$$P_2=A_{21}D_1+A_{22}D_2+A_{23}D_3+A_{24}D_4$$

$$\ldots$$

$$P_6=A_{61}D_1+A_{62}D_2+A_{63}D_3+A_{64}D_4$$

$P_1 \ldots P_6$ indicate the statistical population. D_1、D_2、D_3、D_4 indicate the population density of high building, multilayer building, simple building and compositive building, respectively. Indicate the area of high building, multilayer building, simple building and compositive building in fenshenglu region, etc(Tab.1).

There is 6 equations and 4 unknown. The best population density estimation of every habitation types are calculated with the least square principle(Tab.1). The precise of population estimation can be calculated using formula 3(Tab.1).

Population of any other new region can be calculated since the population density of every kind of habitation type is known. The experiment result is shown in table 2.

Tab. 2 population estimation of new unknown region

region	Total area (k m^2)	Habitation type	Area (k m^2)	Estimation density (person/km^2)	Estimation population	total	Statistical population	Relatively error (%)
qiliyan	0.30611	high building	0.03364	49014	1648	2951	2773	-6.4
		multilayer building	0.05426	14810	803			
		simple building	0.04273	4394	187			
		compositive building	0.17548	1784	313			
dongshun jie	0.31078	multilayer building	0.14124	14810	2091	2456	2572	4.5
		simple building	0.02431	4394	106			
		compositive building	0.14523	1784	259			
jianshelu	0.2509	high building	0.05738	49014	2812	3157	2896	-9.0
		compositive building	0.19352	1784	345			

5 CONCLUSION

This paper pointed out a new land using density method based on least square principle. The method need not sample the population density of every habitation types. The estimation workload of population density of every habitation type is low. The mathematical models is not influenced by random error of sample. The main idea is: (1) habitation types is defined in the study region; (2) the boundaries of every habitation types are lined out based on remote sensing imagery and the area of every habitation types are calculated; (3) mathematical models according to the population data of every sub-region are established; (4)the best population density estimation of every habitation types are calculated with the least square principle when sub-regions number (m) is more than the number (n) of habitation types. (5)The population estimation of any region can be calculated since the population density of every habitation types are known as well.

References

[1] Wang H H. City Population Estimation Method with Satellite Imagery[J]. *Remote Sensing Technology* .1990(3): 48-54.

[2] Wang F Z. Urban Population Estimation—Multispectral Imagery Analysis[J]. *Urban Environment and Urban Ecology.* 1990, 3(3): 42-34.

[3] Langford M, Maguire D J, & Unwin D J. The Areal Interpolation Problem: Estimating Population Using Remote Sensing in a GIS Framework[A]. In Masser I & Blakemore M. (Eds.), *Handling Geographical Information [C]*,55-77. London: Longman, 1991.

[3] Flowerdew R and Green M. Developments in Areal Interpolation Methods and GIS[J]. *Annals of Regional Science.* 1992(26), 67-78.

[4] Sutton P. Modeling Population Density with Night-time Satellite Imagery and GIS[J]. *Comput., Environ. and Urban Systems,* 1997, 21(3/4): 227-244.

[6] Langford M.and Unwin D J. Generating and Mapping Population Density Surfaces within a Geographical Information System. *The Cartographic Journal.* 1994(31), 21-26.

[7] Lin Zongjian, Jin Yimin and Li Chengming. Urban Population Geographical Information System[A]. *The 20th international cartographic conference[C]*, Beijing, 2001, 1279-1282.

[8] Ogrosky C E. Population Estimates from Satellite Imagery[J]. *Photogrammetric Engineering and Remote Sensing,* 1975(41), 707-712.

ROBUST HAZE REDUCTION: AN INTEGRAL PROCESSING COMPONENT IN SATELLITE-BASED LAND COVER MAPPING

B. Guindon [a], Y. Zhang [a b]

[a] Canada Centre for Remote Sensing, 588 Booth Street, Ottawa, K1A 0Y7, Canada
bert.guindon@ccrs.nrcan.gc.ca ; ying.zhang@ccrs.nrcan.gc.ca
[b] Intermap Technologies, 2 Gurdwara Road, Ottawa, K2E 1A2, Canada

Commission IV

KEY WORDS: Haze-reduction, image classification, accuracy assessment

ABSTRACT:

Spatially varying haze is a common feature of archival Landsat scenes currently being used for large-area land cover mapping and can significantly affect product quality. Robust haze reduction that is image-based and involves minimal operator intervention, is therefore a necessary a-priori step to information extraction. The practical implementation of a suitable methodology, based on a Haze Optimized Transform (HOT), is described. The approach is being used in a program to map the Great Lakes watershed with archival Landsat Multi-Spectral Scanner (MSS) imagery. The impact of haze reduction is assessed using inter-scene classification consistency as a ‘surrogate’ measure of user classification accuracy. Consistency comparisons made between sets of raw and haze-reduced scenes indicate that ‘rare’ class identification is most improved by this procedure.

1.1 Introduction

The last decade has witnessed a number of initiatives, for example the National Land Cover Data (NLCD, Vogelmann et al., 2001) and the GEOCover (Dykstra et al., 2000) programs, involving the derivation of regional and national-scale land cover products from moderate to high-resolution satellite images. The Landsat series of satellites has been a primary data source, since it has provided a continuity of coverage for the past 30 years through its Multi-Spectral Scanner (MSS), Thematic Mapper (TM) and Enhanced TM sensors. It is desirable, for a given land cover product, that source imagery be limited to a narrow acquisition window in order to encapsulate thematic information at a given ‘epoch’. Given the constraints of Landsat orbital characteristics and atmospheric conditions, at least a 3-year window is needed to obtain a comprehensive set of low cloud-cover scenes. Even with this relaxed constraint and especially of early ‘epochs’ (e.g. 1970s’s MSS coverage), scenes with visually apparent, spatially-varying haze must be included. This compromise can significantly degrade subsequent image classification and land cover mapping quality. As a consequence, haze reduction must be viewed as a necessary pre-processing step to information extraction.

The desirable characteristics of an ‘operational’ haze reduction module include robustness (i.e. applicable to a wide range of haze conditions), ease-of-use (i.e. minimal operator intervention) and that it be image-based since ancillary atmospheric data will be limited for most archival scenes. It should be noted that only ‘relative’ compensation for haze (e.g. adjustment to the haze level of the clearest portion of a scene) is sought because most image classification algorithms (e.g. unsupervised clustering and supervised maximum likelihood classification) do not require absolute radiometric calibration. Over the past year, research has been undertaken at the Canada Centre for Remote Sensing (CCRS) to formulate and implement such a methodology. This has led to the development of a Haze Optimized Transform, hereafter referred to as HOT, to capture the spatial distribution of haze over Landsat scenes. An implementation of HOT-based haze reduction has been included within the framework of a proto-type workstation currently being used to map synoptic land cover of the Great Lakes watershed with archival Landsat MSS imagery. In this paper we describe the results of a study to assess the impact HOT-based processing by comparing land cover derived from raw versus haze-reduced scenes.

A detailed description of the formulation of the HOT and its example application to Landsat Thematic Mapper data has been presented elsewhere (Zhang et al., 2002a, Zhang and Guindon, 2002b), however, here we present an overview of its salient features. The transform exploits the fact that, under uniform haze conditions, spectral responses of a broad range of common surface classes are highly correlated in the Landsat visible bands. On the other hand, relative response to the presence of haze is wavelength-dependent, being most acute toward the blue end of the spectrum. In the HOT formulation, the clearest areas of a scene are first used to define a ‘clear line’, i.e. a thematic response line in visible spectral space (e.g. bands MSS1 vs. MSS2). The transform then quantifies, for a given image pixel, its spectral deviation from this line. The resulting raster overlay of HOT response then encapsulates the spatially-varying haze component relative to the clearest scene areas.

2. Practical aspects of HOT-based haze reduction

Here we present an overview of our haze reduction procedure with particular emphasis on implementation considerations related to the routine processing of Landsat MSS data. As mentioned earlier, automated image-based processing has been sought. Our current procedure is limited to the radiometric adjustment of the two visible bands (MSS1 and MSS2), in part because of their enhanced sensitivity to haze vis a vis the infrared bands and in part because haze has an additive effect on radiance in this portion of the spectrum.

Figure 1 shows a schematic of the principal processing steps of the procedure.

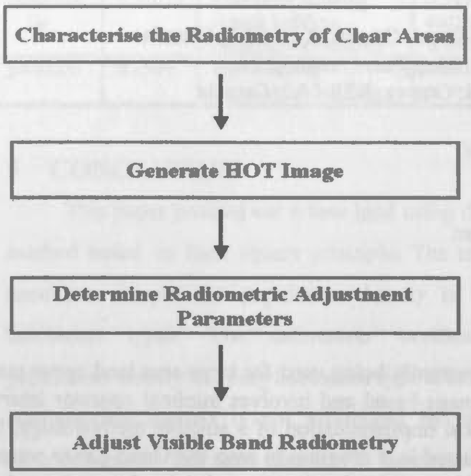

Characterise the Radiometry of Clear Areas
Generate HOT Image
Determine Radiometric Adjustment Parameters
Adjust Visible Band Radiometry

Figure 1. Outline of HOT-based haze reduction

2.1 Characterize the Radiometry of Clear Areas

Our haze adjustment procedure is a relative one in the sense that it attempts to normalize a scene to the radiometry of the clearest portion of a scene. As a consequence, the first step in the process involves collecting a sample of pixels from clear areas to be used to parameterize the 'clear area' radiometry. Sample selection is currently done manually (the only manual step in the overall procedure) by delineating rectangular windows in the scene based on a visual estimation of intra-scene clarity. A key objective of this process is the selection of a sample that spans the full visible band reflectance range, i.e. that it includes a mix of vegetated and non-vegetated land cover.

The clear area sample is analyzed in the visible band spectral space, i.e. (MSS1 (green) vs. MSS2 (red)). In this space data is highly correlated and can be characterized by a 'clear line' defined through least-squares regression.

2.2. Generate HOT Image

In the visible band space, hazy pixels will deviate from the clear line since haze has a greater radiometric influence in the green band than in the red band. The HOT transform uses the observed band 1 and band 2 radiances of a given pixel to compute its orthogonal displacement from the clear line. When this is done for each pixel, a HOT image is generated that quantifies the spatial and intensity distribution of haze. In the case of Landsat MSS, this product is noisy because of a combination of low dynamic range and quantization errors in the visible band data. Noise reduction through spatial smoothing is feasible since haze variations tend to be on the scale of kilometers compared to the 80 meter spatial resolution of the MSS sensor. We accomplish smoothing by generating an image pyramid from the HOT image (Schowengerdt, 1997), selecting a suitable level then expanding that level back to the sampling interval of the parent HOT image through bilinear interpolation. For MSS, experience has indicated that the second pyramid level (i.e. 4 times reduction) provides a robust smoothing solution.

2.3. Determine Radiometric Adjustment Parameters

For those pixels exhibiting a HOT response greater than that of the clear areas, a radiance adjustment (i.e. reduction) will be applied. Based on modelling experiments (Zhang et al. 2002a), a linear relationship between radiance adjustment and HOT level serves as a good first approximation. The scale factor of this relationship is determined using a 'dark target' subtraction approach (e.g. Chavez, 1988). For a given visible band, a series of grey level histograms are generated for pixels in HOT intervals ranging from clear to the haziest levels for which adjustment is practical. The rate of increase of histogram lower bound with HOT defines the required scale factor. It should be noted that water bodies with significant sediment content will also trigger a high HOT response. To overcome this problem, a water mask is created by thresholding the MSS4 image and used to exclude water pixels from the histogram analysis described above.

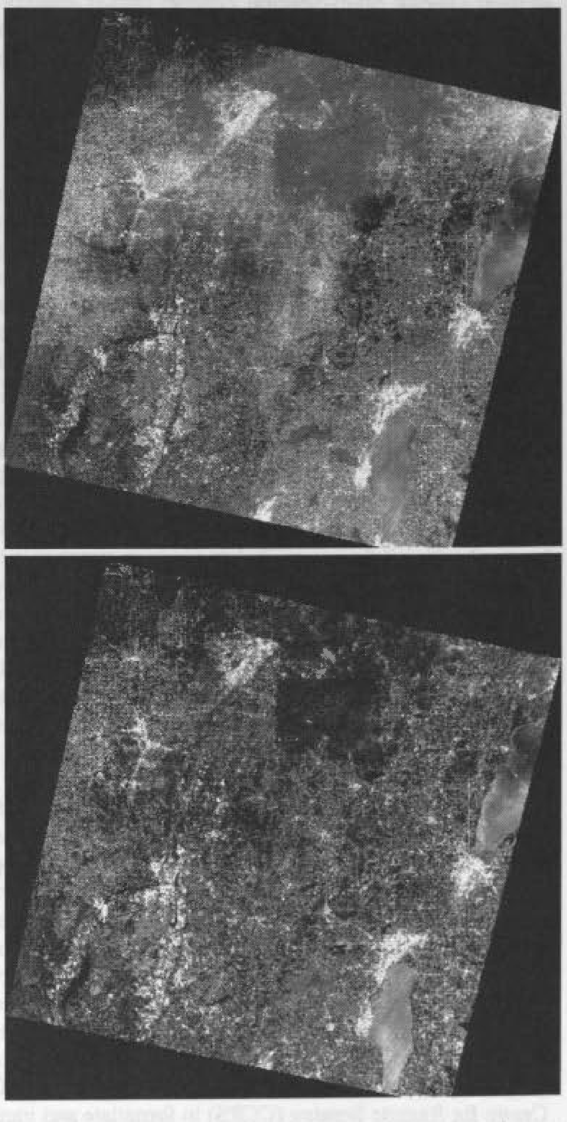


Figure 2. Example of adjustment (lower) from a hazy image (upper)

2.4. Adjust Visible Band Radiometry

With scaling factors in hand, the two visible bands can then be adjusted on a per pixel basis. Water body pixels do not undergo this adjustment. This is not a major limitation since accurate water classification can be achieved even in the presence of high levels of haze. Finally, because of the necessary smoothing operation described above, pixels near the image boundary cannot be correctly processed and must be trimmed from the final product. Figure 2 shows an example MSS1 case with the raw image (top) exhibiting extensive regions of low to moderate level haze that have been successfully suppressed through the adjustment process (bottom image).

3. Classification accuracy impact of haze removal

The assessment described here was conducted as part of a joint project involving the Canada Centre for Remote Sensing and the United States Environmental Protection Agency to generate synoptic land cover products of the Great Lakes watershed. This is being done with archival Landsat MSS data sets for three acquisition epochs, mid 1970's, mid 1980's and early 1990's. Coverage of the complete watershed for each epoch encompasses approximately 50 scenes.

A proto-type system, QUAD-LACC, has been developed to accomplish the mapping task Its methodology is based on independent classification (unsupervised clustering followed by cluster labelling) of each scene and subsequent classification compositing (Guindon and Edmonds, 2002a). In this way, a major feature of Landsat data is exploited, namely, the extensive overlap coverage of scenes from adjacent orbital tracks. This approach was selected for two reasons. First, by utilizing all available imagery, information degradation and loss in individual scenes due to cloud and haze can be partly overcome. Second, classification consistency in overlap regions can be used as an indicator of classification performance that characterizes the spatial distribution of accuracy across large area (i.e. multi-scene) land cover products (Guindon and Edmonds, 2002b).

To assess the impact of haze removal on classification performance, we have selected a set of 35 source scenes from the 1990's epoch and created two sets of classified products, one derived from raw and the other from haze reduced imagery. For the purposes of this study, land cover has been stratified into two broad categories only, forest vs. non-forest land. Within each classified data set, consistency analyses have been undertaken on all cross-track overlap regions. Comparisons have then been made between consistency levels of corresponding overlap regions in the two product sets.

The level of consistency can be quantified in the following manner. Consider the case of the overlap region of two classified scenes, #1 and #2. The consistency of classification of a class A in scene #1 can be defined as the portion of pixels, classed as A in scene #1, that have the same classification in scene #2. For a two-class scenario (classes A and B), the above consistency, C_{1A}, can be expressed as:

$$C_{1A} = \frac{fP_{1A}P_{2A} + (1-P_{1B})(1-P_{2B})}{fP_{1A} + (1-P_{1B})} \qquad (1)$$

where the P's refer to the producer accuracies of each class in each scene (e.g. P_{1A} is the class A producer accuracy in scene #1) and f is the ratio of the numbers of true class A to true class B pixels in the overlap region (Guindon and Edmonds, 2002b). Producer accuracy, P_A, is defined as the probability that a true class A pixel is correctly labelled. This differs from user accuracy, Q_{1A}, which is the probability that a pixel, classed as A, is indeed of true class A. User accuracy can be expressed as:

$$Q_{1A} = \frac{fP_{1A}}{fP_{1A} + (1-P_{1B})} \qquad (2)$$

As has been shown elsewhere and can be seen from a comparison of equations 1 and 2, consistency is a useful surrogate of conventional user accuracy (Guindon and Edmonds, 2002b).

For our rudimentary classes we would expect producer accuracies to be high (typically above 85%). On the other hand, over the Great Lakes watershed, consistency and hence user accuracy can vary significantly because of the variations in f. If we take 'forest' for class A, f can range as low as 0.05 in the largely agricultural/urban south to more than 10 in the boreal forests of the north. In the case of southern scenes, the majority of pixels classed as forest can be commission errors. As a result of commission error randomness, forest consistency can be low (i.e. much less than 0.5) and therefore, as with user accuracy, be very sensitive to even moderate improvements in producer accuracy that may arise from haze reduction.

For a given scene pair and its overlap region, f remains constant but haze processing will alter producer accuracy and hence consistency. While the impact of haze reduction will vary from scene-to-scene and hence overlap-to-overlap region, we can consider some simple scenarios to attempt to gauge its overall impact on the scene population. We will assume that the producer accuracies of the two classes are the same raw imagery and are improved by the same percentage by haze reduction. Initially, four scenarios were considered, namely initial raw producer accuracies (PRAW) of 85 and 90% and incremental improvements (PINC) of 2.5 and 5%. Equation 1 has been used to generate raw-improved consistency 'trajectories' for values of f ranging from 0.05 (low consistency) to 20 (high consistency). The trajectories are shown in Figure 3 along with the unity line (i.e. no change) for reference. All four cases predict similar, moderate levels of consistency improvement.

Figure 4 illustrates the raw versus haze-reduced consistency comparison for the Landsat MSS data set. A number of points can be made regarding these results.

(a) The data set exhibits significant scatter with offsets relative to the unity line which are much larger than those predicted from the four scenarios especially at low consistency levels.

(b) Haze reduction results in significant improvement in all low consistency cases (i.e. cases where raw image consistency was less than 0.5). The impact is more difficult to discern at higher consistency levels from the diagram alone because of scatter. To better understand overall trends, we have grouped data points in intervals of raw consistency and computed the average haze reduced consistency for each. The results, shown in Table 1, indicate that consistency enhancement gradually diminishes with increasing consistency level and that haze processing may reduce consistency at the highest levels.

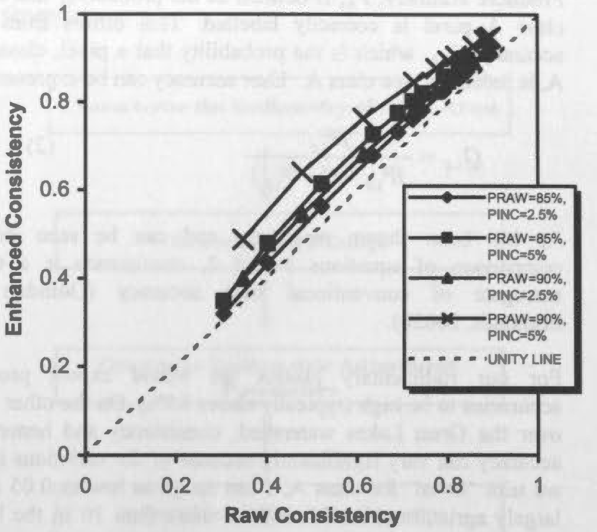


Figure 3. Predicted consistency relationships for model producer accuracy scenarios

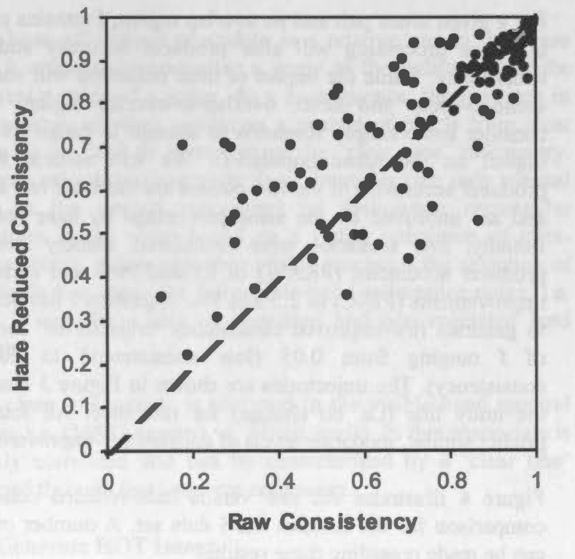


Figure 4. Raw versus haze-reduced consistency for the Landsat MSS data set

(c) Haze reduction results in significant improvement in all low consistency cases (i.e. cases where raw image consistency was less than 0.5). The impact is more difficult to discern at higher consistency levels from the diagram alone because of scatter. To better understand overall trends, we have grouped data points in intervals of raw consistency and computed the average haze reduced consistency for each. The results, shown in Table 1, indicate that consistency enhancement gradually diminishes with increasing consistency level and that haze processing may reduce consistency at the highest levels.

(d) Low consistency is associated with so-called 'rare' classes, i.e. those with small proportional coverage relative to classes with which they are likely to be confused (i.e. low f values). Unless a rare class exhibits a very high level of spectral uniqueness, its producer accuracy will be degraded due to the limitations of the classification process where the goal is to assign a label to a given cluster that is indicative of its dominant class. When a class is under-represented relative other classes of similar spectral properties, many of its member pixels will be members of clusters dominated by other classes thereby reducing its producer accuracy. Based on trial runs using equation 1, we conclude that the observed large improvements at low consistency are explainable by low raw producer accuracies, typically in the range of 50 to 60%, that are increased to the 85% range by haze processing. In addition, as f increases, forest producer accuracy increases to the expected levels of 85 to 90% even in the raw scenes. Consequently, the impact of haze processing gradually diminishes. Independent scene-based classification assessments using ground reference data confirms these model predictions and are being published elsewhere (Guindon and Edmonds, 2002b).

Raw Scene Consistency Range	Mean Haze-Reduced Consistency
0.0 - 0.2	0.43
0.2 - 0.4	0.54
0.4 - 0.6	0.63
0.6 - 0.7	0.72
0.7 - 0.8	0.73
0.8 - 0.9	0.86
0.9 - 1.0	0.92

Table 1. Mean consistency levels of haze-reduced scenes for different ranges of raw scene consistency.

4. Conclusions

Although Landsat data selection has been optimized for many large-area monitoring programs, constituent scenes may be contaminated by haze, thereby jeopardizing effective information extraction. A robust, image-based haze reduction methodology has been described which is suitable for bulk processing applications. Assessment of its impact on Landsat MSS scenes indicate that it can significantly improve inter-scene classification consistency and hence classification accuracy. This is particularly true for classes of low proportional coverage (i.e. 'rare' classes) whose producer accuracies are limited by the framework of the classification methodology.

References

Chavez, P.S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 24, pp. 450-479.

Dykstra, J.D., Place, M.C. and R.A. Mitchell, 2000. GEOCover-ORTHO: creation of a seamless geodetically accurate, digital base map of the entire earth's land mass using Landsat multispectral data. *Proceedings of the ASPRS 2000 Conference*. Washington, D.C., 7 pages (CD publication).

Guindon, B. and C.M. Edmonds, 2002a. Large-area land cover mapping through scene-based classification compositing. *Photogrammetric Engineering and Remote Sensing*, (in press).

Guindon, B. and C.M. Edmonds, 2002b. Using classification consistency in inter-scene overlap regions to model spatial variations in accuracy over large-area land cover products. *Proceedings of the Remote Sensing and GIS Accuracy Assessment Symposium*. Las Vegas, Nevada, Dec. 11-13, 2001, (in press).

Schowengerdt, R. A., 1997. *Remote Sensing: Models and Methods for Image Processing*. Academic Press, San Diego, pp. 264-271.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K. and N. Van Driel, 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat thematic mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, 67, pp. 650-662.

Zhang, Y., Guindon, B. and J. Cihlar, 2002a. An Image Transform to Characterize and Compensate for Spatial Variations in Thin Cloud Contamination of Landsat Images. *Remote Sensing of Environment*, (in press).

Zhang, Y. and B. Guindon, 2002b. Assessment of the effectiveness of HOT-based haze removal in ETM+ images on land cover classification. (in preparation).

Acknowledgements

The authors wish to thank C. M. Edmonds, United States Environment Protection Agency, for providing Landsat scenes of the U.S. portion of the Great Lakes region from the North American Landscape Characterization (NALC) archive.

SEGMENTATION OF HIGH-RESOLUTION REMOTELY SENSED DATA - CONCEPTS, APPLICATIONS AND PROBLEMS

J. Schiewe

University of Vechta, Research Center for Geoinformatics and Remote Sensing, PO Box 1553, 49364 Vechta, Germany,
jschiewe@fzg.uni-vechta.de

Commission IV, WG IV/5

KEY WORDS: segmentation, feature level fusion, multi-source data, high-resolution data

ABSTRACT:

Segmentation algorithms have already been recognized as a valuable and complementary approach that similar to human operators perform a region-based rather than a point-based evaluation of high-resolution and multi-source remotely sensed data. Goal of this paper is to summarize the state-of-the-art of respective segmentation methods by describing the underlying concepts which are rather complex in the case of processing remotely sensed data, demonstrating various applications (automatical object recognition, signal-based fusion as support for visual interpretation, and estimation of the terrain surface from Digital Surface Models), and identifying yet existing problems and further research and development needs.

1. INTRODUCTION

Remote sensing data are an important source for generating or updating GIS databases in a variety of applications. As a reaction to the limitations of the available data sources with respect to their spatial resolution and inherent information content, in the last ten years the focus was laid on the development of advanced sensors, for instance laser scanners (presently, about 65 operational airborne systems world-wide), radar-interferometric sensors (e.g., the Shuttle Radar Topographic Mission), or electro-optical cameras with in-flight-stereo capabilities (e.g., airborne systems like ADS 40, DMC or HRSC-A, and spaceborne systems like Ikonos or Quickbird). In order to increase the information potential for interpretation purposes, multi-sensor systems have been developed for the simultaneous acquisition of image and elevation data (e.g., the TopoSys II sensor).

However, with these sensors the user community faces new problems in the automatical analysis of these types of data:

1. The high spatial resolution of the advanced sensors increases the spectral within-field variability - in contrast to the integration effect of earlier sensors - and therefore may decrease the classification accuracy of traditional methods on per-pixel basis (like the Maximum-Likelihood method). Hence, novel and efficient analysis techniques become a mandatory requirement for efficient processing and analysis.
2. The availability of multi-sensoral or even multi-source data (e.g., existing information from Geographical Information Systems, GIS) is strongly correlated with the necessity for a fusion on data but in particular on feature or decision level (Pohl & van Genderen, 1998). Unfortunately, many standard techniques are not able to handle heterogeneous data sources and context information.

In this context, segmentation algorithms have already been recognized as a valuable and complementary approach that - similar to human operators - create regions instead of points or pixels as carriers of features which are then introduced into the classification stage. The conceptual idea is that each of these regions corresponds exactly to one and only one object class (object oriented approach). Furthermore, segmentation algorithms are able to handle multiple data and information sources, thus performing a fusion on feature level.

Goal of this paper is to summarize the state-of-the-art of the use of segmentation algorithms for various processes applied to remotely sensed data by describing the underlying concepts which are rather complex in this case (chapter 2), demonstrating various applications as carried out at our institute (automatical object recognition, signal-based fusion as support for visual interpretation, and estimation of the terrain surface from Digital Surface Models; chapter 3) and identifying yet existing problems and further research and development needs (chapter 4).

2. CONCEPTS OF SEGMENTATION ALGORITHMS

Segmentation is the process of completely partitioning a scene (e.g., a remote sensing image) into non-overlapping regions (segments) in scene space (e.g., image space).

Respective algorithms have been developed within Pattern Recognition and Computer Vision since the 1980's with successful applications in disciplines like medicines or telecommunication engineering. However, due to the complexity of the underlying object models and the heterogeneity of the sensor data in use, their application in the fields of Remote Sensing and photogrammetry was limited to special purpose implementations only. With the advent of high-resolution as well as multi-source data sources the general interest in segmentation methods has become evident again, and at last significant progress in terms of user awareness was achieved with the introduction of the first commercial and operational software product (*eCognition* by Definiens-Imaging) in the year 2000.

In the following we will briefly describe the general concepts of segmentation methods (section 2.1) and will emphasize the particularities for the evaluation of remotely sensed data (section 2.2).

2.1 General concepts

2.1.1 Principles and strategies: Basic task of segmentation algorithms is the merge of (image) elements based on homogeneity parameters or on the differentiation to neighbouring regions (heterogeneity), respectively. Thus, segmentation methods follow the two strongly correlated *principles of neighbourhood* and *value similarity*.

Generally the following strategies for partitioning a scene into regions are distinguished:

- point-based
- edge-based
- region-based
- combined

Point-based approaches are searching for homogeneous elements within the entire scene by applying global threshold operations which combine such data points that show an equal or at least similar signal or feature value. This threshold can divide the feature space into two or more parts (binarization or generation of equidensites, resp.). The choice of threshold values can be performed statically or dynamically based on histogram information.

Because this grouping has not considered the principle of neighbourhood so far, a connection analysis in scene space is performed in a second step. Here, spatially connected elements (components) of equal value (e.g., grey value "1") are grouped to one region (*component labeling*).

It has to be noted that point-based approaches are less suitable for the evaluation of remotely sensed data due to varying reflection values for a certain object placed at different locations within the real world and the sensed scene (as an example see figure 1, middle).

Edge-based approaches describe the segments by their outlines. These are generated through an edge detection (e.g., a Sobel filtering) followed by a contour generating algorithm. Optionally, the transition from the outlines to the interior region can be achieved by contour filling methods like the watershed algorithm.

As figure 1 (right) demonstrates, the main disadvantage of edge-based approaches is that the edge and also the contour image is strongly affected by noise (in particular in wooded regions, less crucial for artificial objects) which may lead to an unacceptable over-segmentation.

Region-based approaches start in the scene space where the available elements (pixels or already existing regions) are tested for similarity against other elements (see section 2.1.2). Concerning the definition of the initial segmentation the procedures of *region growing* (bottom-up, i.e. starting with a seed pixel) and *region splitting* (top-down, i.e. starting with the entire scene) are distinguished. One disadvantage of the splitting method is that it tends to an over-segmentation because a splitting always produces a fixed number of sub-regions (normally: 4) although two or three of them might actually be homogeneous with respect to each other. As a consequence, one can apply a method combination which leads to the *split-and-merge* algorithm that after every split tests whether neighbouring regions are so similar that they should be re-merged again.

In the following explanations and applications we will mainly rely on region growing approaches.

2.1.2 Homogeneity criteria: In the following the realization of the *principle of value similarity* (or homogeneity, resp.) will be discussed more in detail. Given two elements A and B (i.e. pixels or regions) one possibility for deriving a *homogeneity measure* is to compare a certain feature of A and B (e.g., the grey value) through its Euclidian distance. In addition, it is also possible to consider simultaneously multiple features f_i (i=1, ..., n) of A and B, with the option to introduce individual weights g_i. Hence, a fusion on feature level is realized which gives the corresponding heterogeneity measure Δh by:

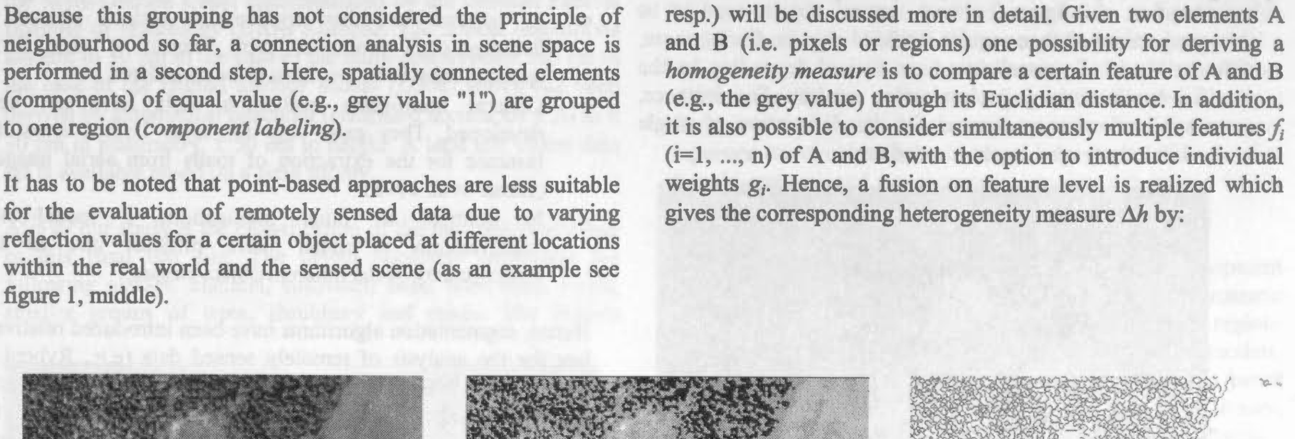


Figure 1. Given image (left) and exemplary point-based segmentation with four classes (middle), and contours as input for edge-based segmentation (right) - data courtesy of TopoSys GmbH -

$$\Delta h = +\sqrt{\sum_{i=1}^{n} g_i \cdot (f_{A,i} - f_{B,i})^2} \qquad (1)$$

As an alternative, the homogeneity measures can be computed before and after an eventual merge of the elements A and B.

With the obtained measure Δh it can be decided whether the elements A and B have to be merged to a larger segment or not. This is done by a comparison with a threshold that controls the size and number of segments and with that the level of generalization of the segmentation process (see also section 2.1.3).

The merging algorithm can also consider further constraints concerning neighbourhood and similarity: In the simplest case element A accepts B if the homogeneity measure is below the given threshold (*fitting*). In contrast, A may accept only that neighbouring element B which fulfills the homogeneity criterion best (*best fitting*). Furthermore an element C is connected to A (which is similar enough to B) only if B and C as well as A and C are similar enough (*local mutual best fitting*).

2.1.3 Evaluation of segmentation: In general the described segmentation methods do not yield a perfect partition of the scene but produce either too much and small regions (*over-segmentation*) or too less and large segments (*under-segmentation*). The first effect is normally a minor problem because in the following classification step neighbouring segments can be attached to the same category a posteriori.

Applying segmentation methods to remotely sensed data we can observe that over- and under -segmentation can occur within a single scene depending on the heterogeneity of objects under consideration. As figure 2 shows, natural objects tend to be stronger partitioned than regular artificial objects. Furthermore, different levels of generalization are desired depending on the specific applications (e.g., evaluation scales). For instance, some applications may demand for the delineation of single trees while others need larger wooded areas.

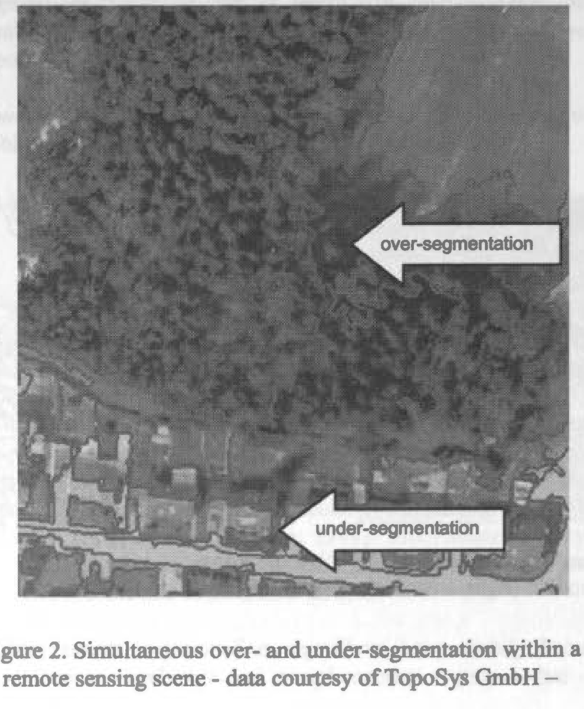


Figure 2. Simultaneous over- and under-segmentation within a remote sensing scene - data courtesy of TopoSys GmbH –

Methods for the evaluation of segmentation results are discussed for example by Hoover et al. (1996), Zhang (1996) or Levine & Nazif (1985). In the last case the authors also present developments for a dynamical determination of the segmentation quality by continuously computing homogeneity measures of all intermediate regions. However, it has be noted that presently the most reliable evaluation method is still a visual interpretation that has to consider the exact geometrical position of the segment borders as well as the membership of one and only one object class to a single region. However, with that the generalization level as well as the homogeneity features and parameters are controlled in a rather subjective manner.

2.2 Particularities for evaluating remotely sensed data

Numerous segmentation methods have already been developed for various applications, including medicines, telecommunication engineering or the analysis of dynamic scenes in neuro-informatics. Consequently, the available software products are originating from these or similar disciplines.

As already mentioned at some places in section 2.1 several reasons lead to the fact that existing methods and implementations from these disciplines can not be directly transferred to the domain of remote sensing, in particular:

- Remote sensing sensors are producing multi-spectral, sometimes also multi-scale input data, so that in contrast to the most often used panchromatic and monoscopic image data in the disciplines mentioned above not only the complexity but also the redundancy increases.
- Manifold additional data (e.g., GIS or elevation data) are available.
- In contrast to other applications various objects of heterogeneous properties with respect to size, form, spectral behaviour, etc. have to be considered.
- General multi-scale evaluation tools have not been developed. They exist only for some limited areas, for instance for the extraction of roads from aerial imagery (Ebner et al. 1998).
- In contrast to other applications a model-based interpretation is much more difficult due to the heterogeneity of the inherent object classes;

Hence, segmentation algorithms have been introduced relatively late for the analysis of remotely sensed data (e.g., Ryherd & Woodcock 1996). As a consequence, the first commercial software packages were introduced not before the year 2000, for example the *Stand Delineation Tool* of the Finnish company Arboreal for forest inventory purposes (Arboreal, 2002) or *eCognition* (Definiens-Imaging 2002) that aims at a more general use.

In conclusion segmentation approaches for the evaluation of remotely sensed data have to be rather complex systems that should

- handle various input data simultaneously (*multi-source aspect*)
- integrate a couple of segmentation strategies serving for all object types which shall be extracted (*multi-method aspect*)
- create various levels of generalization at the same time - due to the fact that different objects are represented best at

different scales (*multi-scale aspect*).

3. APPLICATIONS

In the following we will give an insight into the various applications of segmentation methods with respect to the evaluation of remotely sensed data - also considering the above mentioned specific aspects. Not surprisingly, the emphasis of the usage is on automatical object recognition tasks (section 3.1), but we will also demonstrate alternative applications such as signal-based fusions for visual interpretation purposes (section 3.2), and the terrain surface estimation from Digital Surface Models (section 3.3).

3.1 Automatical object recognition

3.1.1 Brief overview: A couple of scientific work has been undertaken on the use of segmentation methods for the extraction of certain features from close range photogrammetric imagery on one hand, and for the detection of object classes from multi-spectral or panchromatic imagery, for example considering buildings (e.g., Brenner, 2000), buildings and roads (Hoffmann, 2001), or airports (McKeown et al., 1985). Only a limited number of work is concerned with a more detailed classification, for instance Bauer & Steinnocher (2001) perform a recognition of 11 object classes in an urban scene.

3.1.2 Biotope monitoring project: A concrete project at our institute had the purpose to test the applicability of data of the High Resolution Stereo Camera–Airborne (HRSC-A; Wewel et al. 1998) for the classification of biotope types on reaches of Federal waterways characterized by strong relief features. The data are based on a flight mission along a reach of the Main-Danube Canal commissioned by the German Federal Institute of Hydrology (BfG), Koblenz. The spatial resolutions amount to 30 cm in the case of the multi-spectral and 200 cm in the case of the Digital Surface Model (DSM) which has been derived by automatical matching (estimated accuracies ± 20 to ± 30 cm in planimetry, ± 50 cm in height. A land use vector data set is available based on a field survey.

Aim of our study is the classification of the land use/land cover in this rural test site. The object catalogue comprises the following classes: channel, cultivated field, bare field, forest, smaller groups of trees, shrubbery and roads. The objects exhibit a large range of different scales ranging from small shrubs to large cultivated fields.

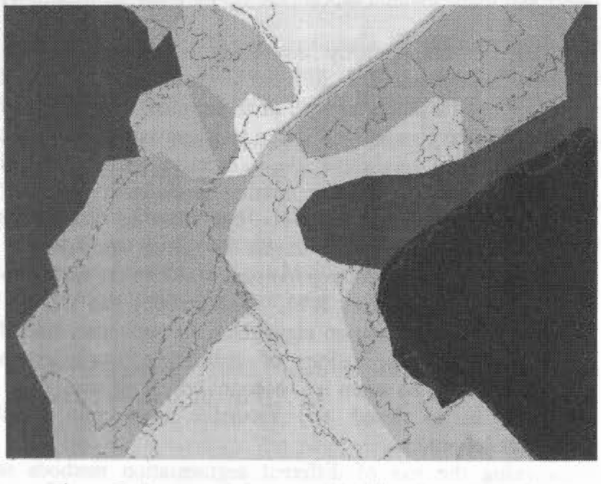


Figure 3. Improved degree of details of segmentation compared to biotope classification based on field survey

Using the hierarchical approach of the *eCognition* software the segmentation is performed at different levels of generalization. The best segmentation results are achieved with different spectral combinations and weighting factors for the different bands with respect to the object classes. As figure 3 demonstrates, a much higher degree of detail could be achieved with the segmentation approach compared to the biotope classification based on the field survey. For a more detailed project description refer to Schiewe et al. (2001).

3.2 Signal-based fusion for visual interpretation

There are a couple of reasons still to deal with aspects of the visual interpretation of imagery. On one hand, most of the (semi-) automatical object recognition procedures do not lead to satisfying results. On the other hand, an increasing number of airborne and spaceborne sensors produce spatially lower resolution but multi-spectral data as well as higher resolution but pan-chromatic data (e.g., Ikonos, Quickbird) so that a signal-based fusion of the above mentioned image data prior to a visual interpretation is very often demanded. Such a fusion shall combine the respective advantages, in particular

- emphasize certain image features (e.g., gradients, colors),
- substitute missing information (e.g., in shadow regions),
- improve geometrical corrections (e.g., by using data of higher geometrical accuracy),
- enable stereoscopic evaluations by merging stereo partners,
- detect changes in multi-temporal data sets.

Unfortunately conventional, point-based operating signal-based fusion methods do not lead to reasonable and sustainable results. One key problem is that a couple of procedures yield a bad color reproduction and a decrease of the global contrast which is mainly due to the neglecting of the fact that the input channels are normally decorrelated to each other. Furthermore important structural information (e.g., edges) get lost very often if multi-spectral and panchromatic imagery are treated in the same manner (Schiewe, 1999).

Our alternative approach which is still under development follows the idea to combine multi-spectral and panchromatic data not point-wise (using global operations) but in a region-based manner applying different and suitable functional models. In a first step, the segmentation partitions the image set based on information about the texture, dominance of certain colors, existence of edges, and high correlation between certain bands.

In the classification step appropriate functional models are attached to the segments according to their specific properties (e.g., contrast). In this context it has been found that additive approaches generally emphasize color information while multiplicative algorithms (e.g., the Brovey transformation) show off structural information (Schiewe, 1999).

3.3 Terrain surface estimation from DSMs

Most elevation evaluation systems (like stereo matching or laserscanning) produce the respective largest values above a position, i.e. a *Digital Surface Model (DSM)*. However, for some applications (e.g., for hydrological modeling) those objects that stand clearly above the terrain surface (e.g., buildings, trees) are not of interest, i.e. the *Digital Terrain Model (DTM)* is required. Furthermore, for object extraction purposes and virtual city modeling the absolute object heights above the terrain surface (i.e., the differences between DSM and DTM) are needed. Because in practise a DTM is not always available, not sufficiently accurate or reliable enough, or too expensive, a substitute has to be estimated (*estimated DTM, eDTM)*, i.e. a *normalization* has to take place. For this task a couple of geometrically based algorithms have been developed (for an overview see Schiewe, 2001), but none of them has succeeded to operationality so far due to the limited quality (especially for inclined terrain) and the missing grade of automatization (especially due to abstract thresholds). Thus, an application dependent combination of methods incorporating rather high interactive efforts for controlling and editing are presently applied in practise.

In the following we want to present a novel region-based, multi-scale approach for the task of terrain surface estimation. Firstly, for the segmentation a proper choice of homogeneity criteria has to be applied. Here we follow the hypothesis that regions which have to be reduced to the terrain surface are characterized by strong altitude gradients and curvatures (Schiewe, 2001). By extending the complexity presented so far by additionally introducing multi-spectral or pan-chromatic image data, we have experienced worse segmentation results due to irregular and inaccurate border lines. This effect is mainly due to a strong affection of image data by noise and shadows.

For the interpretation of the obtained segments we apply a *fuzzy logic* classification approach. There are several reasons for introducing partial rather than crisp memberships in this context (e.g., see Cheng, 2002): On one hand the description of the real phenomenon is neither geometrically sharp (e.g., there is no exactly defined border between forest and terrain) nor standardized (e.g., there are no generally accepted instructions for masking out brush, bridges, etc.). On the other hand the limited spatial sampling rates and measurement errors (the latter being mostly unknown) also lead to indeterminate boundaries.

As classification features the mean of altitude gradients (pointing to forest areas) and the 90% percentile of gradients (pointing to steep edges likes in buildings) can be taken into account. In order to handle the problem of flat roofs (i.e., segments with low gradients), we also consider the 90% percentile of gradients of all surrounding segments (referring to a building edge with high gradients). Finally, the mean difference between first and lust pulse laser scanning measurements, if available, can be taken into account for each segment (also pointing to wooded areas and building outlines).

In some cases it is of interest for the generation of the estimated DTM which object type is associated to a non-terrain area. For instance, for hydrological modeling tasks an interpolation within wooded areas is meaningful while for buildings it is not. Hence, an additional separation of buildings from wooded areas in the classification step becomes necessary. Besides the normalized DSM altitude especially the Normalized Difference Vegetation Index (NDVI) or the spectral texture (in pan-chromatic imagery) are meaningful features.

Figure 4 illustrates the processing results after applying the described region-based approach. The multi-sensor data set which covers a settlement near the City of Ravensburg (Germany) consists of simultaneously acquired laserscanning and multi-spectral image data of the TopoSys II sensor (TopoSys, 2002). As an example, the fuzzification result for the feature "mean gradient" demonstrates the desired high membership values of wooded areas to the class of non-terrain areas as well as the necessity of taking also the gradients of surrounding segments of roof regions into account.

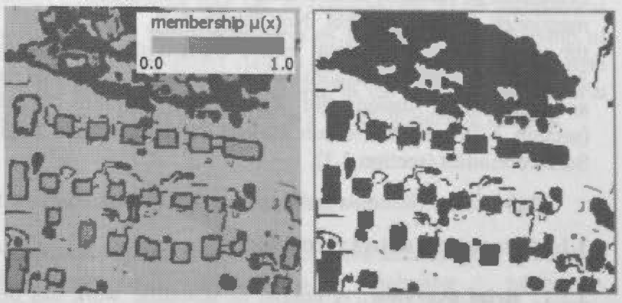


Figure 4. Intermediate results of terrain surface estimation: fuzzification based on "mean gradient" (left) and classification of non-terrain areas (in grey; right) - the test site is identical to that shown in figures 1 and 2

Comparing the achieved results for both tasks - normalization and separation of buildings and other objects - no omission errors have been found. The number of commission errors is below 2% of the entire number of segments. Critical regions within the terrain surface estimation process are small clearings within wooded areas whose slope behaviour in contrast to the surrounding trees could not be sufficiently separated. Applying the segmentation and the following classification not only at one, but at multiple scales, reduces the number of commission errors with decreasing generalization level. However, this scaling process must not be performed as far as possible because the number of omission errors would increase and in the extreme case one would end up with an undesired point-wise classification.

4. PROBLEMS

As already indicated the key problems of an operational use of segmentation algorithms for evaluating remotely sensed data are closely related to the specific aspects as pointed out in section 2.2. While the integration of multi-sensoral and multi-source data is comparably highly developed (with some deficiencies left; see e.g. Schiewe et al., 2001), the realization of the multi-scale and multi-method aspects are yet far away from maturity.

Varying the homogeneity thresholds it is no problem to generate segmentations at different levels of generalization. On the other hand a proper as well as an automatical choice of this level is still not possible and left to an iterative process of visual inspections and modifying the respective parameters. This problem is not only due to missing functionality within the segmentation software products but also to the difficult definition of generalization levels for given applications. It becomes obvious that corresponding work from the cartography domain (e.g., McMaster, 1991; Sester, 2000) has still to be transferred to segmentation algorithms. In particular, rules for the generalization operations of combining (agglomerating) elements have to be taken into account which obviously has to be done in a closed and recursive connection to the classification process.

Concerning the use of different segmentation methods and parameters for delineating different object types it has to be noted that only little achievements have been made so far. It is obvious that this idea implies a fusion of methods or results. In this context Clement et al. (1993) demonstrate a potential realization which still has to be investigated further and has to be transferred to other applications like the evaluation of high resolution and multi-source data. In order to minimize the complexity of the fusion a selection of a minimum number of significant features should be aimed for. Pinz et al. (1996) present their corresponding concept called *active fusion* that integrates prior knowledge and gives recommendations for the further control of the segmentation process.

In general, it is also desirable to perform the segmentation and the following classification in a hybrid rather in a purely data-driven or model-driven manner. In order to introduce human and GIS-based knowledge respective concepts (e.g., semantical nets) have to be linked to segmentation approaches in order to enable a better control and evaluation of the process.

5. CONCLUSIONS

Traditional multi-spectral classification methods on pixel basis are no longer suited for the evaluation of high-resolution and multi-source data from remote sensing. Region-based approaches consisting of a segmentation and a classification step have already proven to be a satisfying alternative solution.

From a conceptual point of view segmentation algorithms for the evaluation of remotely sensed data have to take into account in particular the availability of multi-source data as well as the need for multi-method and multi-scale functionality in order to model the heterogeneous objects under consideration in a flexible and adaptive way. While a feature level fusion of multi-source data is no severe problem anymore, the other two aspects still need a lot of research and development work.

The huge bandwidth of applications of segmentation approaches that has also been outlined here will certainly lead to further progress which is needed in order to use the full potential of the novel remotely sensed data.

REFERENCES

Arboreal, 2002. www.arboreal.fi, last accessed February 20, 2002

Bauer, T. & Steinnocher, K., 2001. Per-parcel land use classification in urban areas applying a rule-based technique. *GIS - Geo-Informationssysteme* 6: 24-27.

Brenner, C., 2000. Dreidimensionale Gebäuderekonstruktion aus digitalen Oberflächenmodellen und Grundrissen (Dissertation) Deutsche Geodätische Kommission, C530.

Cheng, T., 2002: Fuzzy objects: their changes and uncertainties. Photogrammetric Engineering & Remote Sensing 68(1), 41-49.

Clement, V., Giraudon, G., Houzelle, S. & Sandakly, F., 1993. Interpretation of Remotely Sensed Images in a Context of Multisensor Fusion Using a Multispecialist Architecture. *IEEE Transactions on Geoscience and Remote Sensing* 31(4): 779-791.

Definiens-Imaging, 2002. www.definiens-imaging.com, last accessed February, 20, 2002.

Ebner, H., Baumgartner, A. & Steger, C., 1998. Automatische Extraktion von Straßen aus Luftbildern. *Österr. Zeitschrift für Vermessung und Geoinformation.* (85)3: 177-186.

Hoffmann, P., 2001: Detecting buildings and roads from IKONOS data using additional elevation information. *GIS - Geo-Informationssysteme* 6: 28-33.

Hoover, A. et al., 1996. An Experimental Comparison of Range Image Segmentation Algorithms. *IEEE Transactions on pattern analysis and machine intelligence* 18(7): 673-689.

Levine, M.D. & Nazif, A.M., 1985. Dynamic Measurement of Computer Generated Image Segmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2): 155-164.

McKeown, D.M., Harvey, W.A. & McDermott, J., 1985. Rule Based Interpretation of Aerial Imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7(5): 570-585.

McMaster, R.B., 1991. Conceptual frameworks for geographical knowledge. In: Buttenfield, B.P. & McMaster, R.B. (Eds.): Map Generalization: Making rules for knowledge representation. Longman Scientific & Technical. 21-39.

Pinz, A., Prantl, M., Gaster, H & Kopp-Borotschnig, H., 1996. Active Fusion - A New Method Applied to Remote Sensing Image Interpretation. *Pattern Recognition Letters.* 17: 1349-1359.

Pohl, C. & van Genderen, J.L., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *Int. Journal for Remote Sensing* 19(5): 823-854.

Ryherd, S. & Woodcock, C., 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering & Remote Sensing* 62(2): 181-194.

Schiewe, J., 1999. An advanced technique for pixel-based multi-sensor data integration. In: Proceedings of the ISPRS-workshop "Sensors and Mapping from Space 1999", Hannover (Germany), CD-ROM.

Schiewe, J., 2001. Ein regionen-basiertes Verfahren zur Extraktion der Geländeoberfläche aus Digitalen Oberflächen-Modellen. *Photogrammetrie-Fernerkundung - Geoinformation.* 2: 81-90.

Schiewe, J., Tufte, L. & Ehlers, M., 2001. Potential and problems of multi-scale segmentation methods in remote sensing. *GIS - Geo-Informationssysteme* 6:34-39.

Sester, M., 2000. Maßstabsabhängige Darstellung in digitalen räumlichen Datenbeständen. Habilitation, University of Stuttgart, Institute for Photogrammetry.

TopoSys, 2002. www.toposys.com,last accessed February 20, 2002.

Wewel, F., Scholten, F. Neukum, G. & Albertz, J., 1998. Digitale Luftbildaufnahmen mit der HRSC – Ein Schritt in die Zukunft der Photogrammetrie. *Photogrammetrie–Fernerkundung – Geoinformation* 6: 337-348.

Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation. *Pattern Recognition* 29(8): 1335-1346.

A PHYSICS-BASED MODEL FOR ACTIVE CONTOURS: A COMPUTATIONAL ALGEBRAIC TOPOLOGY APPROACH

P. Poulin [a], M.-F. Auclair-Fortier [a], D. Ziou [a], M. Allili [b]

[a] Groupe de Moivre, Département de mathématiques et d'informatique, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada - (poulin, auclair, ziou)@dmi.usherb.ca
[b] Depts. of Comp. Sc. And Math., Bishop's University, Lennoxville, Québec, J1M 1Z7, Canada – mallili@ubishops.ca

KEY WORDS: Curve Deformation, Partial Differential Equations, Physical Laws, Computational Algebraic Topology, Active Contours, Shape-Based Image Retrieval

ABSTRACT:

We present a new method for the deformation of curve based upon a decomposition of the elasticity problem into basic physical laws. We encode the basic laws using computational algebraic topology. Each basic law uses exact global values and makes approximations only when they are needed. The deformations computed with our approach have a physical interpretation. Furthermore, our algorithm performs with either 2D or 3D problems. We finally present results validating our approach.

1. INTRODUCTION

The main problem addressed in this paper concerns the deformation of curves for which several methods have been proposed (Gibson, and Mirtich, 1997, Montagnat, Delingette, Scapel and Ayache, 2000). Physics-based methods (Platt and Badler, 1981, Sclaroff and Pentland, 1995) and methods arising from variational calculus (Bentabet, Jodouin, Ziou and Vaillancourt, 2001, Kass, Witkin, and Terzopoulos, 1987) have been considered to control the deformations.

The numerical solutions obtained with the existing techniques do not always have a physical interpretation. Those arising from variational calculus are usually based upon purely mathematical techniques. In this paper, we propose a new model which includes a physical interpretation of the deformations and offers a systematic method for reflecting some properties of the materials. It also proposes to use exact global values valid over a region instead of considering only local ones. To achieve it, we directly use the physical laws governing the deformations and encode them using a computational algebraic topology (CAT)-based image model. This model makes use of chains, boundaries, cochains and coboundaries to represent the image support and the quantities defined over it.

Our approach has several important advantages. 1) Since the linear elasticity problem is well-known in continuum mechanics then our modelling can be performed wisely and provide some good physical interpretation of the whole de-formation process and of its intermediate steps. 2) The use of global values leads to an algorithm which is less sensitive to noise. 3) The use of coboundaries to link the field quantities defined over pixels of different dimensions provides a modular computational scheme easy to reproduce. The coboundaries are defined according to the suitable basic laws prescribed by each problem. 4) Thanks to the modularity, the approach can easily be generalized to surfaces and volumes.

The approach has been validated in the context of the correction of synthetic and real images of road databases. In the following, and after decomposing the elasticity problem into basic laws, we summarize the CAT-based image model and propose a representation of the problem based upon it. This representation is then used to compute the regularization constraints which are employed to control the deformations. Applications and experimental results are then presented.

2. DECOMPOSITION INTO BASIC LAWS

The first basic laws which are relevant to us are the well-known strain-displacements relations (Poulin, Auclair-Fortier, Ziou, and Allili, 2001):

$$\varepsilon_{ij} = \frac{1}{2}\left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right] \qquad (i,j=1,2,3) \tag{1}$$

where u_i is the change in length of the original body in the direction of x_i.

Our next basic laws concern the force-stress relation. Given distributed body forces ρb_i and traction forces t_i^n applied to a body moving under the velocity field $v_i = v_i(x, t)$, the principle of linear momentum (Boresi, 1965) states that the resultant force acting on a body is equal to the rate of change over time of the linear momentum. Thus

$$\frac{d}{dt}\iiint_V \rho v_i dV = \iint_S t_i^n dS + \iiint_V \rho b_i dV \tag{2}$$

lin. momentum = Forces acting on the body

Using the Cauchy stress formula (Mase and Mase, 1999) :

$$t_i^n = \sum_{j=1}^{3} \sigma_{ji} n_j \tag{3}$$

and Gauss's divergence theorem, we have

$$\frac{d}{dt} \iiint_V \rho v_i dV = \iiint_V (\nabla \cdot \sigma_i + \rho b_i) dV \tag{4}$$

where $\sigma_i = (\sigma_{1i}, \sigma_{2i}, \sigma_{3i})^T$ and σ_{ij} is the stress component in the direction of x_j when the unit normal vector $n = (n_1, n_2, n_3)^T$ is parallel to x_i. Since the element of volume is arbitrary then the integral sign can be taken off and the second basic laws, also known as the equilibrium equations, are obtained assuming a zero velocity field in Eq.4:

$$\nabla \cdot \sigma_i = \rho b_i \tag{5}$$

In order to explain how the internal forces of an elastic body vary when it undergoes deformations, we must take into account a constitutive equation (or material law) which reflect the internal constitution of the materials. Our last basic law is known as the generalized Hooke's law for linear elastic isotropic materials in plane strain situations (Mase and Mase, 1999) and is given by the following equations:

$$\sigma_{ii} = \frac{\mathrm{E}}{(1+\nu)(1-2\nu)} \left[(1-\nu)\varepsilon_{ii} + \nu\left(\varepsilon_{jj} + \varepsilon_{kk}\right) \right] \tag{6}$$

$$\sigma_{ij} = 2G\varepsilon_{ij} = \frac{\mathrm{E}}{(1+\nu)} \varepsilon_{ij} \qquad (i \neq j) \tag{7}$$

where E is the modulus of elasticity of Young, ν the Poisson ratio and $G = \dfrac{E}{2(1+\nu)}$ the modulus of rigidity.

3. ENCODING OF THE BASIC LAWS USING THE CAT-BASED IMAGE MODEL.

An image is composed of two distinctive parts: the image support (pixels) and some field quantity associated with each pixel. This quantity may be scalar (e.g. grey level), vectorial (e.g. force) or tensorial (e.g. Hessian). We model the image support in terms of cubical complexes, chains and boundaries. For quantities, we introduce the concept of cochains which are representations of fields over a cubical complex. A detailed version of the CAT-based image model and some applications in image processing may be found in (Auclair-Fortier, Poulin, Ziou and Allili, 2001, Poulin, Auclair-Fortier, Ziou and Allili, 2001, Ziou and Allili, 2001).

To encode the basic laws of the elasticity problem, the image support is subdivided into cubical complexes. Then global quantities are computed over -pixels via cochains according to basic laws. The constitutive equations 6 and 7 are expressed as linear transformations between two cochains. Since we want to represent two kinds of global values over the image, we use two complexes. The first complex is associated with variables describing the configuration of the system (e.g. displacements) while the second complex refers to the source variables (e.g. forces) (Tonti, 2001).

The first laws (Eq. 1) are modeled as 1-cochains $< \mathcal{D}, \gamma_D >$ for a 1-pixel γ_D over a 3-complex $\mathcal{K}^P$ with $\partial\gamma_D = x^* - x\#$. Considering the normal strains ε_{11}, ε_{22} and ε_{33} in Eq.1 we have an application ε

$$\varepsilon\colon \mathbb{R}^3 \to \mathbb{R}^3$$

$$\mathrm{U} \mapsto \varepsilon(\mathrm{U}) = (\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33})^T = \nabla \mathrm{U} \tag{8}$$

Let us define the cochain $\mathcal{D}$ as

$$< \mathcal{D}, \partial\gamma_D > = \int_{\gamma_D} \varepsilon(\mathrm{U}) d\gamma_D \tag{9}$$

Let us apply a 0-cochain $\mathcal{V}$ to $\partial\gamma_D$

$$< \mathcal{V}, \partial\gamma_D > = \mathcal{V}(x_* - x_\#) = \mathcal{V}(x_*) - \mathcal{V}(x_\#)$$

and use the line integral theorem in Eq.9. Then we have

$$< \mathcal{D}, \partial\gamma_D > = \mathrm{U}(x_*) - \mathrm{U}(x_\#)$$

We thus define $\mathcal{V}(x) = \mathrm{U}(x)$. The cochain $\mathcal{D}$ is then the coboundary of $\mathcal{V}$.

The second laws (Eq. 5) are modelled as 3-cochains $< \mathcal{F}, \gamma_F >$ for a 3-pixel γ_F over a 3-complex $\mathcal{K}^S$

$$< \mathcal{F}, \gamma_F > = \int_{\gamma_F} \nabla \cdot \sigma_i d\gamma_F$$

Applying the divergence theorem to $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3)^T$, we have

$$< \mathcal{F}_i, \gamma_F > = \sum_{\gamma_{S_j} \in \partial_{\gamma_F}} \sigma_i \cdot n_j d\gamma_{S_j} \quad (10)$$

where n_j is the normal vector to γ_{S_j}. Then we define a 2-cochain $\mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3)^T$ such that

$$< S_i, \gamma_{S_j} > = \int_{\gamma_{S_j}} \rho_i \cdot n_j d\gamma_{S_j} \quad (11)$$

Then the cochain $\mathcal{F}$ is the coboundary of $\mathcal{S}$.

We finally need to represent the Hooke's law (Eqs 6 and 7) which links the local values of $\varepsilon(x)$ and $\sigma(x)$. In cochain terms, we want to link the cochains $\mathcal{D}$ and $\mathcal{S}$. Since Eqs 6 and 7 are constitutive equations, we cannot provide a topological expression of them. We use a piecewise approximation) and $\widetilde{\mathrm{U}}(x)$ of U(x) such that for each 1-face γ_D of a 2-pixel γ of $\mathcal{K}^P$ we have

$$\int_{\gamma_D} \widetilde{\varepsilon}(x) d\gamma_D = < D, \gamma_D > \quad (12)$$

Applying the generalized Hooke's law to $\widetilde{\varepsilon}(\mathbf{x})$ satisfying Eq. 12, we have

$$\widetilde{\sigma}_{ii}(x) = \mathrm{E}\left[\frac{(1-\nu)\widetilde{\varepsilon}_{ii}(x) + \nu\left(\widetilde{\varepsilon}_{jj}(x) + \widetilde{\varepsilon}_{kk}(x)\right)}{(1+\nu)(1-2\nu)}\right]$$

$$\widetilde{\sigma}_{ij}(x) = \frac{\mathrm{E}}{(1+\nu)}\widetilde{\varepsilon}_{ij}(x) \qquad (i,j = 1, 2, 3)(i \neq j)$$

at all point x of $\bar{\gamma}$. Eq. 11 is then replaced by

$$< S_i, \gamma_{S_j} > = \iint \widetilde{\sigma}_i(x) \cdot n dS \qquad (i = 1, 2, 3) \quad (13)$$

which depends on the choice of the approximation function $\widetilde{\varepsilon}(x)$.

4. APPLICATION

We apply our approach to a 2D active contours model based upon a Lagrangian evolution of a curve S

$$\frac{\partial S}{\partial t} + \mathbf{K}S = \mathbf{F}_{\mathrm{ext}}(S) \quad (14)$$

where **K** is a matrix computed with the basic laws of Section 2 and encoding the regularizing constraints. Assuming that the initial curve S_0 was in an equilibrium state and that the initial body forces $\mathbf{F_0} = \mathbf{K}S_0$ are constant during the deformation process, we can add these forces to the external forces $\mathbf{F}_{\mathrm{ext}}$ and solve Eq. 14 with an explicit scheme

$$S_{t+\Delta t} = S_t + \Delta t\left(\mathbf{F}_{\mathrm{ext}} - \mathbf{K}\mathbf{U}_t\right) \quad (15)$$

where $\mathbf{U}_t$ is the displacement vector for the curve S at t.

We want to solve Eq. 15 for local U(x) located at the center of each pixel and known initial curve S_0 close to the solution. To achieve it, we first position the two cubical complexes $\mathcal{K}^P$ and $\mathcal{K}^S$ such that the 0-pixels of $\mathcal{K}^P$ correspond to the center of the image pixels the 2-pixels of $\mathcal{K}^S$ coincide with the image pixels. This way of positioning $\mathcal{K}^S$ allows the use of an approximation polynomial of order 1 with the same accuracy as that obtained using one of order 2 (Mattiussi, 1997).

In order to solve Eq. 15, we need global values $\mathbf{F}_{\mathrm{ext}}$ over each pixel of $\mathcal{K}^S$. Since these values are generally known, we did not try to express them in a topological way in the last section. In our examples, we use the line plausibility image obtained using a line detector proposed by Ziou (Ziou, 2000).

We then choose an approximation function $\widetilde{\varepsilon}(x) = (\widetilde{\varepsilon}_{11}(x), \widetilde{\varepsilon}_{22}(x))^T$. For simplicity, we assume that $\widetilde{\varepsilon}(x) = \nabla\widetilde{\mathrm{U}}(x)$ arises from a bilinear approximation of $\mathbf{U}(\mathbf{x})$ and that it satisfies Eq. 12. Then for a 2-pixel of $\mathcal{K}^P$ (Fig. 1), we have

$$\widetilde{\varepsilon}(x) = \left[\frac{D_1}{\Delta} + \frac{D_3 - D_1}{\Delta^2}x_2, \frac{D_4}{\Delta} + \frac{D_4 - D_2}{\Delta^2}x_1\right]^T \quad (16)$$

In order to compute the cochain $< \mathcal{F}, \gamma_F >$, we use four approximation functions $\widetilde{\sigma}_i^a$, $\widetilde{\sigma}_i^b$, $\widetilde{\sigma}_i^c$ and $\widetilde{\sigma}_i^d$ because each γ_F intersects four 2-pixels of $\mathcal{K}^P$, (see Fig. 2). We find the value of the cochain over the four 1-faces of γ_F using Eq. 11:

$$S_i^1 = \int_0^{\frac{\Delta}{2}} \widetilde{\sigma}_i^A\left(\frac{\Delta}{2}, x_2\right) \cdot \vec{\mathrm{i}} dx_2 + \int_{\frac{\Delta}{2}}^0 \widetilde{\sigma}_i^B\left(\frac{\Delta}{2}, x_2\right) \cdot \vec{\mathrm{i}} dx_2$$

and so on for S_i^2, S_i^3 and S_i^4. Using Eq. 10 we have

$$<\mathcal{F}_i, \gamma_F> = S_i^1 + S_i^2 + S_i^3 + S_i^4 \quad (17)$$

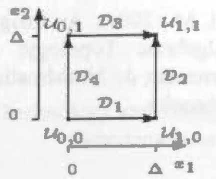
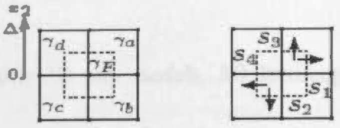


Figure 1. A 2-pixel of $\mathcal{K}^p$ and the topological quantities associated with it.



(a) (b)

Figure 2. (a) γ_F in dashed lines (b) 2-cochain S.

Eq. 17 expresses the internal forces F_1 and F_2 as a function of the displacement **U**. As an example, we present the values of F_1 and F_2 for the 2-pixel γ_F of Fig. 1 with Δ=1:

$$\begin{aligned} F_1 = C\Big[& (3-4\nu)u_{-1,1} + (2-8\nu)u_{0,1} + (3-4\nu)u_{1,1} \\ & + (10-8\nu)u_{-1,0} + (-36+48\nu)u_{0,0} + (10-8\nu)u_{1,0} \\ & + (3-4\nu)u_{-1,-1} + (2-8\nu)u_{0,-1} + (3-4\nu)u_{1,-1} \\ & - 2\mathrm{v}_{-1,1} + 2\mathrm{v}_{1,1} + 2\mathrm{v}_{-1,-1} + -2\mathrm{v}_{1,-1}\Big] \end{aligned} \quad (18)$$

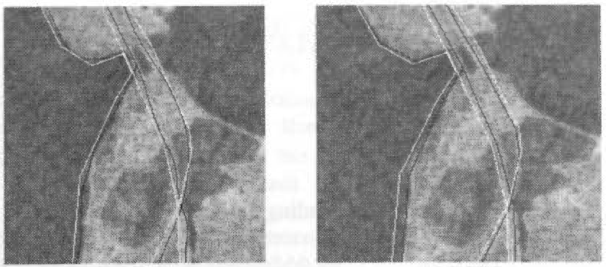
$$\begin{aligned} F_2 = C\Big[& (3-4\nu)\mathrm{v}_{-1,1} + (10-8\nu)\mathrm{v}_{0,1} + (3-4\nu)\mathrm{v}_{1,1} \\ & + (2-8\nu)\mathrm{v}_{-1,0} + (-36+48\nu)\mathrm{v}_{0,0} + (2-8\nu)\mathrm{v}_{1,0} \\ & + (3-4\nu)\mathrm{v}_{-1,-1} + (10-8\nu)\mathrm{v}_{0,-1} + (3-4\nu)\mathrm{v}_{1,-1} \\ & - 2u_{-1,1} + 2u_{1,1} + 2u_{-1,-1} - 2u_{1,-1}\Big] \end{aligned} \quad (19)$$

where $C = \dfrac{E}{16(1+\nu)(1-2\nu)}$ and $\mathbf{U} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}$.

Eqs 18 and 19 set a linear relationship between a pixel and its neighbors. This relation is used to build the stiffness matrix **K** of Eq. 15.

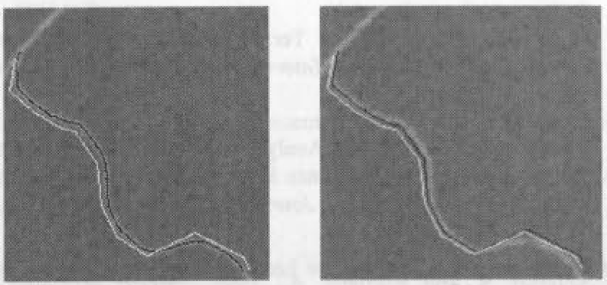
5. EXPERIMENTAL RESULTS

The proposed approach was first used to develop a shape-based image retrieval system (Poulin, Auclair-Fortier, Ziou and Allili, 2001). It was also tested in the context of the correction of high-resolution real and synthetic images of road databases. This section presents some results for this application. We compare the results obtained with our physics-based method (PBM) with those obtained with a finite element method (FEM) (Bentabet, Jodouin, Ziou and Vaillancourt, 2001).



(a) FEM (b) PBM

Figure 3: Road corrections.



(a) (b)

Figure 4: Shape recovery of a curve when the image forces are taken off.

The first image is a section of a RADARSAT SAR image centered on the region of Stephenville (Québec), Canada* with a 25-meter resolution. In this image, the roads correspond to linear features. Figs 3(a) and 3(b) show respectively in black the results obtained with the PBM (E=200 and ν=0.45) and FEM (α=0.03 and β=137.8 for the most left curve). The initialization snake is drawn in white in both figures. One can notice that the PBM most left corrected curve is closer to the exact road than the FEM one in regions of high curvature.

The deformations modeled using the PBM have the interesting property of allowing the objects to recover their original shape when the image forces applied to them are taken off. To illustrate this fact, Fig. 4(a) presents a synthetic image in which the initial and corrected roads are respectively drawn in white and black. Fig. 4(b) shows respectively in black and white the corrected curve when the image forces are taken off and the initial snake. Since the initial body forces $\mathbf{F}_0$ are the only forces applied to the deformed curve, then it recovers its initial shape. However, one can show that the computation of the internal forces is invariant with respect to translation and then the curve can experience a spatial shift as in Fig. 4(b).

This paper presented a new model for the deformation of curves. The proposed approach makes use of both local and global values through a decomposition of the elasticity problem into basic physical laws. These laws are encoded by cochains over cubical complexes and linked together using coboundaries. The coboundary formalism is a generic one and then it leads to algorithms which can easily be extended to higher dimensions. The use of the basic laws provides a physical interpretation of the deformation process. This interpretation was used to develop an active contours model.

---

[*] The image was provided by the Canadian Space Agency under the ADRO-2 program.

ACKNOWLEDGEMENTS

This research was partly founded by the National Sciences and Engineering Research Council of Canada and by the Fonds Québécois de la Recherche sur la Nature et les Technologies. The authors also wish to thank Olivier Delalleau for his contribution in the understanding of the physical model.

REFERENCES

References from Journals:
Kass, M., Witkin, A. and Terzopoulos, D., 1987. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1:321–331.

Mattiussi, C., 1997. An Analysis of Finite Volume, Finite Element, and Finite Difference Methods Using Some Concepts from Algebraic Topology. *Journal of Computational Physics*, 133:289–309.

Sclaroff, S. and Pentland, A., 1995. Modal Matching for Correspondance and Recognition. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(6):545–561.

Tonti, E., 2001. A Direct Discrete Formulation of Field Laws: The Cell Method. *CMES-Computer Modeling in Engineering & Sciences*, 2(2):237–258.

References from Books:
Boresi, A., 1965. *Elasticity in Engineering Mechanics*. Prentice-Hall.

Mase, G. T. and Mase, G. E., 1999. *Continuum Mechanics for Engineers*. CRC Press.

Platt, S. and Badler, N., 1981. Animating Facial Expressions. *Computer Graphics*, 15(3):245–252.

References from Other Literature:
Auclair-Fortier, M.-F., Poulin, P., Ziou, D. and Allili, M., 2001. A Physics-Based Resolution of Diffusion and Optical Flow: A Computational Algebraic Topology Approach. Technical Report 269, Département de mathématiques et d'informatique, Université de Sherbrooke.

Bentabet, L., Jodouin, S., Ziou, D., and Vaillancourt, J., 2001. Automated Updating of Road Databases from SAR Imagery: Integration of Road Databases and SAR Imagery information. information. In *Proceedings of the Fourth International Conference on Information Fusion*, volume WeA1, pp. 3–10.

Gibson, S. and Mirtich, B., 1997. A Survey of Deformable Modeling in Computer Graphics. Technical report, Mitsubishi Electric Research Laboratory.

Montagnat, J., Delingette, H., Scapel, N. and Ayache, N., 2000. Representation, Shape, Topology and Evolution of Deformable Surfaces. Application to 3D Medical Imaging Segmentation. Technical Report 3954, INRIA.

Poulin, P., Auclair-Fortier, M.-F., Ziou, D. and Allili, M., 2001. A Physics-Based Model for the Deformation of Curves: A Computational Algebraic Topology Approach. Technical Report 270, Département de mathématiques et d'informatique, Université de Sherbrooke.

Ziou, D., 2000. Optimal Line Detector. In *International Conference on Pattern Recognition (ICPR'00)*, Barcelona, SPAIN, September 3-8.

Ziou, D. and Allili, M., 2001. An Image Model with Roots in Computational Algebraic Topology: A Primer. Technical Report 264, Département de Mathématiques et d'informatique, Université de Sherbrooke.

DESIGN AND IMPLEMENTATION OF THE HIGH-PERFORMANCE 3D DIGITAL LANDSCAPE SERVER 'DILAS'

Stephan Nebiker

Basel Institute of Technology and Management / Fachhochschule beider Basel (FHBB), Department of Geomatics Engineering, Gründenstrasse 40, CH-4132 Muttenz, SWITZERLAND, s.nebiker@fhbb.ch

Commission IV, WG V/5

KEY WORDS: 3D city models, 3D landscape models, DBMS, multi-resolution, multi-representation, real-time visualisation, internet

ABSTRACT:

Recent advances in the fields of 3D data capturing, storage capacities, database technologies and web-based 3D-visualisation are gradually enabling the establishment and exploitation of large 3D landscape and city models on a regional or even national scale. These landscape models will enable exciting new information and entertainment services in the Internet. However, there is a lack of solutions for managing, updating and serving such large 3D worlds. The research project dilas (Digital Landscape Server) presented in this paper addresses these issues.

This paper presents concepts addressing the vastly contrasting requirements of very large spatial objects, on the one side, and large numbers of complex and possibly dynamic 3D objects on the other. The chosen modelling concept for 3D objects combines a versatile 3D object model, multi-representation and object serialisation. This information is managed in a commercial object-relational DBMS using a hybrid data model. The results of the research project are currently being integrated into the commercial version of dilas – one of the first fully scalable solutions for the management and visualisation of very large 3D landscape models.

KURZFASSUNG:

Fortschritte in den Bereichen 3D-Geodatenerfassung, Speicherkapazitäten, Datenbanktechnologien und in der webbasierten 3D-Visualisierung ermöglichen mittlerweile den Aufbau regionaler bis nationaler 3D-Landschaftsmodelle und Stadtmodelle. Diese virtuellen Landschaftsmodelle werden eine Reihe neuer Informations- und Unterhaltungsdienste im Internet ermöglichen. Zur Zeit fehlen jedoch noch Lösungen zur effizienten Verwaltung und Nachführung derart grosser 3D-Landschaftsmodelle.

Die in diesem Beitrag beschriebenen Konzepte ermöglichen Lösungen für die stark unterschiedlichen Anforderungen sehr grosser Geoobjekte einerseits und sehr vieler, komplexer 3D-Geobjekte andererseits. Das vorgestellte und erfolgreich implementierte Lösungskonzept für 3D-Geoobjekte kombiniert ein leistungsfähiges 3D-Objektmodell mit einer Mehrfachrepräsentation und einer Objektserialisierung. Für die Speicherung dieser 3D-Geoobjekte kommt ein kommerzielles, objekt-relationales Datenbankverwaltungssystem Einsatz. Die Forschungsergebnisse werden zur Zeit in die kommerzielle Version von dilas integriert – einer der ersten voll skalierbaren Systemlösungen zur Verwaltung und Visualisierung beliebig grosser 3D-Stadtmodelle.

1. INTRODUCTION

1.1 Motivation

Increasingly efficient 3D data capturing methods are boosting the creation of city-wide, regional or even national 3D landscape and city models. These 3D models provide an ideal basis for future spatial information and entertainment services in the (wireless) Internet (Nebiker, 2001). The landscape and city models, which enable these services, are reaching an enormous size and an increasing complexity. Thus, the efficient management and web-based visualisation of such large 3D models incorporate a number of major challenges. One is the maintenance and updating of very large 3D geodatabases, another is the streaming and visualisation of 3D data over networks with still very limited bandwidths.

The representation of 3D landscape and city models requires a variety of spatial data types. These include orthoimagery and terrain data, vector-based 3D and 2D geo-objects, object textures, 3D scene objects, animations and hyperlinks. These data types have very different characteristics and requirements in terms of management and visualisation. The spectrum ranges from very large spatial objects, such as orthoimagery and high-resolution DTM data, with data volumes in the order of Terabytes to large numbers of complex and possibly dynamic 3D objects.

1.2 Related work

A number of ongoing research projects in the field of landscape management and 3D-GIS are either focussing on the management of very large imagery and DTM databases, or on the modelling and management of complex 3D objects, e.g. (Pfund, 2001), (Zlatanova and Tempfli, 2000) or (Wang, 2000). The latter projects address the aspects of 3D modelling and 3D topology, but they typically revert to conventional data management concepts based on relational database models. However, the relational model is not well suited for managing complex spatial objects and does not provide the scalability required to efficiently manage and serve large numbers of 3D objects as they are typically encountered in regional or national 3D landscape models.

1.3 Aim

The goal of the dilas project is to identify the main requirements in 3D landscape management and to develop a prototype system integrating these diverse and contrasting requirements into an operational and truly scalable solution. The dilas project succeeds the GRIDS project (Nebiker, 1997), (Nebiker and Relly, 1999) and builds on the since commercialised GRIDS Server (GEONOVA, 2001). The "dilas" project is funded by industry partners and by the Swiss Federal Office for Professional Education and Technology.

2. DILAS CONCEPTS

2.1 Overview

The following section highlights some of the key concepts developed and implemented as part of the dilas project: a flexible 3D object model, a multi-representation and multi-resolution approach for the different object types, a storage concept for 3D and raster objects and XML-based process rules.

2.2 3D Object Model

One of the key concepts of the dilas project is a generic, fully object-oriented model for 3D geo-objects. This object model incorporates a topologically structured 3D geometry model, which supports most basic geometry types. The geometry model is based on a 3D boundary representation. It incorporates the capability for multiple levels of detail (LOD) as well as texture and appearance information required by 3D visualisation packages. The 3D object model is suitable for representing any spatial topic (e.g. buildings, bridges, power-lines).

The dilas 3D object type is supplemented by a number of spatial data types used for representing very large mosaics of high resolution terrain and texture data:

- raster maps
- orthoimagery
- terrain and surface models (regular grids)
- terrain and surface models (irregular point clusters), e.g. for managing very large laser scanning height data sets

2.3 Multi-Representation and Multi-Resolution

Two key issues in the efficient management and visualisation of large 3D models are multiple representation and multiple resolution. Different multi-representation strategies were developed for the spatial object types used in dilas. The original multi-resolution approach for managing very large raster mosaics (Nebiker, 1997) was further refined and extended to all mosaic types listed above.

3D objects are represented by 3D bounding boxes, 2D object boundaries and the actual 3D geometry. The first two representations are essential for efficient query operations and are automatically derived from the main 3D representation.

2.4 Storage Concept for 3D Objects

The goal for handling and manipulating 3D objects was to provide an optimum modelling flexibility in combination with an excellent object query and retrieval performance. The developed concept is based on the following components:

- a 3D object representation in Java and XML
- a 3D object serialisation and de-serialisation
- a persistence framework built on top of the DBMS
- spatial data types for 3D and raster objects within an object-relational environment

A number of these mechanisms are adapted from modern object-oriented programming environments. The object serialisation approach, for example, permits to map very complex objects to a simple, but highly efficient storage mechanism. The storage mechanism is based on a type extension for 3D objects which encapsulates the actual LOB-based object storage. Attributes which are frequently queried and accessed are automatically extracted from the serialised objects and stored in relational or object-relational form. With the combined use of object serialisation and LOBs for storing 3D objects the expensive object reconstruction process inherent to the relational model can be avoided. This leads to a performance improvement in the order of magnitude.

The persistence framework developed in dilas adapts concepts from the Java Data Objects (JDO) extension. It permits a fine-grained control over changes to the 3D object properties.

2.5 XML-based Process Rules

The processes of importing, structuring, generating and validating 3D city models are quite complex and typically differ from organisation to organisation, e.g. different level assignments, exchange of geometry only versus exchange of actual 3D objects etc.

The goal of accommodating these diverse requirements led to the development of a mechanism using 'XML-based process rules'. The benefits of this rule-based approach are:

- The possibility of formally specifying valid processing options (e.g. data import options) through the means of different XML Schemas.
- The easy adaptation of process rules or the creation of new process rules by a project leader or system administrator and the possibility of easily integrating these rules into the user interface.
- A rigorous validation of user-defined process rules by means of standard XML tools and mechanisms.

3. DILAS SYSTEM ARCHITECTURE

One of the design goals of the dilas project was to rely on state-of-the art commercial database technologies. The current prototype system is using an Oracle 9i DBMS. The dilas prototype system consists of the modules dilas Server, dilas Manager, dilas 3D Modeler and dilas Scene Generator (Nebiker, 2002).

3.1 dilas Modules

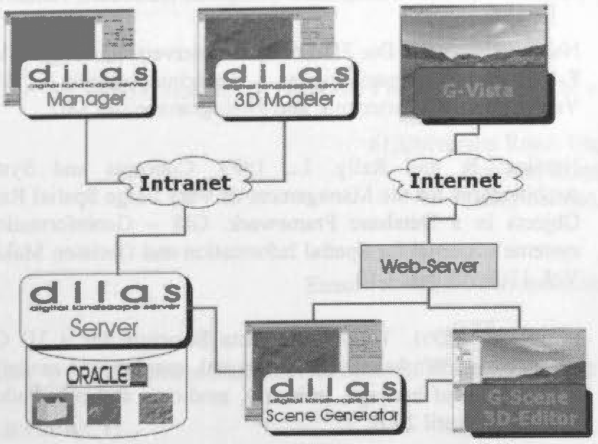


Figure 1: dilas System Architecture.

The Server and Manager module make up the core components of the system, which address the aspects of storage management, 3D scene management and querying, representation mapping as well as 3D scene export and import.

The 3D Modeler component is built into MicroStation V8, the latest CAD version of Bentley Systems. The 3D Modeler module maps between the dilas 3D object model and the MicroStation V8 geometry model. Through the MicroStation Java API dilas 3D Modeler has full access to the CAD geometry model and to the abundance of construction and import/export functionality available within MicroStation V8. Currently, dilas 3D Modeler incorporates functionality for the editing of 3D objects, the automatic generation of 3D buildings from roof models or 2D map data and for the interactive texturing of 3D objects.

The dilas Scene Generator plays a key role in enabling the web-based visualisation of very large landscape and city models using the high-performance 3D-visualisation software G-Vista (GEONOVA AG). dilas Scene Generator generates web-based multi-gigabyte 3D scenes with large numbers of 3D objects.

3.2 The Integration of 2D and 3D

One of the key factors in making 3D city models and landscape models a technical and commercial success will be the integration of 3D landscape management solutions with existing 2D GIS environments.

In dilas this 3D-2D integration is achieved by adapting the OGC Simple Feature data model and by extending it with the spatial data types listed in the previous section. This approach yields a number of benefits:

- the vast amounts of existing 2D geodata can also be accessed and exploited in 3D
- the 3D geometry, for example, can be treated as a spatial attribute of a conventional 'GIS feature'
- the 2D representation of a 3D object is visible as a read-only attribute in any OGC SF compliant GIS

4. RESULTS

4.1 Feasibility Studies

The dilas project was preceded by a number of studies by (Eugster and Henz, 2001) and (Schaad and Schärer, 2001) which proved the feasibility of the 'dilas approach' and its superior performance and scalability over conventional data management solutions.

4.2 Management and Generation of Large 3D Landscapes

In June 2001, the project team and its industry partners GEONOVA and G-Graphix launched probably the world's first interactive web-based visualisation of a nation-wide 3D scene with resolutions down to the sub-meter level (Gibbs, 2001). The 3D scene for this "Flight through Switzerland" was generated using the first version of dilas Scene Generator. Since then a number of new 3D geoinformation services have been launched using results from the dilas project.

The service shown in Figure 2 features a web-based 3D visualisation of the 2002 Winter Games in Salt Lake City – embedded in a 3D landscape of the entire State of Utah. Key components of the dilas project are now used to manage and generate web-based 3D scenes with texture and height data in the range of hundreds of Gigabytes in size.

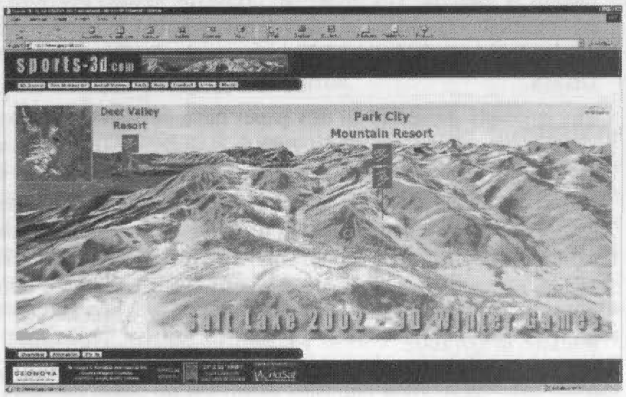


Figure 2: 3D geoinformation service featuring the 2002 Winter Games of Salt Lake City (www.sports-3d.com)

4.3 Management and Generation of 3D City Models

The fact that dilas integrates the entire 3D city models in a single DBMS environment offers a range of new possibilities in the efficient generation of 3D city models. The following building generation models are supported at the moment:

- automatic building model generation based on 3D roof geometry information and on a digital terrain model
- automatic building model generation using building ground plane information (e.g. 2D building outlines from cadastral data sources) together with a digital surface model and a digital terrain model

The latter approach permits a highly efficient and fully automatic generation of very large city models in cases where surface models, typically from airborne laser scanning, are available (see Figure 3).

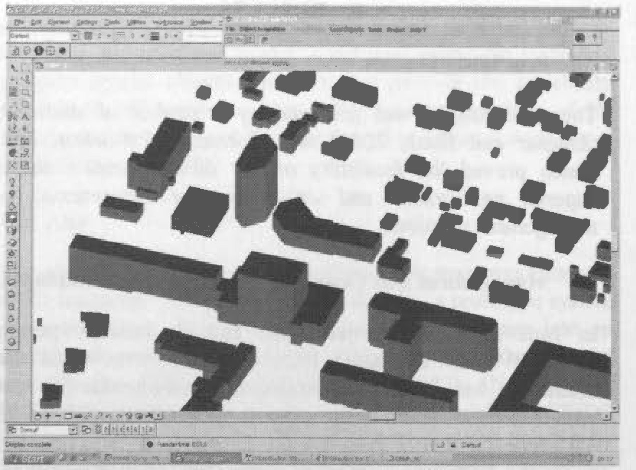


Figure 3: *dilas 3D Modeler* with a section of a 3d city model which was automatically derived using ground plane data together with a DTM and DSM (source data: Service des systèmes d' information et de géomatique (SSIG), Genève)

5. CONCLUSIONS

The dilas project has shown that an efficient, flexible and highly integrated management of very large 3D landscape and city models is feasible. The presented prototype system provides an environment supporting all data types required to handle large 3D landscape and city models within a single object-relational DBMS.

The concepts presented in this paper have been evaluated on a number of real-world projects with thousands of 3D objects and very large texture and height data sets. In these projects the feasibility of the 'dilas approach' and its excellent performance and scalability in comparison with conventional data management solutions could be clearly demonstrated. The results of the dilas project are currently being integrated into a commercial 3D landscape server which is expected to be released in Q2/2002.

Ongoing and future research activities are focusing on the streaming of 3D object data to a large number of mobile clients and on the optimal handling of LODs for 3D objects. Further important aspects affecting the dilas project are the future definition of 3D types in the OGC SF framework.

6. REFERENCES

Eugster, H. and Henz, T., 2001. Geodatenverwaltung in Oracle 8i auf der Basis von XML. Diploma Thesis, FHBB Fachhochschule beider Basel, Muttenz.

GEONOVA, 2001. GRIDS - Geospatial Raster and Image Database System. GEONOVA AG, Muttenz, Switzerland (http://www.geonova.ch).

Gibbs, W.W.N., 2001. A Wide Web of Worlds. Scientific American, October 2001.

Nebiker, S., 1997. Spatial Raster Data Management - A Database Perspective. PhD Thesis, ETH Zürich, Zürich, 179 pp.

Nebiker, S., 2001. dilas (Digital Landscape Server) - Management and 3D-visualisation of large terrain data sets, OEEPE/ISPRS-Workshop 'From 2D to 3D - Establishment and Maintenance of National Core Geospatial Databases, Hannover.

Nebiker, S., 2002. Die 3D-Landschaftsservertechnologie dilas - Ein anwendungsorientiertes Forschungsprojekt, VPK Vermessung, Kulturtechnik und Photogrammetrie, 3/02.

Nebiker, S. and Relly, L., 1999. Concepts and System Architectures for the Management of Very Large Spatial Raster Objects in a Database Framework. GIS – Geoinformationssysteme - Journal for Spatial Information and Decision Making, Vol. 12(4/99): pp. 4-10.

Pfund, M., 2001. Topological Data Structure for a 3D GIS, Italy-Canada Workshop on 3D Digital imaging and modeling, applications of heritage, industriy, medicine & land., Padova, Italy, 3.-4. April 2001.

Schaad, G.-R. and Schärer, P., 2001. Webbasierte Geodatenvisualisierung auf der Basis von XML und SVG. Diploma Thesis, FHBB Fachhochschule beider Basel, Muttenz.

Wang, X., 2000. CyberCity Spatial Information System (CC-SIS) - Project Description. http://www.geod.ethz.ch/p02/projects/3DGIS/CC-SIS_intro.html.

Zlatanova, S. and Tempfli, K., 2000. Modelling for 3D GIS: Spatial Analysis and Visualisation through the Web. International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B4, Amsterdam 2000.

7. ACKNOWLEDGEMENTS

The author would like to thank the Commission of Technology and Innovation (KTI) of the Swiss Federal Office for Professional Education and Technology for the support of this project[1]. Thanks are also due to the industry partners GEONOVA AG, G-Graphix and Flotron AG for their support and close collaboration.

---

[1] KTI-Project Nr. 5195.1 'dilas – Digital Landscape Server'

SEARCHING AND BROWSING INTERFACE FOR IMAGE DATABASES VIA INTERNET

Zeal Su[1] Jung-Hong Hong[2]

1.Ph.D. Student 2. Associate Professor Department of Surveying Engineering, National Cheng Kung University

#1,University Road, Dept. of Surveying Engineering
National Cheng-Kung University
Tania, Taiwan, Republic of China

Email : zeal@www.sv.ncku.edu.tw junghong@mail.ncku.edu.tw

KEY WORDS: Image Database□Browsing□Human-Computer Interface□Spatial Cognition

ABSTRACT:

With its easier availability, finer resolution and wider coverage of spectrum in recent years, remote sensing technology has emerged as a fast way for users in a variety of domains to acquire updated information. With such popularity, the needs nowadays are much greater than before, and we are even expecting an era that thousands of images may be produced per day. Although such a large number of images can certainly bring huge impacts to help us realizing the world, the accumulated data volume will sooner or later become a bottleneck for future image database. The idea of this research is to introduce human spatial concepts towards the selection of images from image databases. The core of suggested mechanism is the formalization of spatial relations based on humans' spatial cognition. With the aids of spatial browsing capability, images that covering interested area can be selected and later retrieved. This paper integrates knowledge for image selection and human spatial concept, which is just the beginning of this still-growing research topic.

1. BACKGROUND

With the technique improvement on easier availability, finer resolution and wider coverage of spectrum in recent years, remote sensing images have emerged as the most efficient way for users in a variety of domains to acquire updated information about the earth. Thousands of images from a variety of satellite platforms are produced and available via Internet nowadays. The fast development of Internet technology provides an extremely powerful environment for the sharing and distribution of remote sensing images (and of course the number of images may be overwhelming for users to handle). There are many famous image databases on Internet, such as USC-SIPI Image Database and NASA Image eXchange (NIX) etc. These databases all contain an enormous volume of images and the number of images keeps on growing everyday. Although such a large number of image databases provide a more complete coverage to the earth, it is more and more difficult for users to search images from such abundant data source. We proposed to introduce human spatial concepts into the image selection and browsing process in this paper, the ultimate goal is to reduce the difficulty of dealing with overwhelming amount of images.

Image database is an important and valuable data source to GIS. An image database is often a complex data search mechanism, such that users have to be well trained before they can operate the database management system effectively. Even images satisfying users' constraints are found, the selection of the most appropriate images may be another obstacle. An interface based on the analysis of users' spatial searching and browsing behaviour is certainly helpful to reduce the complexity of human-computer interaction and consequently improve the querying efficiency of remote sensing images via Internet. . For example, the request of images about the city of New York can be expressed as 'Select images that *cover* the city of New York.' Since spatial relationships, such as 'cover' in the above example, are frequently used in our daily lives, it would be easier for users to express such type of queries.

The basic concept of this paper is to incorporate knowledge of human spatial cognition into the interface for image selection and browsing. The core of the suggested mechanism is therefore based on the computer formalization of spatial relations that is applicable in the image searching and browsing process. For the current huge image archives, a typical situation is too many images may qualify users' constraints and it may require users a tremendous amount of time to visually inspect all the images before a conclusion can be reached. A 'smart' browsing interface based on the image selection knowledge from domain expertise can reduce the time required. For example, an image completely cover the interested area should be given a higher priority than those images only covering a part of the interested area. This suggests that the image mechanism should be capable of distinguishing the spatial relationships of 'entirely contain' from 'partly contain', though theoretically their formalizations are all based on the geometric intersection of two polygons (an image and an interested area). Therefore, users can still use 'cover' as their constraint, but the system should be able to act 'intelligently' to provide the queried result for further browse.

We are still in the early stage of this research and this paper can only present some preliminary results of the developed mechanism. Though far from practical use, we wish the discussion here can set up a foundation for our future research. Section 2 discusses the related analysis on spatial searching and browsing behaviours. Section 3 explains the basic mechanism of our system and section 4 demonstrates some test results. Finally section 5 conclude our major findings and explore possible developments in the future.

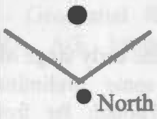
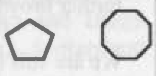
2. SPATIAL BROWSING BEHAVIOUR

For computers to 'understand' what humans want and act accordingly, it is clear further investigation on human knowledge modelling and spatial concept formalization is necessary. Human behaviour analysis and knowledge acquisition is therefore an urgent demand (Sestito & Dillon, 1994). The knowledge formalization and implementation in computers must include how data is represented and organized into considerations. Frank (1992) divided the creation of spatial data into three levels: spatial concept abstraction, spatial data modeling and spatial data structure. While humans may largely apply their spatial knowledge in the conceptual level (e.g., image schemata), the execution in computers is nonetheless often in the data structure level. It is therefore necessary to consider how to incorporate human spatial cognition into the computer system (Frank and Mark, 1992).

Humans frequently use spatial relationships to describe spatial phenomena in their daily lives (Egenhofer and Herrin, 1993). These spatial descriptions can be as simple as a simple spatial predicate, e.g., North of, or as complicated as the combination of several types of spatial relations, e.g., close + North of + disjoint. While humans usually express spatial relations in qualitative concept, most GIS systems are quantitative-based and fail to directly support the processing of qualitative spatial relations. An appropriate mapping between the two and mathematical formalization are necessary for any mechanism that tries to bring qualitative spatial relations into it user interface or query processing. The advantage is, however, if a system would allows users to express their queries in qualitative spatial relations, users can communicate with this system easier with spatial knowledge they already possess. Nevertheless, it is worthy to note even though humans can communicate with each other with qualitative spatial relations well, computer systems don't have such a spatial knowledge for interpretation and reasoning.

Frequently used spatial relation can be classified into three major categories: metric, direction and topology relations (Table 1).

Table 1.Spatial Relations types

Type	Metric	Direction	Topology
Description	Distance relation between A and B	Direction relation between A and B in specific direction system	Describing the related layout relation between A and B
Example	"It is Far between New York and Los Angle"	"Canada is North of United States"	Taiwan and Japan is disjoint
Graph	Far	 North	 Disjoint

While most DBMS deals with the 'search' behaviour of data satisfying users' constraint, 'browse' behaviour draws more and more attention with the rapidly increasing amount of data on the Internet. The Merriam Webster Dictionary define the word 'browse' as "to look over or through an aggregate of things casually especially in search of something of interest." Stephenson □ 1988 □ classifies the behaviour of human-computer interactions based two considerations:

- If users understand exactly what they need.
- If users know to acquire data from the system

Stephenson defined altogether six types of HCI behaviours: retrieving, searching, browsing, exploring, scanning and wandering. Figure 1 illustrates these six behaviors based on the above two considerations. Among them, only retrieving, searching and browsing are of major interests in this research. According to Stephenson's definition, these three behaviors are:

Retrieve: users know what data is and also know how to get the data through system interface.

Search: users know what data they want, but do not know exactly how to operate systems to get the data.

Browsing: users are not quite sure what they need and cannot specify an 'exact' constraint. The system must help users to narrow down the range of possible data.

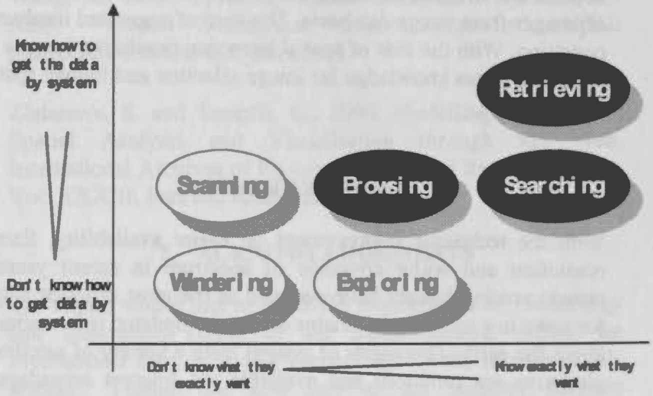

Know how to get the data by system
Retrieving
Scanning
Browsing
Searching
Don't know how to get data by system
Wandering
Exploring
Don't know what they exactly want
Know exactly what they want

Figure 1. Interaction models between human and computer □Stephenson ,1988□

When users query an image database, the search of images satisfying users' constraints is completed by DBMS and retrieved for further visual inspection. The huge volume of image data will be a problem even for database administrators. Since the volume of qualifying images may be overwhelming, it becomes an important issue for systems to provide an effective searching and browsing mechanism to help users to precisely specify their requests and quickly browse images that may be of interests. The design of 'browsing environment' is based on the following two major principles:

a) Help users to identify required data step by step.
b) Provide a mechanism to sort qualified images based on image selection knowledge.

3. IMAGE DATABASE AND BROWSING MECHANISM

Every remote sensing image vendor has his own design on the images provided. The basic common characteristics of remote sensing images include:

(1) Specification
Every type of image has its own design on such characteristics as format, sensors, spectral bands, satellite orbits, etc. The specification of images may influence if this type of images is

chosen. For example, when requesting an image providing an overview of Taiwan, it is no necessary to search aerial photos because of their relatively small size when compared to the area of Taiwan.

(2) Spatial coverage
Every image has its own spatial coverage on the ground. Any image search must be based on the 'geometric intersection' test between the area of interests and the spatial coverage of images.

(3) Time
Every image is acquired at a specific time. Many image queries include temporal constraints, such as 'find images that is taken after the 921 earthquake.'

Note what we really need is the content of images, the above description merely provide us a way to select images we need (e.g., images cover the city of New York). Unfortunately, users can only determine if an image satisfying his or her needs only after they have a chance to visually inspect the image. Besides, the information is hidden in the images, users must have a capability to correctly interpret images before they can determine if the image can be used. More help from systems to aid users' interpretation is certainly preferred, for example, superimpose the boundary of the area of interests on the image.

Figure 2 shows the system framework of our proposed system. Four important concepts will be described in the following:

- Image database is a physical storage device for storing raw images (and possibly their re-sampled, low-resolution images). Users can retrieve re-sampled or full-sized images at their will for further visual inspection.
- The image metadata database provides all the necessary information about image to the searching and browsing mechanism, so that the qualifying images can be reduced to a reasonable number before visual inspection begins.
- A spatial searching and browsing mechanism receive constraints created in spatial browser, translate the constraints to appropriate spatial predicate based on spatial knowledge, request metadata based on knowledge about image, then determine the images qualified.
- The user interface for browsing mechanism allows users to specify their spatial constraints based on their spatial concepts. The queried result is passed to the interface for users to make further choice.

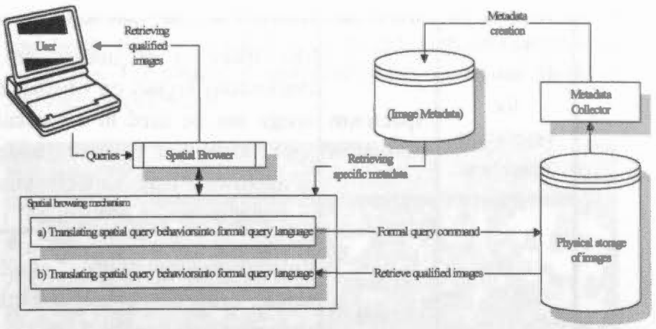

User
Retrieving qualified images
Metadata creation
(Image Metadata)
Metadata Collector
Queries
Spatial Browser
Retrieving specific metadata
Spatial browsing mechanism
a) Translating spatial query behaviorsinto formal query language
Formal query command
b) Translating spatial query behaviorsinto formal query language
Retrieve qualified images
Physical storage of images

Figure 2. Proposed system structure

3.1 Image Metadata

A general definition of metadata is "data about data". It is the description information about existing data used to help users to have more understanding toward the data before they really use it. No matter it is a long statement or a list of elements, an ideal metadata must be able to clearly explain the basic nature of the data it describes. Table 2 lists the suggested image metadata elements in our system:

<table>
<tr><th colspan="2">Type</th><th>Discussions</th></tr>
<tr><td colspan="2">Spatial information</td><td>Spatial metadata records the ground coverage of the respective image. This information can be used to calculate spatial relations□metric, direction and topology relations □ between interested region and image. Though the coverage specification may vary from one type of image to another, as long as the coverage is a rectangle, it can be represented by the coordinates of its four corners.</td></tr>
<tr><td colspan="2">Temporal information</td><td>Temporal metadata records the time an image is acquired. It therefore represents a snapshot of the phenomena in real world. It is directly available from satellite on-board data in a yy/mm/dd/hh/mm/ss format. Many DBMS and GIS software support DATE or TIME data type to store temporal metadata. Temporal data can be used for queries like 'find images acquired in day time.'</td></tr>
<tr><td>Reference information
Some image native attributes may be the constraints for image selection. For Example,</td><td>Spatial resolution</td><td>This element refers to the correponding ground size of an image pixels. For example, the resolution of LANDSAT images ranges from 30m to 120m. Since this element provides description how 'detailed' the image describe the real world phenomena, it is an important judgment factor for image browsing.</td></tr>
</table>

we will adopt IR images for vegetation detection	**Spectrum resolution**	The spectrum of images has a tremendous impact on whether the image can be used in a particular application. For example, images of microwave length are often used in water detection application.
	Sensor type	Related sensor name, satellite height, orbit parameters and other sensor information.
	Cloud Coverage	Cloud coverage can be used as image selection knowledge for visual verification sequence. Images with higher cloud coverage are given a lower priority.

Table 2. Adopted image metadata items and description

For better management and query efficiency, metadata for all the images are collected and stored in an independent database. We can therefore narrow down the possible images from the huge image archive by only using metadata database, and later retrieve only the images that meet users' demands.

3.2 Spatial Browsing Mechanism

There are several important missions for spatial browsing interface, including

- Retrieving necessary data items from image metadata database and introducing a friendly and user-oriented intuitive user interface to users.
- Processing spatial queries and constraints
- Executing queries
- Retrieving qualified images and sorting them based on image selection knowledge

Users can specify query constraints and browse the retrieved images in the designed interface. We summarize some necessary characteristics for ideal spatial browsing interface:

- Auxiliary data is necessary. For example, an index map helps users to have a better understanding about the spatial relationship between images and the selected regions.
- Thematic data, such as maps with annotations, help users to define the extent of interests.
- Ideal and intuitive interface for spatial constraints creation is necessary. Users can specify spatial constraints by point, line or regions, and further integrate them with the use of spatial relations.
- System must be able to process users' spatial relation queries received and retrieve necessary metadata (e.g. image coverage extent). The display of qualifying images must be based on advanced cartographic knowledge.
- In the situation where more than one image qualifying given constraints, the result must be appropriately sorted based on developed image selection knowledge before presenting to users.

3.3 Spatial relation formalization

To provide users an intuitive interface for querying images, spatial relation formalization is necessary. No matter what type of spatial constraint is, the most important requirement regarding image queries is 'the qualified image must cover the area of interests', represented by:

QUALIFIED_IMAGE_TEST(candidate image) = TRUE

IF COVER(candidate images , region of interest) = TRUE

In other words, there are two main factors to evaluate when considering if an image is spatially qualified.

- The spatial coverage of candidate images and area of interests
- The definition for spatial relation "COVER"

Candidate image--The spatial coverage of the candidate images can be directly retrieved from metadata database. If further constraint like resolution or even application is specified, some types of images may be ignored.

Area of interests—There are two approaches to create area of interests, either by directly creating an area of interests or by using the combination of reference objects and spatial relations. The first approach is similar to the basic selection tool most GIS software provides (point, line, polygons). The second approach would require the interpretation of the geometric meaning of the selected spatial relationships (e.g. the *east part* of the United States). Reference object can be point, line or region type.

COVER relation—This relation can be defined with the geometric intersection algorithm. However, the developed algorithm must be applicable to all of the possible situation, i.e., point-polygon, line-polygon and polygon-polygon. Table 3 summarizes the types of spatial relation involved in browsing behaviours.

Spatial relation type	Example
Metric relation	Expressing the "far" or "near" concept. Metric relation is always used simultaneously with reference object in our system.
Direction relation	Direction relation is defined on the basis of reference object as well. Area of interests can be created by the combination of reference object and specific spatial relation. (e.g. East of the city of New York). The definition of direction relation changes with the dimensionality of reference objects.
Topology relation	The frequently used topology relation in browsing behavior is **COVER**

Table 3. Spatial relations for image database browsing

4. IMPLEMENTATION

This section demonstrates some of our preliminary results regarding to the image searching and browsing via Internet. Since we are in the early stage of the research, some of the developed spatial predicates are still incomplete and most of the data is only simulated. We separate the system development into two parts, the first part relates to Internet technology, while the second part deals with the development of spatial predicates. For Internet part, ESRI ArcIMS was chosen as the test platform for the image database. ArcIMS is the first server-side solution for Internet GIS development, further system development and customisation is possible with the support of JAVA language. A server-side solution, such as the 3-tier structure of ArcIMS, has the advantage of protecting data from illegal downloads. To be able to provide higher customisation ability, the Internet spatial query interface is developed with JSP and Servlet, which are client independent so that users can browse without download plug-ins. It saves Microsoft SQL Server 2000 is chosen as the database platform to process alphanumeric data and we also introduce robust JDBC interface module for Chinese characters displaying and querying. At the moment, the specific implementations in this paper include two separate parts, and these two parts will be integrated in the future research:

- Implementation of Internet GIS interfaces
- Spatial browsing mechanism kernel test

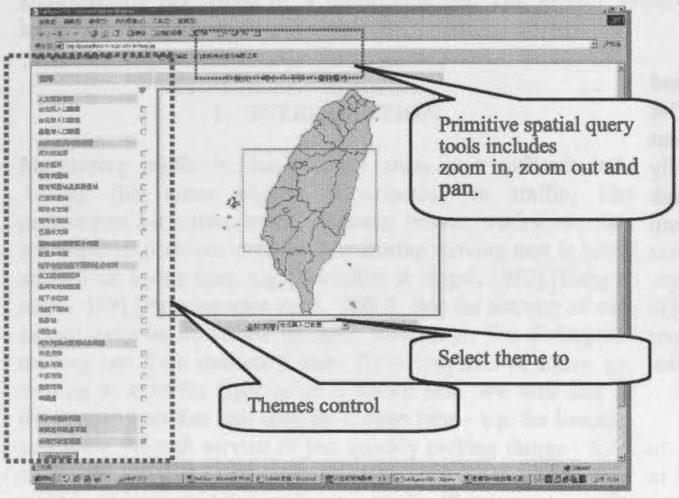

Primitive spatial query tools includes zoom in, zoom out and pan.
Select theme to
Themes control

Figure 3. Internet client interface

In order to implement proposed system structure, there should be a client connector for users to login system with common Internet www browsers. We adopt the JSP + JavaScript solution to have better query efficiency and a more completed expansion capability on user client interface. A primitive test system interface is shown in Figure 3. It supplies a set of primitive spatial browsing tools (e.g. zoom, pan) to help users to specify region of interests. We implement these primitive spatial interaction tools by JavaScript. For pure JSP framework can do little things about client user interface modification.

For faster calculations and better security on system kernel source code, we plan to implement the second part with CGI (Common Gateway Interface). This part is still under development with JAVA and shall be integrated with the first part soon. In order to test the spatial relation module, we create an image database with 5000 records of simulated image metadata. Both the time and spatial coverage of images are simulated randomly. Figure 4 illustrates the distribution of simulated images by using color brightness to represent time difference. The system is temporarily developed and tested in ArcView environment.

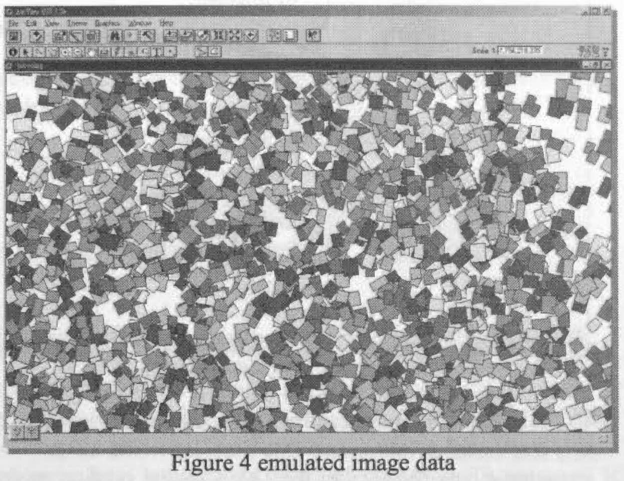


Figure 4 emulated image data

Figure 5 shows a query regarding the spatial predicate 'along', which is often used for linear objects or the boundary of polygons, e.g., images along a highway. Theoretically this can be done with a geometric intersection test, but it is certainly not equal to simply an 'intersection' test. In the browsing interface, the linear constraint will be automatically divided into two categories of line segments, one category covered by images and one category not. Users can therefore have a clear understanding about how images can be used in their applications.

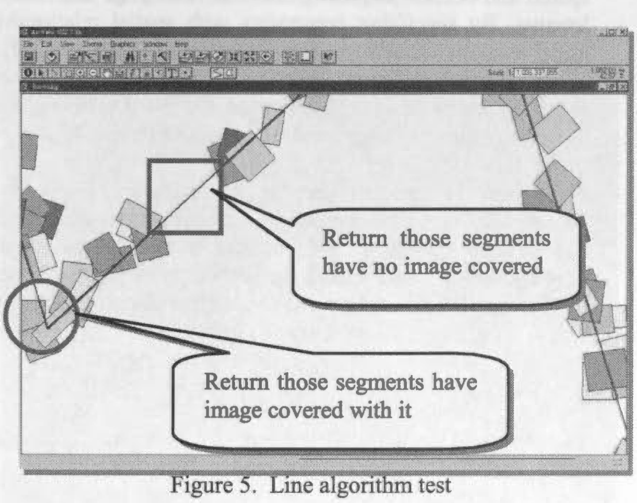

Return those segments have no image covered
Return those segments have image covered with it

Figure 5. Line algorithm test

Figure 6 shows a query regarding to the spatial predicate 'cover', which can be used for point, line, and polygons. This is again based on the geometric intersection tests. Note in this case the area of interests is much larger than the size of any image, so no image can entirely cover the selected area. Under such circumstance, some of the areas may be covered by a number of images, while some areas may not be covered at all. Like the above case, the selected area must be divided into two categories for users further browse.

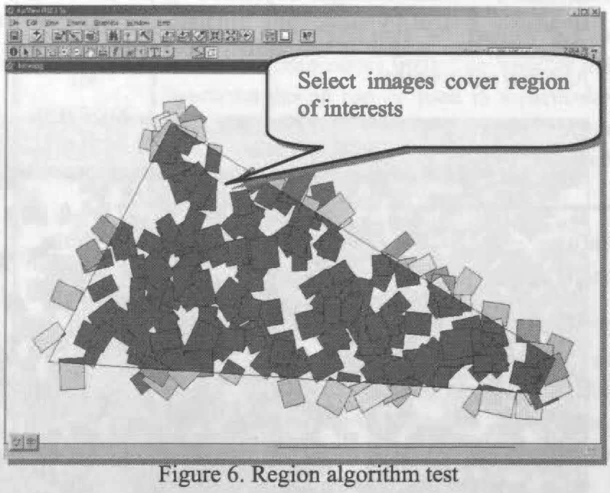


Figure 6. Region algorithm test

The next step is to integrate these two components together and it is still under development. One problem to the integration is to improve the efficiency of the JAVA spatial relation module. We will discuss the integration steps in future papers.

5. CONCLUSIONS

Remote sensing image has become the most important and efficient source for the collection and update of GIS data. With such a huge volume of images produced everyday, an efficient image searching and browsing mechanism and a friendly interface is necessary. The Internet image-browsing framework proposed in this paper is an attempt to incorporate human spatial and domain knowledge towards the image selection via Internet. By specifying constraints with spatial relationships, users can interact with image database easily and intuitively with the help from smart browsing mechanism. There are some important issues for smart browser implementation in the future, including efficiency improvement and distributed structure.

This paper is the first part of this research. We hope to successfully integrate corresponding components described in this paper together. A JAVA spatial relation kernel is under development. It may increase the portability of proposed spatial mechanism. Distributed structure is another import issue for this research for future image databases utilization.

6. REFERENCES

David M. and Frank A.,□1992□NCGIA Initiative 2□Language of Spatial Relations Closing Report

Egenhofer, Max and Herrin, J. R., □ 1993 □ Querying a Geographic Information System *Human Factors in Geographical Information System*□Eds. D. Medyckyj-Scott and H. Hearnshawn□,Belhaven, London, pp 124 – 135.

Frank A. U., □ 1993 □ The Use of Geographical Information System□*The User Interface is the System*.Human Factors in Geographical Information System□Eds. D. Medyckyj-Scott and H. Hearnshawn□,Belhaven, London, pp 3 – 14.

S. Sestito & T. S. Dillon, Automated Knowledge Acquisition , Prentice Hall, 1994

Stephenson Geoffrey A., □1994□, Knowledge Browsing Front-Ends to Statistical Database. Statistical and Scientific Database Management. pp 327 – 337.

Jung-Hong Hong□1998□□Primitive spatial relationships in GIS and their applications in human-computer interactions. Proceedings of the 17th Conference on Surveying Theory and Applications, National Cheng Kung university□pp.1100-1109

NASA photograph image space aeronautics airplane shuttle. http://nix.nasa.gov/browser.html (accessed 3 Jan. 2002)

The USC-SIPI Image Databasehttp://sipi.usc.edu/services/database/Database.html(accessed 3 Jan. 2002)

ESTIMATING VEHICLE ACTIVITY USING THERMAL IMAGE SEQUENCES AND MAPS

U. Stilla, E. Michaelsen

FGAN-FOM Research Institute for Optronics and Pattern Recognition, Gutleuthausstr. 1, 76275 Ettlingen, Germany
stilla@fom.fgan.de, mich@fom.fgan.de

KEY WORDS: Thermal images, Large scale vector maps, Spot detection, , Grouping

ABSTRACT:

Monitoring traffic in dense build-up areas is a difficult task. In this contribution we describe a method to assess the activity of vehicles based on airborne image sequences from an infrared camera. At the resolution of approximately one meter vehicles appear as elongated spots. In urban areas many additional other objects have the same property. To discriminate vehicles from these objects we fuse information from IR-images and vector-maps. Besides evidence from overlapping areas of frames an IR-sequence is accumulated in the scene. We also perform grouping of vehicles into rows along the margins of the roads. These tasks can be automatically performed by a production net. The generic productions for grouping become feasible through the use of map knowledge as context.

1. INTRODUCTION

Monitoring traffic in dense urban areas is a difficult task. Among the many objects participating in traffic, like pedestrians, bicyclists, trams, subways, busses, trucks, etc. cars are a special problem in cities. Monitoring moving cars is being studied for a long time, e.g. [Dreschler & Nagel, 1982],[Haag & nagel, 199],[Partsinevelos et al., 2000]. But the activity of cars cannot only be described by their movement. We distinguish moving cars from stationary ones. These may still be active, i.e. waiting at a traffic light or in a traffic jam. We also like to distinguish cars that halt only for a short time - e.g. for loading, in a drive through service or just quickly picking things - from those who are parked for hours or days. Thus the temperature of a car is an important feature for its activity. Temperature can be captured by IR-sensors.

There is mutual evidence between cars and certain GIS-segments. Cars will be present only on roads and parking lots. While some authors suggest to use car detectors as hints for parking lots [Quint, 1996] or roads [Hinz & Baumgartner, 2001], we use the map as support for searching cars. For oblique views and building recognition a 3-d GIS database is used [Stilla et al., 2000], but in dense urban areas the cars will be visible only in nearly vertical views, and roads and parking lots are flat. While in [Stilla, 1995],[Stilla et al., 2000] we focussed on the geometric aspects, we here emphasize the importance of the radiometric information. So conventional 2-d large-scale vector maps are sufficient.

Stationary passive cars appear as dark stationary spots, because they are colder than the surrounding. They will usually occur along the margin of roads or in parking lots, and they will be grouped into rows. Fig. 1 shows such a situation. Active vehicles will appear as moving bright spots on the roads. A stationary bright spot within a row of stationary dark spots can be interpreted as a car that is still warm (has been moved short time before) or as warm spot on the bare concrete giving the hint, that there has been a vehicle short time before, that moved away. Rows of bright spots in the roads are probably cars, that are waiting at a traffic light or in a traffic jam. Thus the percentage of bright versus dark spots gives a good estimation of the activity of cars in the scene.

Fig.1 shows a typical example of a thermal image of an urban site with rather low car activity. While some buildings and cars have good contrast to the background, roads do not necessarily have clear margins in this kind of imagery

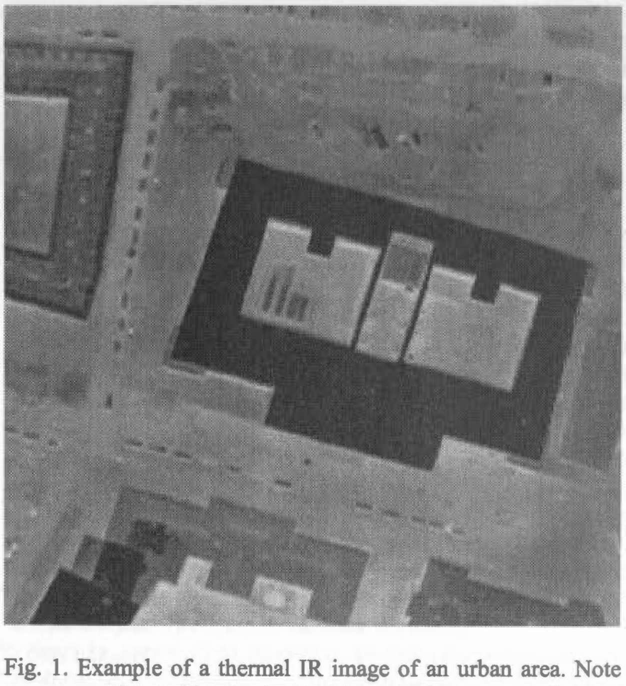


Fig. 1. Example of a thermal IR image of an urban area. Note the parking cars appearing as dark spots grouped in approximately equally spaced rows along the margin of the roads.

Several authors have proposed to use perceptual grouping in urban areas with the contextual aid of maps. E.g. in [Quint, 1996] parking lots known from the map are used to analyze monocular aerial images and to group cars into rows. The knowledge is represented in a semantic net, and the instantiation of corresponding structures in the data is done by A* search. Here we use a production net to represent the knowledge and a pruned bottom-up search.

2. IMAGE PROCESSING

In this section we present the proposed method and show example results: In Sec. A the Operator is explained, that we use for spot cue detection. Sec. B presents the cluster formation, that assembles pixels into spots with considerable mass and second moments. The fusion with map information, that filters out most of the clutter spots is described in Sec. C. We then focus on the grouping process in Sec. D. In Sec. E we recommend to also fuse the information from different images in the overlapping regions. While this section focuses on the application, the following Sec. III gives a brief methodological overview.

2.1 Iconic spot detection

For detection of cues for cars a spot detector [Kohnle et al., 1993] is used. For each pixel p=(x,y) it looks for the maximum gray value md along the margin of a square of radius d (Pixels (x+j,y+i), where Max (|i|, |j|)=d). See Fig. 2a. The radius d varies from 1 to a maximum, which is a parameter of the operator. Of all maxima md the minimal value is compared to the gray value in the center. If the central value is significantly higher, then there is evidence for a hot spot at this position, and the operator sets this difference as value at position (x,y) of its output. Otherwise it sets the output to zero. Dark spots are found in an analogue way.

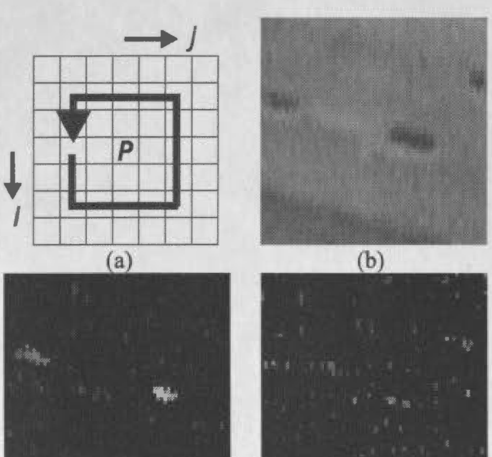


Fig. 2. Spot detector and its application to an IR-image: a) Pixel locations considered in the inequalities of the operator; b) section of an IR-Image, c) result of the cold spot detector; d) result of the hot spot detector.

Fig.2 shows the action of this operator on a section of the image displayed in Fig. 1b shows the section from the center, Fig. 2c shows the cold spot image and Fig. 2d the hot spot image. For clarity Fig. 2c and Fig. 2d have been adjusted in brightness.

2.2 Spot Formation by Clustering

Non-zero pixels are entries to the symbolic reasoning system presented in this contribution. The first step in this system is the formation of spots from sets of adjacent pixels of consistent temperature (hot or cold). Apart from this temperature a spot gains a location, a mass and second moments as attributes.

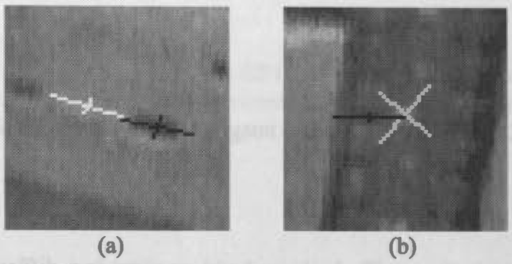


Fig. 3. Examples of spots found by the cluster formation: a) A strong cold spot (a car) on the right, and a weak warm spot on the concrete on the left; b) a cold and a warm spot on a roof.

Fig. 3 shows examples of such spots. Fig. 3a displays one example of a cold spot (black) caused by a vehicle and one example of a hot spot (white) caused by still warm road surface in a place, where a car has moved away short time ago. The crosses indicate the position, orientation and eccentricity of the spots. The mass of the cold spot is one order of magnitude higher then the mass of the hot spot. While these are examples for the kind of spots we are interested in, Fig. 3b shows examples of spots with similar or more prominent attribute values, that do not correspond to vehicles or phenomena caused by vehicles. These are due to structures on a roof in the upper left region of the example image. The cold one is very eccentric but has only little mass (comparable to the hot one in Fig. 3a, the hot one has high mass, but its eccentricity is so low, that the orientation has barely any meaning. Among all spots in the image, the ones corresponding to vehicles like in Fig. 3a are a minority, while clutter like in Fig. 3b forms the majority. This motivates the use of map knowledge to concentrate the search on the road margins and parking lots.

2.3 Additional Map Information

Large scale vector maps of roads provide a proper means to exclude most of the non-vehicle spots. Fig. 4 shows such a map overlaid as white lines on the image. The regions of interest inferred from the map are displayed in dark gray-tone, while the non-interest regions are displayed much brighter.

For the registration of map and image attention must be paid to the interior distortions of the camera (in this case a quadratic distortion with one parameter was used) and to the exterior distortions of the view (here we assumed the scene to be flat and thus used a projective homography with eight parameters).

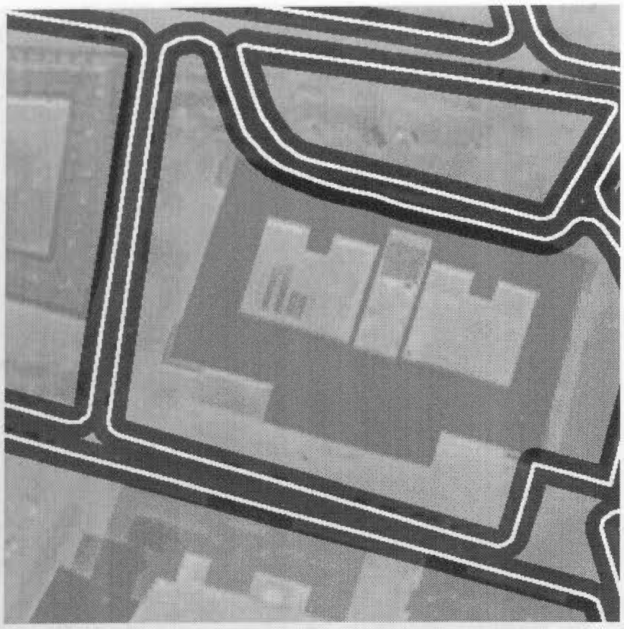
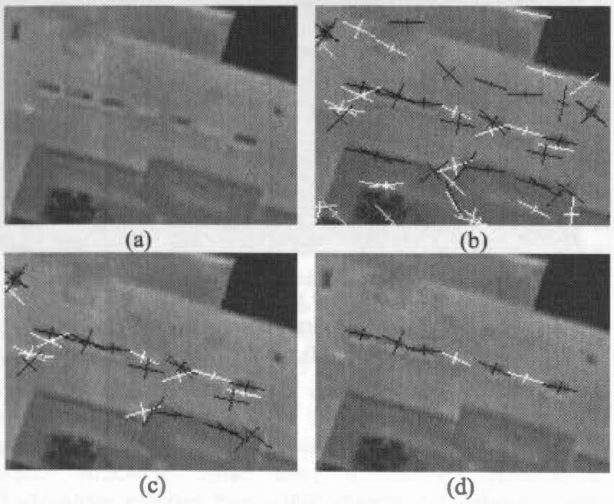


Fig. 4. Interest regions imposed by a map on the IR-image. The road margins of the map are displayed as white lines. The resulting interest regions are shaded much darker then the regions that are excluded from the search for cars.

2.4 Grouping Spot-Rows along the Road Margins

Cars tend to be placed in equidistantly spaced rows along the margin of roads. This criterion allows to discriminate them from other spot-shaped objects. Grouping of such spots into rows of arbitrary length is a generic operation. Fig. 5a shows a section of an IR image containing a row of cars.



(a) (b)

(c) (d)

Fig. 5. The benefit of grouping: a) A section of an IR-image; b) all spots constructed in that region; c) spots in the interest region given by fusion with the map; d) car-spots remaining consistent with the row model after grouping.

All detected warm and cold spots in this section are displayed in Fig. 5b. Spots caused by cars constitute only a subset of these. Fig. 5c shows those spots, that are sufficiently close to a road margin. The grouping starts only from spots, which exceed a minimal mass. The grouping direction is constrained by the road margin. All spots are taken into account, that exceed another lower threshold. Only those spots fitting into the straight and equidistantly spaced row model are grouped.

Still there may be several alternatives of grouping, e.g. if two spots are close to another in a location consistent with the model (see Fig. 5c, most right member of the row). Among the alternatives one group is selected based on an assessment, which is calculated from the number of spots, total mass, regularity in spacing and straightness and consistency in orientation of the spots. Fig. 5d shows the best group containing seven spots.

2.5 Processing image sequences

Compared to standard aerial imagery the image size of the IR images used here is small (standard video). An appropriate resolution on the ground implies that only a small section of the scene can be captured by one image (frame). To monitor a broader scene a sequence of images has to be recorded. The analysis should not be performed on the images independently.

Rather we suggest to project image spots of sufficient mass into a fixed Cartesian coordinate system in the scene and perform a clustering of the spots in the scene. Utilizing map information and grouping is then performed in the scene instead of the image. Large structures, e.g. very long rows of vehicles, that don't fit into a single image can be analyzed.

Parameters and tolerances can be defined in real scene measures (e.g. meter), which is more appropriate than in the image (e.g. pixels). For cameras with non-projective distortions straight lines in the scene are not necessarily straight in the image. This poses additional problems to non-local grouping techniques like the row formation, if the pixel coordinate system is used.

Image sequences allow a multiple measurement of the attributes of a spot. Averaging mass, orientation and eccentricity reduces the noise. Image sequences in time allow to determine velocity as additional attribute of a spot.

3. THE PRODUCTION NET

The automatic extraction of vehicle rows described in the previous section is implemented in a blackboard based production system [Stilla, 1995],[Stilla & Michaelsen, 1997]. A production captures a structural relation of the object model. The production defines how a given configuration of objects is transformed into a single more complex object. In the condition part of a production, geometrical topological and other relations of objects are examined. If the condition part of production holds, an object specific generation function is executed to generate a new object.

The hierarchical organization of object concepts and productions can be depicted by a production net which displays the part-of hierarchies of object concepts. Concepts represent object types and define a frame for concrete objects (instances) which are described by their attribute values .

The production net for this contribution is shown in Fig. 6. Production p1 generates objects i_SPOT from sets of objects i_PIXEL which are close together and consistent in temperature. Production p2 generates objects s_SPOT in an analogue way from sets of objects i_SPOT. While p1 captures a pure part-of relation staying in one domain (the image), production p2 has concrete-of semantic and changes domains from image to scene.

Production p3 initiates the grouping process. Here the link from the object m_LINE of the map domain has a context semantic, while the link to the object i_SPOT is a part-of link. Production p5 performs the generic perceptual grouping of objects i_SPOT into objects i_ROW and uses objects m_LINE as context. Productions p4 and p6 are analogue to productions p3 and p5 working in the scene domain.

The recursive productions p5 and p6 need to use the object concept m_LINE as context. This prevents them from grouping all possible rows of objects i_SPOT or s_SPOT. Otherwise there would be very many possibilities, making the process infeasible.

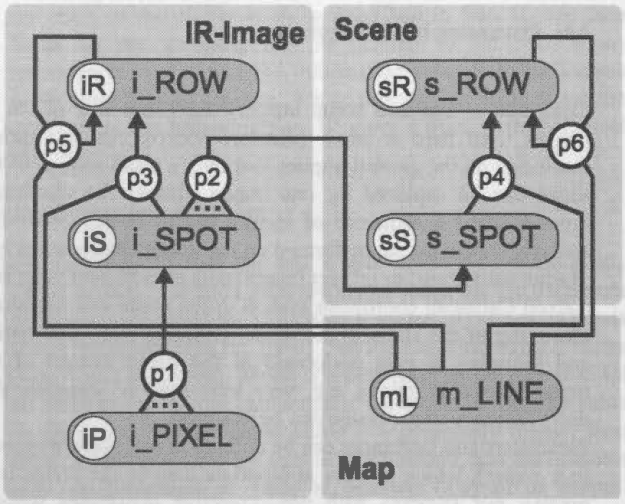


Fig. 6. Production net for automatic vehicle row extraction for activity estimation. Such production nets give a good overview over the workflow. Cycles in the net give a hint, where there is danger of potentially high search effort.

4. RESULTS

The image in Fig 7 shows a scene section obtained further to the left side of the section displayed in the image in Fig. 1. The same groups of cars are seen in the right margin of Fig. 7 and the left margin of Fig.1 In Fig. 8 this group appears in the upper central part. This figure displays all spots which are accumulated in the scene from approx. 40 images and the objects m_LINE of the map. Cold spots are drawn in black and warm spots are drawn in white. Many of the hot spots result from structures on the roofs. The percentage of hot spots in relation to all spots is 34%.

Using the constrains described above only the spots displayed in Fig. 9 remain as cars. All Spots from roofs and some spots from roads disappear. The percentage of active cars in relation to all cars is 3%.

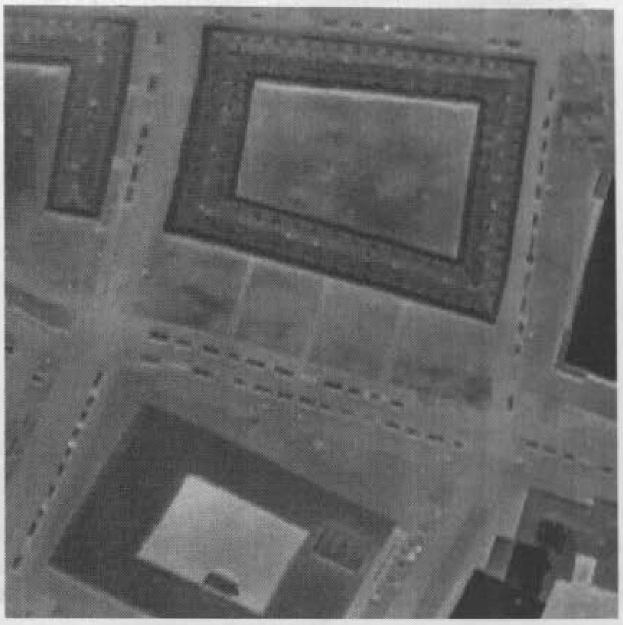


Fig. 7. Other IR image of the same sequence. The images shows a scene section left of scene section displayed in Fig. 1.

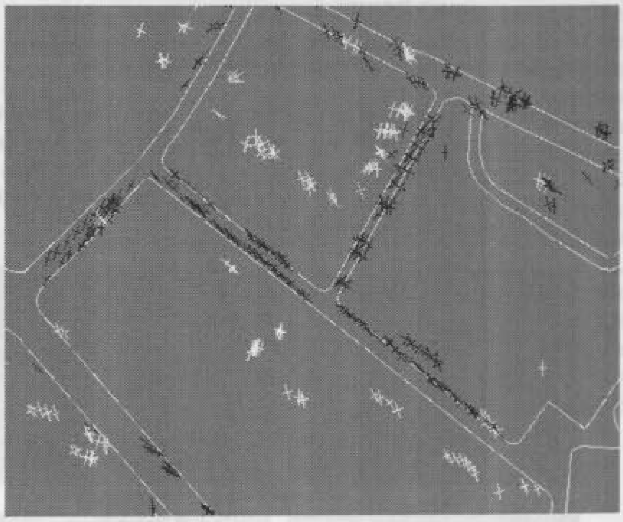


Fig. 8. Section of the scene with cold and hot spots (comp. Fig. 5b). The cars are contained, but also many warm spots on the roofs.

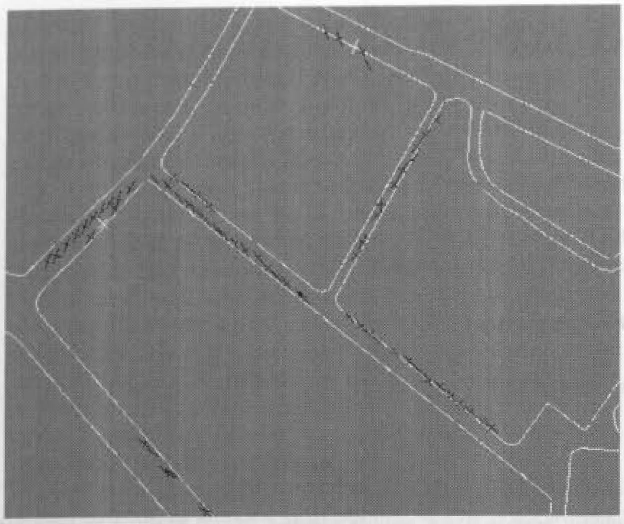


Fig. 9. Result of grouping along the margin of the roads in the scene (comp. Fig 5d). All spots correspond to cars. Only two spots are warm. The activity is estimated low.

5. DISCUSSION

Crucial for the success of such a system is an adequate choice of the parameters [Michaelsen & Stilla, 2000]. One example of such a parameter is the width of the interest regions around the road margins. This should be chosen with respect to the maximal errors of the registration and additionally with respect to the distance from the road margins in which cars typically appear. If it is chosen too small with respect to these criteria, a large portion of cars will not be detected. If this parameter is chosen too liberal, spots on roofs will be mistaken for cars.

First results show that not all cars are found. The parameter setting was tuned to avoid false alarms. E.g., we only accept groups with at least four cars. Estimating the number of cars would require a different parameter setting. The estimation of the ratio of active versus all cars only needs a sample of the cars present in the scene.

The group of cars on the right margin of the road in the lower left region of Fig. 7 results in a distributed cluster of spots in the scene (Fig. 8, left). It leads to two groups of cars on either side of the road (Fig. 9, left). We blame this fault to inexactness in the determination of the external camera orientation. Because GPS-INS data were not provided with this IR image sequence, we estimated the orientation data on every tenth image manually and interpolated in-between. Provided that orientation data are recorded with sufficient accuracy such faults can be avoided.

Perceptual grouping and the utilization of additional data sources like maps can help in the automatic interpretation of airborne high-resolution thermal image sequences. Even in highly structured urban areas it becomes possible to discriminate vehicles from other objects and determine their activity. Working in the scene instead of the image has advantages, e.g. evidence from multiple images can be accumulated.

The test data set shows a scene with rather low activity. To make assessments on how the method works on scenes with high activity further data have to be processed, e.g. scenes containing major roads with considerable traffic and scenes containing parking lots of shopping centers in the opening hours.

In scenes with moving cars tracking the cars and the determination of their velocity will be another important discriminative attribute. Active vehicles will move preferably on roads at speeds, for which a reasonable bound is given. So hot spots of considerable mass and with their velocity fulfilling these criteria can be counted as active vehicles, even if they are not members of a row.

6. REFERENCES

Dreschler L, Nagel HH (1982) Volumetric model and trajectory of a moving car derived from monocular TV frame sequence of a street scene. CGIP, 20, 199-228.

Haag M, Nagel HH (1999) Combination of edge element and optical flow estimates for 3D-model based vehicle tracking in traffic image sequences. IJCV, 35:3, 295-319.

Hinz S, Baumgartner A (2001) Vehicle detection in aerial images using generic features, grouping and context. In: Radig, B., Florczy S., Pattern recognition, DAGM 2001, LNCS 2191, Springer, Berlin, 45-52.

Kohnle A, Neuwirth W, Schuberth W, Stein K, Höhn DH, Gabler R, Hofmann L, Euing W (1993) Evaluation of essential design criteria for IRST systems. 76/SPIE Vol.2020 Infrared technoloy XIX.

Michaelsen E, Stilla U (2000) problems in geometric modeling and perceptual grouping of man-made objects in aerial images. International archives of photogrammetry and remote sensing, vol. 33, part B3, 577-583.

Partsinevelos P, Agouris P, Stefanidis A (2000) Modelling movement relations in dynamic urban scenes. International archives of photogrammetry and remote sensing, vol. 33, part B4, 818-825.

Quint F (1996) Recognition of structured objects in monocular aerial images using context information. In: Leberl F, Kalliany R, Gruber M (eds) Mapping buildings, roads and other man-made structures from images. IAPR TC-7 workshop in Graz, Oldenburg, Wien, 213-228.

Stilla U (1995) Map-aided structural analysis of aerial images. ISPRS Journal of photogrammetry and remote sensing, 50(4):3-10

Stilla U, Michaelsen E (1997) Semantic modeling of man-made objects by production nets. In: Gruen A, Baltsavias EP, Henricsson O (eds) Automatic extraction of man-made objects from aerial and space images (II). Basel: Birkhäuser, 43-52

Stilla U, Soergel U, Jaeger K (2000) Generation of 3D-city models and their utilization in image sequences. International archives of photogrammetry and remote sensing, vol. 33, part B2, 518-524.

AUTOMATED TECHNIQUES FOR SATELLITE IMAGE SEGMENTATION

A. Guarnieri*, A. Vettore*

*CIRGEO (Interdepartment Research Center for Geomatics)
via Romea 16 – 35020 Legnaro (Padua) – Italy
Phone: +39-049-8272688; Fax: +39-049-8272686
e-mail: antonio.vettore@unipd.it; cirgeo@unipd.it

Commission IV, Working Group IV/5

KEY WORDS: satellite image classification, image segmentation, clustering

ABSTRACT

In this paper a combined method between "classical" and automatic approach for remote sensing image analysis is presented. Typically, satellite images are used in order to detect the distribution of vegetation, soil classes, built-up areas, roads, and water body as rivers, brooks, lakes, ecc. Referring for example to Landsat-TM images, the identification of such aspects is performed through the "classical" approach of image classification. Basically, it deals with the use of pseudocolors and/or combinations of various spectral bands to acquire different thematic layers from the images. In this work a further processing step is introduced, namely a segmentation algorithm is applied to color images in order to improve the image analysis both from qualitative and quantitative point of view. This algorithm belongs to the class of operations performed in the automatized unsupervised analysis of color images. Last recent advances in the field of computer science and CPU performance have lead to a great reduction of data processing times, allowing therefore to apply also in the field of Remote Sensing more complex algorithms. The proposed segmentation algorithm is based on a feature-space approach and implements two processing techniques: the "histogram thresholding" [1] and the "clustering" [2]. Some interesting results applied to Remote Sensing images will be provided.

1. INTRODUCTION

So far, tipical approach to remote sensed data analysis has been based on the use of different linear combinations between available spectral bands or creating new ones. In this phase satellite data are processed through the activity of a human operator, which using RGB filters tries to identify several classes of elements appearing on them. These classes are therefore a way to group homogeneus land features, such as urban areas (roads, buildings, etc.), vegetation (woods, cultivated and uncultivated soils) and water areas (rivers, lakes, etc.). These layers are often represented in terms of pseudocolor images, in order to better highlight a specific feature distributed along the land.

The basic concept underlying that procedure is similar to the image segmentation, the first step for the digital image processing normally adopted in the field of Computer Vision. This operation is performed through the partitioning of an image just in homogeneous and separated regions. So far, segmentation techniques were applied only to gray scale images, though the color information would allow a more complete image representation. In fact, the application of this method was limited mainly by the computational time spent for color data processing, larger than the one needed for gray color data. Today, recent advances in the field of computer science and CPU performance, have lead to a great reduction of data processing times, allowing therefore to apply these segmentation algorithms to the field of Remote Sensing, as well. Several color image segmentation algorithms are nothing but the development of previous gray color procedures, others are instead new *ad-hoc* techniques for color data, which take into account the physic relationship between light and coloured materials. Such algorithms work in well defined color spaces, such as RGB , HSI or HSV. Anyway these reference frames are not uniform, i.e. color differences of same entity, as perceived or "measured" by human eye, are not converted in similar distances among the points representing such colors in the above mentioned spaces. These problem has been overcome by introduction of uniform color spaces, such as C.I.E.L*u*v* and C.I.E.L*a*b*.

Adopted color image segmentation algorithms can be classified as follows:

- Feature-space based, working on the space of coloured figures in the image
- Image-domain based, i.e. they analyse the image geometry and color
- Physics based, involving the physic relationships between light and materials

In this work, a segmentation algorithm belonging to the first class was applied to remote sensed images in order to improve qualitatively and quantitatively the result of the "classical" approach. This means, assuming that different classes of land features are already identified through classical methodology, a further analysis step is introduced by application of proposed algorithm to refine the results of previous phases. The main advantage of this approach rely on the fact that the algorithm is unsupervised, therefore it doesn't require any *a-priori* information and can be fully automatized.

The paper is structured as follows: in section 2 a brief overview of feature-space based techniques is reported, then in section 3 the proposed algorithm is explained. Some results of its application to remote sensed images are presented in section 4, while section 5 deals with the conclusions.

2. THE FEATURE-SPACE BASED TECHNIQUES

Assuming that color is a constant property of the surface of each object appearing on the image, the image segmentation can be addressed through two following different strategies: *clustering* and *histogram thresholding*.

In the first technique, image pixels are firstly mapped on a certain color space, in order to convert pixels to points. Then these points are grouped in different sets (clusters) on the ground of color information of each corresponding pixel. In this way, given the above mentioned assumption, the different objects of the image can be discriminated in terms of these clusters or cloud of points. The distribution of the points inside each cluster depends mainly upon the color change, due to shading effects and noise of the acquisition device. It should be noted that the clustering technique belongs to the *unsupervised* classification algorithms, since no *a-priori* knowledge about the image is required. An example of clustering implementation is provided by the k-means algorithm: it is widely used not only for color image segmentation but also for applications involving vectorization and data compression.

The histogram thresholding algorithm belongs to another class of segmentation techniques, early applied to gray scale images. In this method image pixels are not mapped on a color space, but rather some *ad-hoc* histogram of color figures, such as the Hue, are generated. Through that model, objects on the image will be identified as peaks of the histogram, while the background will correspond to its depressions. In the field of color images a thresholding algorithm involves a bit more complex implementation, since it has to work in a 3D color space, meaning that a 3D histogram has to be taken into account. Furthermore, in this case histogram profiles become quite jagged with spurious peaks, which make the segmentation more ambiguous.

The proposed method is based on a combination of the clustering and histogram thresholding techniques. In summary, given a remote sensed image, the representative color are firstly identified by looking for the major color groups, through the histogram thresholding of the Hue information. Then, the larger clusters in the planes of constant Hue are determined, through the k-means clustering algorithm.

3. THE SEGMENTATION ALGORITHM

The segmentation process works in the C.I.E.L*u*v* uniform color space, provided with euclidean norm $\|L^*u^*v^*\| = |(L^*)^2+(u^*)^2+(v^*)^2]^{1/2}$. In this space a cilindric coordinate reference system was introduced (H_{uv}^*, C_{uv}^*, L^*), whose Hue angle is defined as $H_{uv}^*=\arctan(v^*/u^*)$, the crominance as $C_{uv}^*=[(u^*)^2+(v^*)^2]^{1/2}=L^*\bullet S$ and the saturation as $S=([(u^*)^2+(v^*)^2]^{1/2})/L^*$. The clustering method, based on anisotropic diffusion, is a non-linear filtering technique, which performs a more high selective smoothing in omogeneus regions and almost null on the egdes, while it retains all the edge-related information. In this algorithm Hue H(x,y,t) and Saturation S(x,y,t) are represented as one complex quantity, the "crominance " function K(x,y,t)=S(x,y,t)exp(j H(x,y,t)), which is in turn diffused, clustered and segmented. The modeling of Hue and Saturation in the same function takes into account the physic relationship existing between them. It is well known that Hue changes are negligible for low values of Saturation, but noticeable for high values. The same operations are applied to the lightness function L*, which is however processed separately.

The combination of this two parallel segmentation tasks leads to a partitioned color image. The overall scheme of developed algorithm is showed in Figure 1.

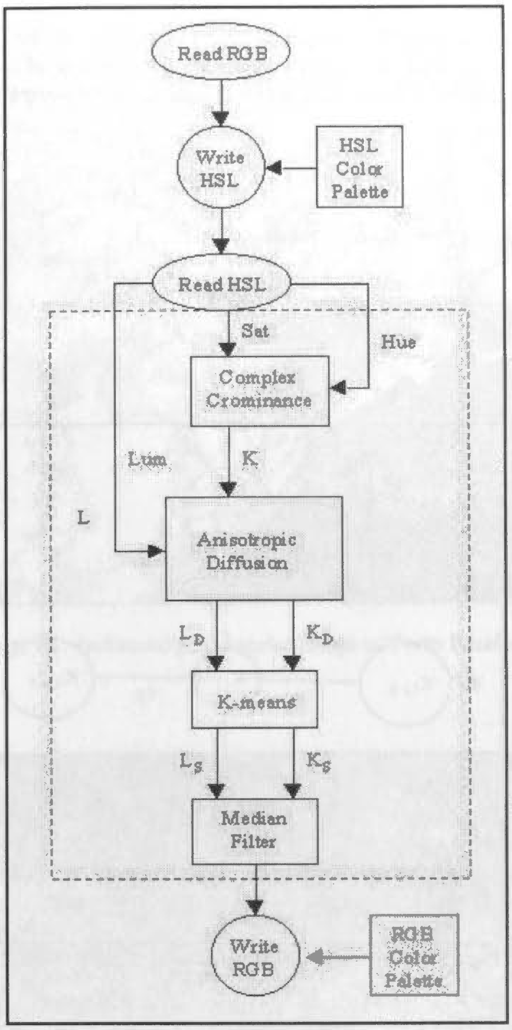


Figure 1: Scheme of the segmentation algorithm in the dashed box

The anisotropic diffusion has been numerically implemented through the partial derivative equation of heat diffusion, as stated below:

$$\partial K(x,y,t)/\partial t = div[c(x,y,t)\cdot\nabla K(x,y,t)] \qquad (1)$$

where *div* is the divergence operator, while ∇ is the gradient computed respect with the spatial variables.

Such equation can be discretized through a square lattice [3], with the the complex Crominance value K(x,y,t) associated to the vertices and the conductance coefficient c(x,y,t) associated to the arcs, (see Fig. 2), as follows:

$$\begin{aligned}[K_{i,j}]^{t+1} &= [K_{i,j}]^t + \lambda[c_N\cdot\partial_N K_{i,j} + c_S\partial_S K_{i,j} + \\ &+ c_E\partial_E K_{i,j} + c_W\partial_W K_{i,j}]^t\end{aligned} \qquad (2)$$

where $0 \leq \lambda \leq 0.25$ is required for the stability of the numeric scheme, N,S,E,W are symbols of the four vertices of the lattice and symbol δ defines the four differences *nearest-neighbour*:

$$\begin{aligned} \partial_N K_{i,j} &= (K_{i-1,j} - K_{i,j}); \\ \partial_S K_{i,j} &= (K_{i+1,j} - K_{i,j}); \\ \partial_E K_{i,j} &= (K_{i,j+1} - K_{i,j}); \\ \partial_W K_{i,j} &= (K_{i,j-1} - K_{i,j}); \end{aligned} \tag{3}$$

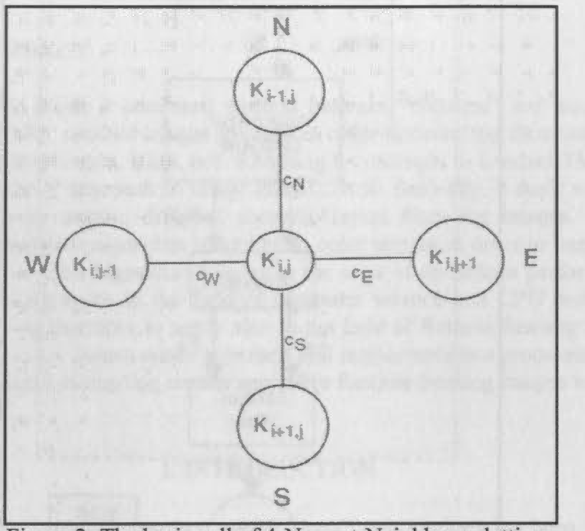


Figure 2: The basic cell of 4-Nearest-Neighbours lattice

The conductance coefficient is upgraded at each iteration as follows:

$$\begin{aligned} [c_N]^t_{i,j} &= g(|(\partial_N K_{i,j})^t|); \\ [c_S]^t_{i,j} &= g(|(\partial_S K_{i,j})^t|); \\ [c_E]^t_{i,j} &= g(|(\partial_E K_{i,j})^t|); \\ [c_W]^t_{i,j} &= g(|(\partial_W K_{i,j})^t|); \end{aligned} \tag{4}$$

The g(•) function can be modeled according to one of the following forms:

$$\begin{aligned} g(\nabla k) &= \exp(-(\|(\nabla k)\|)/A^2) \\ g(\nabla k) &= 1/(1+(\|(\nabla k)\|)/A^2) \end{aligned} \tag{5}$$

The first function is best suited to highlight edges provided with high contrast respect with the ones to low contrast, while the second form discriminates better between large and small regions. In this algorithm the second function has been chosen, in order to privilege the generation of large regions on the image, while the A costant value is dinamically computed for each iteration and set as 5% of maximum value of $|(\delta K)|$.

Finally, the image segmentation is obtained by separately partitioning of Crominance $K_d(x,y,t)$ and Lightness $L_d^*(x,y,t)$ and then combining the results through the *k-means* algorithm.
Given these settings, the proposed algorithm, based on anisotropic diffusion, becomes easy to be implemented, featuring a local behavior, i.e. the amount and the kind of smoothing are locally determined by the values of complex crominance and are adopted for each image region.
The choice for values of parameters T_1 and T_2 plays a major role in the setup of clustering process. The first, T_1, determines the average distance among the clusters, while the second, T_2, refers to to the average radius of a single cluster, if considered as a circle, and it is not possible to fix an *a-priori* value suitable for all images. Anyway an estimate of T_1 can be obtained from data distribution, estimating the radius of circle ρ_ω in which a certain percentage of the data ($\omega = 95\%$) are contained.
Tests peformed on remote sensed images showed that setting $T_1=\rho_\omega/2$ and $T_2=\rho_\omega/4$ leads to excellent results on a large set of images and, in the same time, provides a limited number of segments (often less than 7), which well represent the color information of the image. Parameter T_2 can be freely set, provided that $T_2 < T_1$ in order to avoid the overlapping of clusters.

4. TEST AND RESULTS

In this section some examples of the application of the segmentation algorithm are reported. The remote sensed data were acquired by the Landsat-TM satellite on 12/26/1996 and on 05/03/1997. In the first phase, satellite images were processed using the classical approach, working on three data typologies: "natural color" images, infrared pseudocolor images and enhanced pseudocolor images.
The first class provides the user with photographic likewise viewing of the terrain, similar to color airphotos, while the second allows to get a more in depth analysis, discriminating between different categories of the same object class. For instance it is possible to detect different forms of vegetation, since the reflectance changes are strongly related with the morphologic structure of the leaf. Enhanced pseudocolor images are obtained by combination of different bands and can be useful employed to detect the soil humidity, urban and cultivated areas, and so on.
Figure 3a shows a Landsat-TM "natural" image taken during spring 1997, as result of application of RGB flter to TM1, TM2 and TM3 bands. Figure 4a shows a view of Porto Baseleghe lagoon (Venice, Northern Italy) by the mouth of Lemene river, which was taken during winter 1996. In this case the spectral components of the image were enhanced through following bands combinations: band TM7/5, where TM7/5=(TM7-TM5)/(TM7+TM5), band TM/6 as (TM7-TM6)/(TM7+TM6) and finally band TM3/1 as (TM3-TM1)/(TM3+TM1). The first band enhances the reflectance changes due to humidity and soil composition, the second discriminates between vegetation and urban areas, while the last band allows to distinguish between damp or dry soils, on the ground of their composition (organic material, kind of rocks).
Results of application of segmentation algorithm to these images are showed in figures 3b and 4b, in which clusters were early depicted in pseudocolors but in this paper have to be converted them in gray scale. Though it is not easy to appreciate by eye the performance of the algorithm, the results can be summarized as follows. Referring to Figure 3, the segmentation of this image allowed to discriminate perfectly between the wood area (in red in the resulting segmented color image) and cultivated and dry soils (respectively in green and blue), as

reported in Figure 3b. As regards the Figure 4b, it shows how well urban areas, the sea and vegetation areas could be segmented in comparison with corresponding image of Figure 4a.

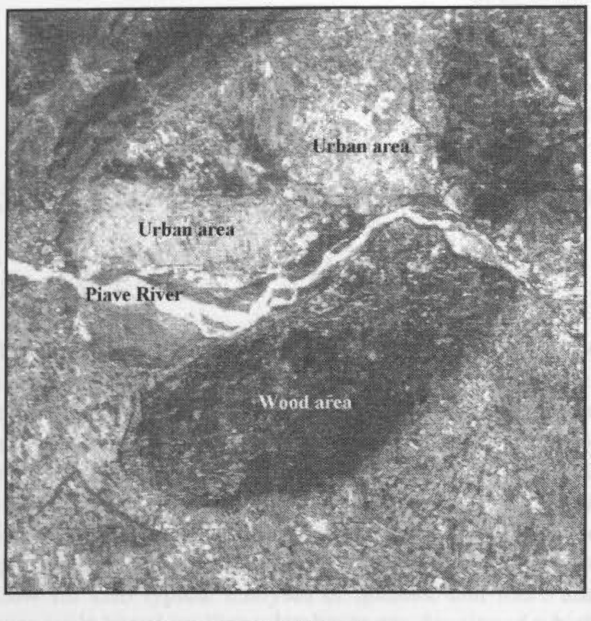

Urban area
Urban area
Piave River
Wood area

Figure 3a : Natural color image

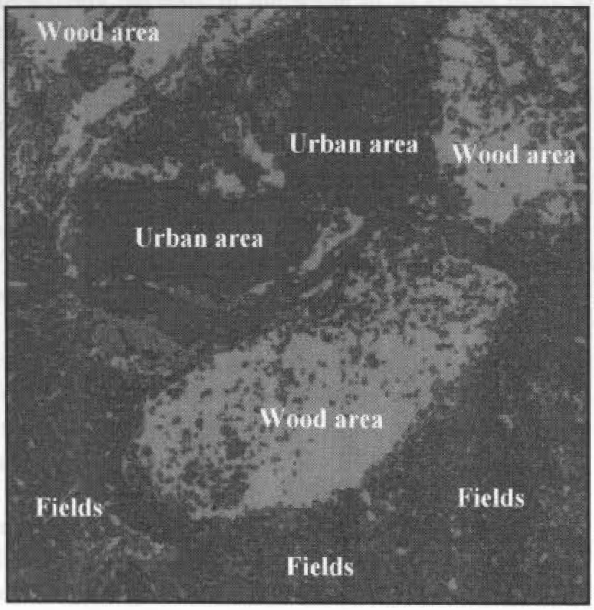

Wood area
Urban area
Wood area
Urban area
Wood area
Fields
Fields
Fields

Figure 3b : Segmented natural color image

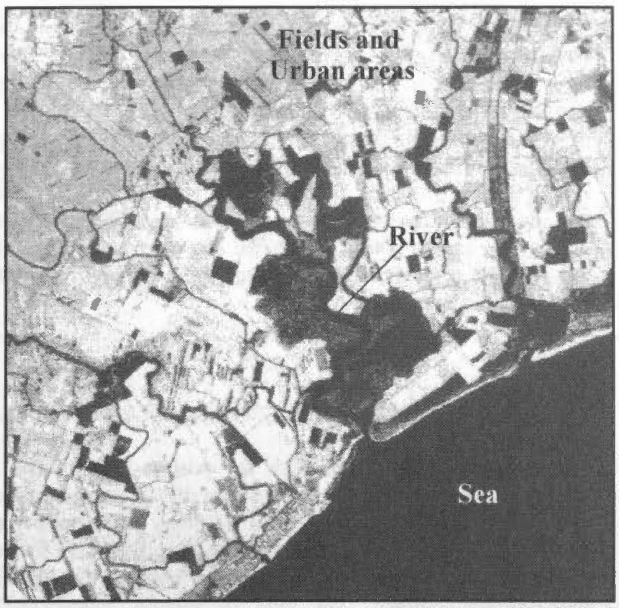

Fields and
Urban areas
River
Sea

Figure 4a : Enhanced pseudocolor image of Porto Baseleghe

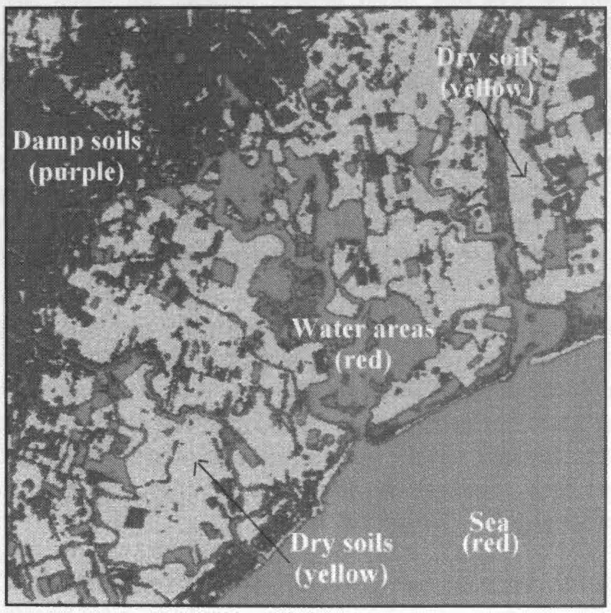

Dry soils
(yellow)
Damp soils
(purple)
Water areas
(red)
Dry soils
(yellow)
Sea
(red)

Figure 4b : Segmented image of Porto Baseleghe

5. CONCLUSIONS

In this work a combined method for remote sensed color images segmentation was presented. The method is based on the clustering and histogram thresholding technique, early developed in the field of Computer Vision for image processing. In this case the proposed algorithm was introduced as a second step in the tipical workflow of satellite image analysis in order to improve both the classification results of "classical" approach and the georeferencing of detected thematic areas. The advantage of proposed method relies in the fact that it is "unsupervised", i.e. it doesn't require any external human control or a-priori information. Therefore the procedure can be fully automated and can be easily implemented in any GIS application.

Performance of the algorithm were assessed by comparison between the results of "classical" Remote Sensing image analysis with the ones of the presented method, which were applied to the same set of satellite images. As showed in previous section, the segmentation algorithm could be successfully applied also to pseudocolor images, providing excellent results if the image is composed by wide and well defined regions. Unfortunately, if the satellite image is fragmented, and shows very similar colors distributed in a large amount of small regions, the segmentation algorithm provides often an ambiguous and unsatisfactory result.

ACKNOWLEDGMENTS

This work was developed with the project "Techniques for automatic processing of data acquired by integrated system" partly financed by MURST (Italian Ministry of University and Research) in year 2001 as project of relevant National interest. National coordinator: Giorgio Manzoni, head of the Research unit Antonio Vettore.

REFERENCES

[1] Lucchese L., Mitra S.K., . 1998. An algorithm for unsupervised color images segmentation. *Proc. of 1998 IEEE 2nd Workshop on multimedia signal processing*, Redondo Beach, CA, USA, pp. 33-38.

[2] Lucchese L., Mitra s.K., 1999. Unsupervised segmentation of color images based on k-means clustering in the chromaticity plane. *Proc. of IEEE Workshop on content-based access of images and video libraries (CBAIVL '99)*, Fort Collins, CO, USA.

[3] Perona P., Malik J., 1990. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. on PAMI*, Vol 12, No 7, pp 629-639, July 1990.

QUERY FUNCTIONALITY FOR 3D VISUAL DATABASES

Thomas Hilker, Michael Hahn, Dietrich Schroder

University of Applied Sciences, Stuttgart
Faculty: Surveying and Geo-informatics, Schellingstr. 24, 70174 Stuttgart
m.hahn.fbv@fht-stuttgart.de
www.fht-stuttgart.de

Commission IV, WG IV/5

KEY WORDS: Integration, 3D visual database, GIS, Visualisation, Query functionality Abstract

ABSTRACT:

Three-dimensional visual databases, created for landscape or 3D map visualisations, consist of integrated data sets including DTMs, orthorectified aerial or satellite images and optionally topographic or thematic maps. Objects models of man-made objects like buildings might be linked to those databases as well as models of natural objects like tree representations. If those databases are optimised for visualisation purposes they generally do not allow for GIS typical analyses or queries. But more and more tools based on those databases are developed which allow for 3D photorealistic visualisation of corresponding virtual landscape or city models. Viewers allow to interactively navigate in the scene or to enable users to virtually participate in simulated events.

The main purpose of this research is to work on the link between highly efficient 3D visualisation systems and GIS. Two different approaches are pursued:

1) Integration of query functionality to a 3 D visualisation tool
2) Integration of a 3D visualisation functionality into a conventional GIS.

The software development for both approaches integrates two different visualisation software packages, G-Vista (offered by G-Graphix) and Terra SDK (offered by Skylinesoft) with Intergraph's GeoMedia.

Query functionality and 3D visualisation tools:
Both visualisation tools are based on Microsoft COM technology. For this reason an Active X-container called GeoViewer is developed in Visual C++, embedding Active X controls of the visualisation software. GeoViewer allows to link and query GIS databases comparable to the functionality of conventional GIS. The result of those queries may have attribute related information, which will be presented to a user in a dialog box, while the location related result will activate the visualisation tool, eg. by generating a flight path and showing a virtual 3D flight to this place. GeoViewer can handle both, conventional GIS databases containing BLOBs as well as simple databases which may have not more than Points of Interest (POI) described by their spatial xyz co-ordinates.

3D visualisation functionality and conventional GIS:
The second approach basically consists of a GIS customization integrating the 3D visualisation tools mentioned above. For this purpose, a COM server is developed in C++ which can directly be executed within the GeoMedia environment. Analysis carried out with the GIS may lead to information which through the integrated geo-data visualisation functionality can be presented with a capability of the spatial - visualisation tool.

Query functionality for 3D visualisation tools is of interest for a lot of applications like real-estate marketing, virtual tourism, location based services, urban planning and others. All this applications will benefit from the fact that both, GIS and 3D visualisation tools, offer Internet services by streaming the data through the net.

ON THE USE OF HIERARCHIES AND FEEDBACK FOR INTELLIGENT VIDEO QUERY SYSTEMS

Kristin Eickhorst, Peggy Agouris

Dept. of Spatial Information Engineering and
National Center for Geographic Information and Analysis
University of Maine
348 Boardman Hall
Orono, ME 04469-5711
{snoox, peggy}@spatial.maine.edu

Working Group IV/5

KEY WORDS: Querying, Video, Metadata, Hierarchy, Intelligent Feedback, Data Cubes

ABSTRACT:

Current query systems for video databases rely heavily on structured schemas and often require user annotation of data. These systems need to be made more flexible and accessible to the diverse organizations that utilize them. By building on current work with lifelines, and incorporating new structures for organizing metadata, we can make great strides towards accomplishing this goal. This paper examines current methods for working with video queries, and explores how the use of hierarchical organization can improve on these methods. We look at how data cubes can be combined with image pyramids and scale space theory to facilitate this hierarchy. Finally, metadata structures and intelligent feedback systems are discussed and their uses both before and after the query are explained in the context of our system. This system represents a new way of looking at the information contained within a video sequence and the resultant ways in which the information should be organized.

1. INTRODUCTION

As the medium of video is used with increasing frequency for both business and personal purposes, there has been a search for new and better ways to structure video databases for querying. The system that we are constructing builds on currently existing technologies, and incorporates user feedback as well. In this paper, we present our system in five parts: Section 2 provides an overview of the current video database methods in use today, and also gives a flow chart of our system for comparison. Section 3 focuses on the use of lifelines to gather basic information about objects in the video. Section 4 examines data cubes as a means of organizing the data within a video sequence. Section 5 offers a new hierarchical way of looking at metadata based on existing FGDC standards. Finally, Section 6 discusses intelligent feedback systems, and the ways in which they may be implemented to increase accuracy of video retrieval.

2. CURRENT SYSTEMS

2.1 Schema-Dependent Query Systems

The methods currently being used to perform video queries are often schema-dependent. A schema is a highly structured description of all predicates or relations in a database. Using schemas locks the user into a set pattern of looking at the data. For example, several systems classify a body of video clips as a collection. This collection is further subdivided into clips of varying length and subject matter based on the specific needs of the user and the preferences of the database administrator. One of the most common schema-dependent methods is indexing by stratification, in which each element of interest is associated with a specific time interval, and these elements are then annotated (Hacid et al, 2000). Groupings can be made of elements that exist at discrete temporal intervals, in a manner known as temporal cohesion.

The problem with traditional systems like this is that they all involve a great deal of annotation. This is not yet an automated system, so a great deal of work is required on the part of the database administrator. Another problem that comes into play with annotated systems is that the annotator's interpretation of a scene may differ greatly from that of the user. For example, when dealing with video of a parade, the annotator may be interested in the types of floats passing down the street, while the user may be more interested in a particular tree in the background. Annotated systems cannot possibly extract all the information present in a given scene, and thus are inherently limited in their resultant querying capabilities.

2.2 Object-Oriented Modelling

There is also a movement toward object-oriented modelling, which does not require the use of a schema. In this case, a sequence of video frames can be modelled as an object, with associated attributes and attribute values to describe their contents (Oomoto & Tanaka, 1993). This kind of model is closer to our proposed system, as it includes the possibility of inheritance and thus a hierarchical type of organization. Any objects located within a specific time interval are able to inherit properties unique to that interval.

While it approaches what we would like to accomplish with our system, even this type of organization has some limitations.

There is no way to specify relationships between objects within a time interval, though all may inherit properties from a parent. Many users may be interested in either the connections between two particular objects or in an overall description of the behaviors of all objects in a scene. Our system approaches this problem by examining the lifelines of specific objects within the video and extracting attributes that help with these comparisons.

2.3 Current System

Our system consists of several interrelated parts, which can best be explained with the aid of a flow chart:

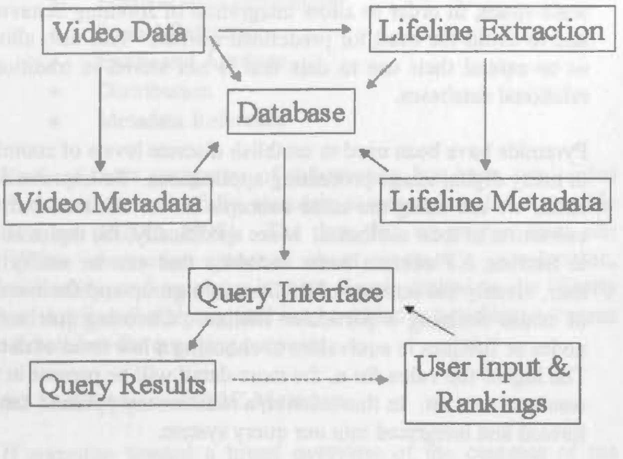


Figure 1. Components of our system

The system's input is video data. Lifelines of objects within the video are extracted in order to obtain additional information about the video contents. These two data sources (video and lifelines) have metadata associated with them, either through annotation by collectors of the video data, or through extraction and computational methods. The video and lifelines along with their respective metadata are all stored in a master database.

This database has a query interface, by which users can express their preferences for metadata values and can also give each category of metadata a subjective ranking. These inputs are used to determine the most likely video clips for a specific user's needs. Once the results have been viewed, the user has the option of updating either metadata input or rankings. This could also be done in an automated manner by the computer.

The remainder of this paper discusses each of the components of our flow chart as follows: Section 3 is concerned with lifeline extraction. Section 4 discusses the ways in which data about these lifelines can best be stored in the database, utilizing data cubes and other hierarchical methods. Section 5 is concerned with metadata and its storage within the database. Section 6 concludes with a look at the query interface and potential feedback mechanisms that can be utilized as well.

3. LIFELINES

3.1 Lifelines

A lifeline can be broadly defined as a sequence of the spatial locations (x,y,z), of an object over a time interval (t_1,t_2), during which an object has moved from one location to another. Lifelines have properties that lend themselves well to defining relationships between objects. During video analysis, extracted lifelines are defined by their nodes, which denote changes in the attributes of cardinality and acceleration of any given object. Groups of lifelines can be compared in terms of either their geometries or their attributes (Stefanidis et al, 2001). When these groups are compared, additional attributes such as topology or average separation can be computed and used as a basis for making decisions about the relationships between objects. Analyzing these lifelines provides an abstraction mechanism for summaries of video content.

One way this can be accomplished is by development of tuples that describe the contents of our database in terms of lifelines and their attributes. Examples of such tuples include:

- Group (grp_id, topology, lifeline_1… lifeline_n)
- Lifeline (ln_id, geometry, acceleration, cardinality)
- Geometry (spatial, temporal)
- Spatial (node_position_1… node_position_n)
- Temporal (node_position_1… node_position_n)

where the user determines n in all instances.

While many tuples could be formulated with our system, those above have been selected as most appropriate for our particular applications. We are certainly interested in the geometry that defines each particular lifeline, and attributes of acceleration and cardinality help describe where the object has changed its movement trends. Including topology when grouping lifelines provides a means of relating each lifeline to others around it. All of these attributes are used in our system. While it is instructive to look at each of these tuples on their own, it is also important to realize that they can be organized in a hierarchical manner. This hierarchy can be expressed in terms of data cubes or their extensions, which will be discussed in Section 4 of this paper.

By breaking the geometry down into spatial and temporal components, we allow the user some freedom in choosing to emphasize one of these dimensions over the other. Of course, it is also possible to look at a combination of spatial and temporal information, which most closely represents the overall content of the video sequence in question.

In looking at spatial coordinates (x and y dimensions), we are primarily interested in the movements of a given object in the video space and the corresponding attribute of cardinality, which changes at each of the nodes. When examining temporal coordinates (z dimension), we are more interested in when the object changes its velocity than in its specific path. In this case, the attribute of acceleration becomes important. An example of the use of lifelines in our system may be instructive.

3.2 Lifeline Example

Topology and other group-oriented attributes can be examined when lifelines are formed into functional groups based on the user's needs. For instance, if a video camera is set up on a tall building to monitor the traffic flow on surrounding roadways, it may be of interest to determine the locations of bottlenecks before the traffic backs up too much.

If each car in the scene is picked out as a separate object and lifelines are constructed in real time, cars on a given road can be grouped together and their topology or average separation can be computed. A group of cars with a low average separation may represent a potential bottleneck, in which case a navigation system could suggest alternate routes for motorists, or a change can be triggered in the timing of traffic lights or lane allocation.

Once all the key entities and their attributes have been extracted from a video sequence of interest, we must turn our attention to how they can be organized to best facilitate queries. A current method that has gotten much attention is the data cube, a tool often used for decision-making processes in businesses. We feel that it can also help us organize our video data.

4. DATA CUBES

4.1 Data Cubes

Data cubes work on relational databases to provide a means of facilitating query performance. These cubes are composed of dimensions, representing categories of data. Within data cubes, dimensions are organized into hierarchies, or levels. The data values to be analyzed are known as the measures of the cubes, and these can be analyzed through queries involving combinations of dimensions and hierarchies (Harinarayan et al, 1996).

To further our query system, we are concerned with the construction of such cubes. The tuples set forth in the description of lifeline abstraction in Section 3.1 will be used as a working example and can be easily adapted for this task. In formulating data cubes, one uses fact tables, which contain data about the topic of interest. Following the traffic example given previously, we will concentrate on group behaviors as our topic of interest. Thus, our fact tables would consist of multiple tuples of the type:

- Group (grp_id, topology, lifeline_1… lifeline_n)

In order to populate our database with these tuples, we would specify a roadway and collect all lifelines with nodes falling within the boundaries of the road. These lifelines would then compose one group, and topology could also be computed.

Once all desired groups of lifelines have been extracted and added to the fact table, we can then look to a data warehouse, where dimension tables are stored. Dimension tables contain information related to the components of the fact table. In our system, the following would be stored in dimension tables:

- Lifeline (ln_id, geometry, acceleration, cardinality)

The query system would refer to a list of lifeline tuples in order to obtain a full picture of all the lifelines within a given group. Data within these dimension tables can easily represent dimension hierarchies, when there are functional dependencies among the attributes of the dimension tables (Mumick, 1997). In our example, ln_id functionally determines geometry, acceleration, and cardinality. Geometry can then be broken down further into its spatial and temporal components. The measures of the cube would be the attributes of acceleration, cardinality, and topology, as well as geometry (node positions).

With the components of the cube in place, a set of likely queries can be defined, such as a list of the nodes at which any object within a group is accelerating. With a small enough set of data, it may be possible to catch all of the most requested queries. If the system is extended to encompass locations of stationary objects within the scene though, the data cube may become too complicated to anticipate all possible user requests.

4.2 Supplements for Cubes

The major problem with the above cubes is that queries must be predefined. We propose a combination of data cubes with other existing mechanisms of data organization, such as pyramids and scale space, in order to allow integration of zooming behaviors and to avoid the need for predefined queries. This also allows us to extend their use to data that is not stored in traditional relational databases.

Pyramids have been used to establish discrete levels of zooming in many digital image processing applications. To improve data cubes we are using the same concepts in our lifelines and the extraction of their attributes. More specifically, the tuples listed in Section 3.1 contain some variables that can be set by the user, namely the number of lifelines in a group and the number of nodes defining a particular lifeline. Choosing just a few nodes or lifelines is equivalent to choosing a low level of detail. The higher the value for n, the more detail will be present in the resultant data set. In this manner, a rudimentary pyramid can be formed and integrated into our query system.

Scale space is a concept by which an image or other object of interest can be examined at any point on a continuum of scale, not just at discrete stops. This allows for zooming to exactly the desired level of detail within an image (Lindeberg, 1994). One of the goals of our research is to incorporate scale space into our query system and even bypass our relational tables. In order to accomplish this, we would need to look not only at the nodes that define lifelines, but also at the coordinates of the lines themselves between these nodes. Future research will focus more intensely on this aspect of our query system.

Extracting lifeline geometry and attributes and organizing the resultant data are the first steps in our video query system. However, it is not wholly effective on its own. The user of any such system will inevitably be interested not only in these descriptors of the dataset, but also in metadata about the quality of the observations being made and the source of the data. The next major component in our system builds on traditional metadata schemes in order to address these concerns and to begin the querying process.

5. HIERARCHICAL METADATA

5.1 Content Summaries

We realize that it is often impractical, and sometimes impossible, to list the entire contents of any data set. Thus content summaries are used to get an idea of which data set will be most appropriate for a given application. These summaries list the most important pieces of information about each data set in question, so that quick decisions can be made about their relative usefulness (Hardy & Schwartz, 1996).

In order to facilitate content summaries and their querying, we propose the construction of new metadata structures. Each level of the video hierarchy discussed in the previous sections needs its own metadata components, because we are able to view successively more detail as we move from groups to lifelines to individual nodes. It would not make sense to examine metadata about specific nodes when the user is only interested in an overall description of the data set.

The Federal Geographic Data Committee's (FGDC) metadata standards are already structured in a somewhat hierarchical manner, with the following categories on their topmost level:

- Identification
- Data Quality
- Spatial Data Organization
- Spatial Reference
- Entity and Attribute
- Distribution
- Metadata Reference

Each of these categories of information has been subdivided further in order to fully elucidate the metadata for a specific data set (FGDC, 1998). The data that we extract from our lifeline abstraction process about geometry, acceleration, cardinality, and topology is most closely linked to the "Entity and Attribute" category, and can be added as an additional piece of metadata for querying purposes.

5.2 Additions to FGDC Metadata

If someone wanted a broad overview of the contents of the dataset, FGDC metadata on topics such as distribution or metadata reference would be appropriate. Once the user decides that the data set may be of some interest and starts zooming in to look at specific entities though, information about these entities and their attributes and about data quality may become more important.

For example, suppose a bank robbery had taken place, and the user wanted to track all cars that accelerated and turned toward the highway within a five-minute window after the crime is committed. In this case, the most important metadata would deal with entities and attributes on the FGDC level. It would also be helpful to know how accurate we could expect the results to be. Both temporal and geometric accuracy would need to be scrutinized at the level of specific nodes. This accuracy may differ from that of the groups or lifelines, due to error propagation (Clarke, 1998). Thus it is important to specify a level of interest, either through data cubes or one of the alternative zooming methods presented in Section 4.

Our system is designed to ask the user which level of the hierarchy is most important. Then the appropriate metadata categories could be presented for input and analysis through feedback systems, such as those presented in Section 6.

6. FEEDBACK SYSTEMS

6.1 Basic Operation

The last major component of our current work is the use of intelligent feedback systems, whereby a user could express preferences for such qualities as coarseness of spatial resolution, frequency of event occurrence, and overall accuracy of data. These preferences would be gathered before a query is processed and would be used in mapping a path through the hierarchical metadata for use in processing.

In our system, the user is presented with a list of descriptions for each of the seven FGDC metadata categories, including a sample of the subdivisions within each category. The main categories can then each be given weights between zero and one, with one meaning that only exact matches between the requested metadata and that found in the metadata file would be deemed acceptable. A weight of zero would mean that any response in the metadata file would be acceptable.

Any of the seven categories given a ranking above zero is then subdivided into the next level of the hierarchy, and these could be weighted as well. For instance, if the user were interested in the results of the Data Quality category, these levels would be presented:

- Attribute Accuracy
- Logical Consistency Report
- Completeness Report
- Positional Accuracy
- Lineage

The user can then assign weights to these subdivisions. This continues until either all elements on the lowest level are given weights of zero, or no further subdivisions can be made.

One of the biggest challenges of the system would be in determining whether a match had actually been made or not. In this case, algorithms would need to be developed to compare similarity of responses (Sharma 1997). It is relatively simple to compare two numerical responses, and geometric responses can be mapped out and compared by computing distances. It is much more complicated if the category being addressed is answerable only with text however. This area of our system is still under development.

Once similarity measures had been determined, they would then be multiplied by the weights given by the user such that:

$$S = w_1 s_1 + w_2 s_2 + \ldots + w_m s_n \quad (1)$$

where S = total similarity
w_i = weight assigned by user
s_i = similarity measure for given metadata

These weights are based on the importance that each user gives to the different categories of metadata, and are used to select a group of video clips that most closely match the desired criteria (Jain, 1994). While an expert user may want to utilize the system fully by choosing all possible weights, the non-expert may desire a more automated system.

Consider a database administrator at a police station, who is in charge of collecting surveillance tapes at local cultural and sporting events, to be used in searching for suspicious activities. This individual would be interested in nearly every aspect of the available metadata, in order to be sure that the source, accuracy, and even the spatial reference data was such that the videos could be utilized by the members of the police department. In this case, all possible metadata categories would be available for ranking.

Individual police officers, on the other hand, would be much more interested in the entity and attribute information that could be extracted from the tapes. They would probably not care much about where the tapes originated or even about the accuracy of the data to be extracted. If the video is available for them to view, it is because the database administrator has deemed it acceptable in these categories. This leaves the individual officers free to concentrate on matching very specific attribute information instead. Defaults could be accepted for the other categories, and only those metadata categories that are of interest at the time would be ranked.

Once all the rankings had been completed and the similarity indices had been computed, the user would be presented with a list of the video clips that most closely fit the given criteria.

6.2 Advanced Systems

After the query results had been presented to the user, additional adjustments could be made to the preferences via user feedback. The users would be allowed to rank the clips that were presented as possible solutions. These rankings would then be used to gauge the criteria being used to make decisions about what metadata are important.

For example, if an individual has given data quality a very high ranking, but then chooses clips near the bottom of the ranked list as being more desirable than those near the top, perhaps lower values for data quality should have been chosen. Internal adjustments could be made to the weights for each category and new sets of possible solutions could be constructed.

This feedback system would allow the user to obtain the best possible results from the query process. Additional features that could prove useful to users include a default set of weights for given applications, as well as storage of preferences for future queries by the same user (Keogh & Pazzani, 1999). In either of these cases, the weights would be handled behind the scenes, unless the user specifically wanted to modify them.

Going back to our original example of a camera mounted on a tall building, information about the data quality, distribution, or metadata reference would probably not be as important as information about entities and attributes within the video sequences. We could assume that the user of this system would already know who had collected the data, and would be more interested in the attributes, so as to determine where a potential bottleneck might be found.

Anyone who uses this system often enough for the same purposes could store profiles with associated weights, so that they would only need to log in, choose their task, and give their preferences for metadata results. The system would do the rest.

7. CONCLUSIONS

The combination of feedback systems with modified data organization tools and new hierarchical structures for metadata represents a powerful new environment for querying video databases. While there is still work remaining to be done both on the underlying theories and their implementation, we believe that the proposed system has the potential to make video querying faster and more effective.

REFERENCES

Clarke, J., 1998. Modelling uncertainty: a primer. Technical Report 2161/98, University of Oxford, Dept. Engineering Science.

Federal Geographic Data Committee (FGDC), 1998. FGDC-STD-001-1998 "Content Standard for Digital Geospatial Metadata (Revised June 1998)", Washington DC. http://www.fgdc.gov/standards/documents/standards/metadata/v2_0698.pdf.

Hacid, M., C. Decleir, J. Kouloumdjian, 2000. A database approach for modelling and querying video data. *IEEE Transactions on Knowledge and Data Engineering*, 12(5), pp. 729-750.

Hardy, D., M. Schwartz, 1996. Customized information extraction as a basis for resource discovery. *ACM Transactions on Computer Systems*, 14(2), pp. 205-216.

Harinarayan, V., A. Rajaraman, J. Ullman, 1996. Implementing data cubes efficiently. *Proceedings of the ACM SIGMOD Int. Conference on Mgt. of Data 1996*, pp. 205-216.

Jain, R., A. Hampapur, 1994. Metadata in video databases. *ACM SIGMOD Record*. 23(4), pp. 27-33.

Keough, E. M. Pazzani., 1999. Relevance feedback retrieval of time series data. *Proceedings of the 22nd Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval.*

Lindeberg, T., 1994. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2), pp. 225-270.

Mumick, I., D. Quass, B. Mumick, 1997. Maintenance of data cubes and summary tables in a warehouse. *Proceedings of the ACM SIGMOD Int. Conference on Mgt. of Data 1997*, pp. 100-111.

Oomoto, E., K. Tanaka, 1993. OVID: design and implementation of a video-object database system. *IEEE Transactions on Knowledge and Data Engineering*, 5(4), pp. 629-643.

Sharma, V., 1997. Organization and visualization of metadata for spatial information in a digital library. Master's Thesis, University of Maine, Orono, ME.

Stefanidis, A., P. Partsinevelos, K. Eickhorst, P. Agouris, 2001. Spatiotemporal lifelines in support of video queries. *Proceedings Twelfth International Workshop on DEXA*, pp. 865-869.

ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation through grants DGI-9983445 and ITR-0121269.

WG IV / 6:

LANDSCAPE MODELLING AND VISUALIZATION

HIGH RESOLUTION DIGITAL SURFACE MODEL (DSM) GENERATION USING MULTI-VIEW MULTI-FRAME DIGITAL AIRBORNE IMAGES

B. Ameri, N. Goldstein, H. Wehn, A. Moshkovitz, H. Zwick

MacDonald Dettwiler and Associates (MDA)
13800 Commerce Parkway, Richmond, BC. Canada
Tel: (604) 231-2308, Fax: (604) 278-2117
bameri@mda.ca, norman@mda.ca, hw@mda.ca, amoshkov@mda.ca, hhz@mda.ca

Commission IV

KEY WORDS: Multiple image matching, DSM, Graph theory, Sensor calibration, Digital cameras

ABSTRACT:

FIFEDOM (Frequent-Image-Frames Enhanced Digital Ortho-rectified Mapping) is a multidisciplinary ongoing research project aiming to provide low-cost and high information content digital mapping by developing an intelligent digital airborne image acquisition strategy, and novel image processing techniques. In particular, FIFEDOM is designed to exploit the high degree of overlapping of the digital aerial imagery to generate accurate map products. FIFEDOM utilizes high-overlap digital image acquisition to provide multiple-look-angle reflectance for each pixel in a scene. The FIFEDOM project concentrates on four technology areas, Image Acquisition System and Sensor Calibration, Radiometric Balancing, Digital Surface Model (DSM) generation, and Bi-directional Reflectance Distribution Function (BRDF) model extraction.

This paper emphasizes the FIFEDOM DSM module. It introduces an innovative, and breakthrough methodology for automated DSM generation. This is achieved through utilization of multi-view, multi-frame highly overlapped digital images. This capability improves the result significantly by generating a very dense, high quality, and reliable Digital Surface Model. The paper also discusses sensor calibration and radiometric-balancing, as they are prerequisite steps for the successful completion of the DSM task.

1. INTRODUCTION

Since the seventies, an increasing amount of research work has been reported aiming to replace the manual time consuming image mensuration process with a semi/fully automated operation (Helava 1976, Förstner 1982, Ackermann 1984). This investigation leads to a series of more complex and successful commercially available systems, for e.g. automated DEM generation or AAT. In spite of a high degree of automation in those systems, still a significant amount of work is required for post-processing and editing of extracted e.g. elevation data especially in problematic areas such as forest or built-up areas. This is due to a number of reasons such as occlusion, height discontinuities, repetitive patterns, shadows, and lack of texture, to name a few. The proposed work is an attempt to overcome some of these issues especially in built-up areas where the user community would like more frequent updates if the cost could be contained. The importance of this project to the geomatic industry is based on the reality that digital airborne cameras are now practical for mapping. Modern CCD performance and cost characteristics make this a cost-effective approach, and multiple digital image pairs are not expensive compared to traditional aerial photography.

The paper introduces an innovative methodology for automated DSM generation. This is achieved through utilization of multi-view, multi-frame highly overlapped digital images. Existing tools are based on stereo image matching techniques, where only two overlapping images are processed. This leads to unsatisfactory results, e.g. in built-up areas, where occlusion is a major issue. The method proposed by FIFEDOM overcomes this problem, if not completely, then at least partially, by applying novel multiple image matching techniques. In this manner, if an area/point is occluded in one image, it is still likely to be seen in other overlapping images, captured from different view angles, and it is possible to reconstruct a 3D representation of the occluded terrain surface. This capability improves the result significantly by generating a very dense, high quality, and reliable Digital Surface Model (DSM). The resulting DSM is subsequently (beyond the scope of this paper) used to create Digital Elevation Models (DEM) and other related image products e.g., true ortho-rectified images.

The paper also briefly discusses other FIFEDOM technology areas, especially sensor calibration and radiometric-balancing, as they are prerequisite steps for the successful completion of the DSM task.

2. OVERVIEW

Figure 1 illustrates a context diagram of the FIFEDOM data processing system. The inputs are D1 12-bit per band raw colour imagery, an ascii file containing GPS/INS data that accompanies the acquired imagery and is generated by the acquisition system's support software, and a product specification determining the processing flow and which of the outputs to generate. The outputs are the camera's radiometric and geometric calibration results (section 3.3), BRDF results (section 5), radiometrically balanced image frames (section 4) and DSM results (section 6).

Figure 2 illustrates the relationship between the DSM module and other FIFEDOM components. The module receives the refined model parameters, the single channel of radiometrically balanced images, and a coarse DSM as input and delivers a TIN/GRID dense DSM as output.

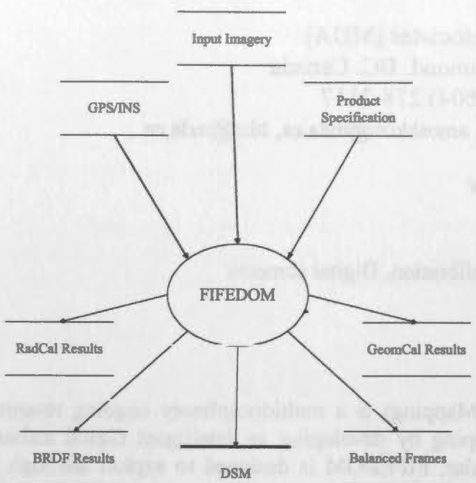


Figure 1: FIFEDOM system context diagram

The refined model parameters are estimated in a post bundle adjustment (BND) process where the initial values are provided using the onboard GPS/INS system during the flight mission. The radiometrically-balanced images are the output of Radiometric Balancing (RadBal) module of FIFEDOM where, using sparse tie points (TIPs), the corrective gains and offsets of the individual images are estimated and applied.

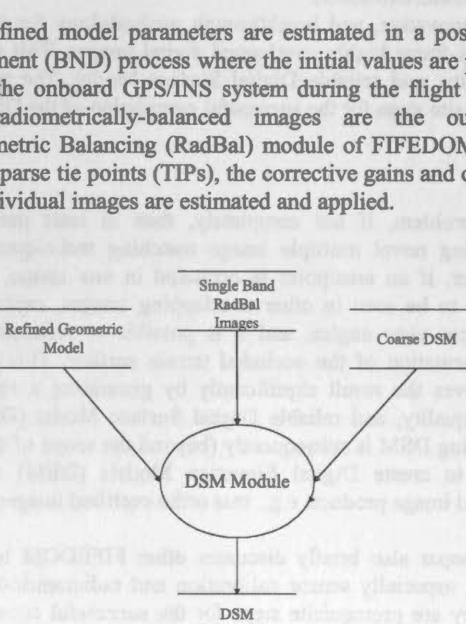


Figure 2: DSM module external interfaces

This process is discussed in more detail in section 4. In addition, the sparse TIPs are used to generate a coarse DSM representing an approximation of terrain surface, which is used as a starting point for DSM generation. The balancing is applied in one band only (green band) and is sufficient for 3D-point extraction, but not necessarily to form a globally balanced and calibrated mosaic of the data set. The actual and a more sophisticated RadBal process is applied to all 3 bands in a later stage, where the output of the DSM module is used as an input to that stage.

The following sections contain a detailed description of the above processes.

3. FIFEDOM ACQUISITION SYSTEM

This section describes how imagery from the D1 camera is prepared, as input to the DSM, BRDF and Radiometric Balancing undertakings. There are two steps involved:

1. Extracting imagery from the camera.
2. Applying calibration (geometric and radiometric)

The following sections elaborate on these tasks.

3.1 Acquisition

One factor in choosing the Nikon D1 is its rugged construction. The camera has been mounted on a twin piston engine, Piper Navajo Chieftain airplane, based in Victoria, BC by the Range and Bearing Environmental Resource Mapping Corporation, who also acquire the project's data sets, and supply the GPS/INS positioning for each frame of imagery.

The camera is mounted so that the long dimension is along track. The acquisitions are executed to obtain 90% overlap in the along track direction, and 60% side lap. This ensures that each point on the ground will be in at least 20 different image frames with 20 different view directions.

3.2 Camera

The Nikon D1 camera lens has a nominal focal length of 14mm and a single focal plane, with the CCD R, G and B pixels interspersed in a standard Bayer pattern (Figure 3). The focal plane has 1324 rows by 2012 columns, half the pixels are green, one quarter are red and one quarter are blue. A thin film low-pass filter covers the entire focal plane, to reduce the aliasing effects within each band.

3.2.1 Image format/preparation: The camera is able to generate various image formats – jpeg, tiff and raw. The format used in the FIFEDOM project is the raw format, as this is simply a dump of the focal plane to a disk file, without any balancing or special effects being applied. The raw imagery allows us to do a stable radiometric calibration (that is valid for a fixed set of camera settings).

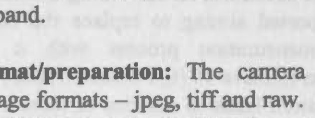


Figure 3: Bayer pattern of CCD on the D1 focal plane.

The raw image is in a proprietary format, which Nikon calls NEF, which is actually (non-standard) TIFF. The FIFEDOM project derives several imagery formats from the NEF file: Two of these are:

1. *Bilinearly Interpolated.* This is the workhorse format for generating DSMs and orthophoto mosaics. Each band is bilinearly resampled to the grid of the focal plane, resulting in the three bands being geometrically registered with each

other. The resulting imagery loses a 1-pixel border around the focal plane, and has dimensions 1322x2010.

2. *Separated RGB.* In order to provide the project with imagery that is unsullied by the bilinear interpolator, we are also able to extract the three bands from the NEF file, and place them into their own separate image files. No two bands are registered, but all the pixels are unaltered from the NEF file.

3.3 Calibration

An important goal of the FIFEDOM project is to command accurate placement and interpretation of the imagery. For this, the geometric and radiometric calibrations are essential. Another goal is that the calibration procedures be easily repeatable, so that it is convenient to

1. Verify the stability/accuracy of the calibration, and
2. Try different camera settings.

To a large extent, our geometric calibration procedures are easily repeatable. We are currently experimenting with various radiometric calibration approaches.

3.3.1 Geometric: The geometric calibration is divided into the two phases of camera interior and camera alignment calibration, which are discussed in more detail, below.

Camera Interior: The highly overlapped (low parallax) imagery allows tiepoints to be found completely automatically, simply with image correlation, and with no need of an Earth elevation model. We assume a standard distortion model for the camera (radial and tangential distortions), and use the tie points to constrain the distortion parameters. One caveat is that the focal plane tends to be parallel to the ground, so that a small number of GCPs are also used, to remove the focal length ambiguity. We have achieved 0.1 pixel accuracy in this way.
Camera Alignment: The GPS/INS describe the pose (position and orientation) of the GPS/INS frame, but not necessarily the pose of the camera. To this end, after the camera interior calibration has been applied, the tiepoints are used again, this time to calculate the pose of the camera relative to the GPS/INS frame. This increases the accuracy of the relative positioning of the frames in subsequent data sets, and improves the robustness and correctness of the tie point generation.

3.3.2 Radiometric

The goal of radiometric calibration is to calculate the radiance at the camera for each pixel of an image frame. This is one step in the chain of calculating the ground reflectance of each pixel, which is essential for BRDF investigations and for the absolute comparison of the acquired data across different data sets and even different sensors. The process can be divided into two parts, *relative* and *absolute.*

Relative calibration imposes a flat field on each band, i.e. the radiance values are known up to an (unknown) scale factor that is the same for every pixel of the band. Absolute calibration determines the gains (1 per band) that convert the relative calibration to true radiance units. We are experimenting with various approaches:

1. Calibration panels,
2. Overcast sky shots (with and without a diffusing opal glass filter), and
3. Integrating sphere.

In light of the criteria of correctness, convenience and cost, we are currently evaluating these approaches, but are, for now, using the diffuse sky shots to calculate the relative calibration, which is a set of gains for each band; as depicted in Figure 4.

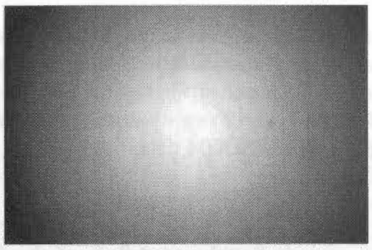


Figure 4: Typical image of 1/gains for relative calibration

4. RADIOMETRIC BALANCING

There are two main goals for radiometric balancing in the FIFEDOM project:

1. Supply a single band of balanced imagery for DSM generation. The green band is used, as it has the highest spatial resolution of the 3 bands.
2. Produce seamless RGB mosaics that are pleasing to the eye, while maintaining, as much as possible, the proper spectral properties of the imagery.

The radiometric balancing approach is empirical, to remove residual errors from the 3 radiometric processes that precede the radiometric balancing step. These 3 processes are:

1. Radiometric calibration: Resulting in camera radiance.
2. Atmospheric correction: Resulting in ground reflectance for the pixel-specific sun and view angles, and
3. BRDF removal: Resulting in ground reflectance for a sun angle and a view angle that are the same across all pixels and images in the data.

Moreover, it is not always the case that all 3 of these processes will be applied to a data set. In such a situation, the radiometric balancing must still result in a balanced set of images, albeit not necessarily tied to engineering units.

The radiometric balancing approach is to lay a coarse grid on each image, with a gain/offset at each grid location. The tie points are used to constrain the gains and offsets, which are then applied to each image. At non-grid locations, the gain/offsets are obtained by bilinear interpolation from those at the grid locations.

5. BRDF

FIFEDOM data sets allow a user to view a given region on the ground from many different view angles. For example, for a typical FIFEDOM data set, there are about 26 distinct frames, which view a region from different zenith and azimuth angles. Thus, since the Nikon D1 is an RGB camera, this corresponds to an image with 78 "bands". This large number of bands can be compared with hyperspectral images. Moreover, since the FIFEDOM data have 12-bit radiometric resolution, even small changes in view-angle dependent image brightness can be detected.

This great wealth of information about the view angle dependent reflectance of the ground can be fitted to a bi-directional reflectance distribution function (BRDF) model. Since this BRDF model for an object is characteristic of that object, it can serve as the basis for a scene classification.

FIFEDOM explores a wide range of the issues associated with BRDF-based land classification. For example, for forestry applications it is hoped that physical parameters of the forest canopy such as LAI, crown closure, tree geometry etc. can be derived from the FIFEDOM data.

6. DSM

This section describes the workflow of the FIFEDOM DSM module and discusses its major components and their interrelationships. The process is performed based on a hierarchical coarse to fine strategy, where the final output is a seamless DSM that covers the entire area. The process consists of four fundamental steps as follows:

1. Image Pyramid Generation – images are down sampled to different resolution layers.
2. Point Feature Extraction – a list of interesting points is extracted individually for each of the images.
3. Multiple Points Matching – the extracted points in the different images are matched based on similarity measures and N–partite graph solution.
4. Multiple Forward Intersection – match points consisting of 2 or more points enables to reconstruct the 3D location of the object point on the terrain. Reconstruction of the object point and rejection of false points is achieved through a Least Square solution.

6.1 Initial Search Space

One of the fundamental tasks in the proposed DSM generation algorithm is to identify and to measure conjugate (homologous) points in two or more overlapping images, a process known as feature/image matching. There are many interesting aspects to image matching such as determination of initial values and suitable approximations, the selection of exact correspondence from multiple solutions, outliers detection and removal, and mathematical modelling of terrain surface, to name a few. In this section we focus on the constraints, which restrict the space of possible matching solutions, such as setting bounds on the search area, which also begins the image matching process rather close to the true solution.
The FIFEDOM camera is a frame camera with a central perspective projection. Central perspective projection provides a very powerful constraint, namely that of epipolar geometry. Given two images and a 3D point in object space the epipolar geometry is defined as a plane containing the object point and the two projection centers of both images. If the relative orientation of two images is known, for a given point in one image the epipolar line in the other image can be computed, and the corresponding point must lie on this epipolar line. Thus, the image-matching problem is reduced from a two- to a one-dimensional task. In order to facilitate matching along epipolar lines in a multiple image environment such as FIFEDOM, the epipolar geometry should be constructed on-the-fly using the relative orientation parameters of the images. The epipolar geometry constraint is applied in the process of Multiple Points Matching (see section 6.3), where the process is searching for the candidate homologous points within a search area along the epipolar lines. This constraint is fundamental in reducing ambiguity and computational cost. Even if only approximate values for the parameters of relative orientation are known, the epipolar constraint should be used in order to restrict the search space in the direction perpendicular to the base line.

To further reduce the initial search space in the matching process a hierarchical coarse-to-fine strategy is used. In this process images are represented in a variety of resolutions, leading to an image pyramid. The images are organised from coarse to fine pixel resolution, and results achieved on one resolution are considered as approximations for the next finer level. A coarser resolution is equivalent to a smaller image scale, and a larger pixel size. Thus, the ratio between the flying height and the terrain height increases as the resolution decreases, and local disturbances such as occlusions become less of a problem (Hepkie 1996). In this work a low pass Gaussian kernel is used to generate the image pyramid of FIFEDOM images.

6.2 Point Feature Extraction

In the previous section we discussed some simple but powerful methods to reduce the search space. Now, the issue is the selection of appropriate matching primitives. In fact, the distinction between different matching primitives is probably the major difference between the various matching algorithms. This is true because this selection influences the whole process of the matching. These primitives fall into two broad categories, *windows* composed of grey values (area based matching), and *features* extracted in each image a priori (feature based matching). We selected a feature based matching approach in this work, where the Förstner operator is used to extract salient points as matching primitives in each image individually prior to the matching process (Förstner 1986). The advantage of the point selection is obvious. It leads to a great information reduction, as we only have to deal with a set of points and not with all pixels in the images.

Each extracted point is characterized by a set of attributes. These attributes are the key elements in the success of the upcoming process of Multiple Point Matching, as they are input parameters to a *weight function* (Equation 1), which actually establishes the similarity measures between the candidate homologous points in the overlapping images (see section 6.3 for details).

Note that the interest operator should be carried out on each resolution level of the image pyramid separately, since points can vanish or be displaced from one level to the next due to the low pass filtering which is inherently present when decreasing the resolution. The lists of selected points and their associated attributes for every image are passed to the next module to establish the correspondence relationships between homologous points in overlapping images.

6.3 Multiple Point Matching

The objective of the multiple point matching is to determine precise locations of homologous points from n images ($n \geq 2$). All the points should contribute simultaneously to the solution to exploit a major advantage that digital image matching offers. This results in a higher redundancy for the matching problem and thus a greater reliability is achieved for the results (Baltsavias 1991). The following two sections discuss the overall concept of the multiple point matching problem and the potential solution to this problem.

6.3.1 Multiple Point Matching Problem: Figure 5 illustrates the multiple point matching problem. We have 5 overlapping images, all of which contain the same building. Point p_1 in image 1 is a point of interest (the upper right corner of the building). The multiple point matching problem in this case is to find the set of points in all the other images that best match point p_1 in image 1. Each of the images 2 to 5 have a set of candidate points which potentially match point p_1. These are the red and green points (the blue points are candidates for other points, but not for point p_1). The image 2 has two candidate points, and images 3, 4, and 5 each have 3 candidate points. It is impossible to have a candidate point from the same image as point p_1.

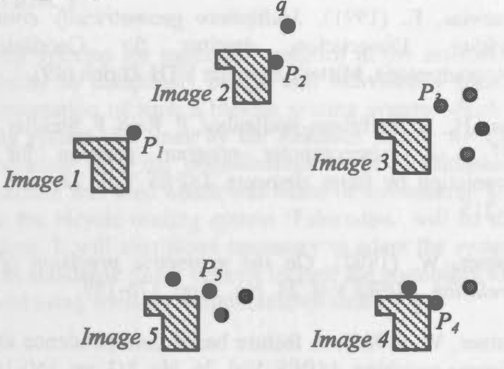


Figure 5: Multiple point matching problem

By simply looking at the figure, we see that points p_2, p_3, p_4, and p_5 (drawn in red) are the set of points that best match point p_1 in image 1. But how do we determine algorithmically that this is the optimal solution?

6.3.2 N-partite maximum-weight clique problem: The first step in solving the problem is to associate a weight, or a measure of similarity, between p_1 and each candidate point for p_1 in the other images. The similarity weight is computed based on the weight function f_{ij}:

$$w_{ij} = f_{ij}(s_1, s_2, s_3, \ldots) \qquad \begin{matrix} i = 1,2,\ldots n \\ j = 1,2,\ldots n \\ i \neq j \end{matrix} \qquad (1)$$

Where w_{ij} is the correspondence weight, s_1, s_2, s_3, ... are the similarity measures between the candidate homologous points such as normalized cross correlation, and n is the number of overlapping images that contain candidate points for p_1.

Once we have generated the weights for every pair of candidates points, we then construct an undirected weighted n-partite graph $G=(V, E)$ where V is the set of all distinct candidate points and E is generated by considering the weight of all the points previously matched. The Figure 6 shows the resulting partially drawn graph. All the edges connecting candidate points are drawn in for point p and q, along with their weights. There would certainly be more edges connecting candidate points for points p_2, p_3, p_4, and p_5, but the graph would be far too crowded

Once this graph G is constructed, we solve the problem of multiple points matching by finding a set of vertices, which form a *maximum weight clique.*

Thus we can reduce the multiple point matching problem to the maximum-weight clique problem. The edges that connect the maximum-weight-clique are drawn with solid lines in the Figure 6. All other lines are dashed. A similar approach is reported by Tsingas (1994) to solve the matching problem in the digital point transfer process.

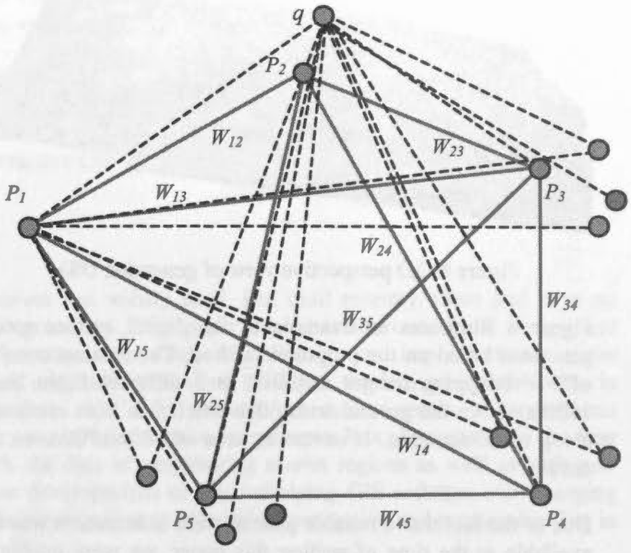


Figure 6: n-partite undirected weighted graph

6.4 Multiple Forward Intersection

The idea is that the 3D position of the matching points in object space (terrain/model co-ordinates) should be determined from the simultaneous contribution of the 2D co-ordinates of the homologous points in n (n ≥ 2) images. This can be obtained directly by spatial forward intersection of corresponding space rays as the result of a least squares solution, as depicted in Figure 7. The mapping relation between the point in 3D object space and its perspective projection in 2D image space is represented by the classical collinearity equations (Ameri 2000). The required mapping parameters are derived through a bundle adjustment process in earlier phases, which form the refined model parameters and are input to the DSM module.

The cloud of 3D points generated in each pyramid level n is used as a coarse approximation of terrain heights for the next level $n+1$. Therefore a simple terrain modelling process, for the time being, is advised in order to filter out the erroneous points and convert the cloud of 3D points into a possibly outliers-free TIN data structure, and from that to a GRID structure.

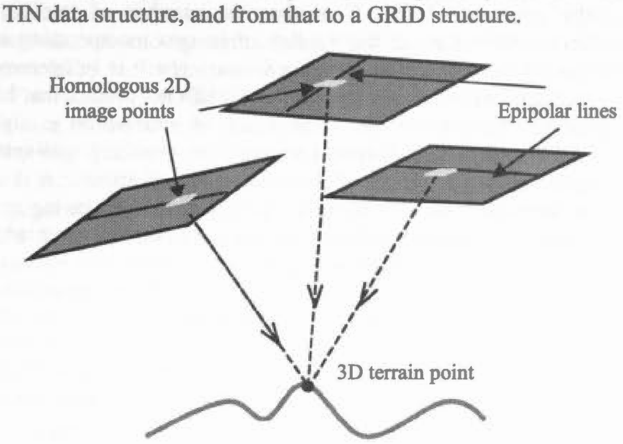


Figure 7: Multiple Forward Intersection

A more robust approach to terrain surface modelling with finite elements technique is reported by Ebner et al. (1980), where surface curvatures are introduced as constraints in order to regularize the model surface in areas where 3D measured points are missing.

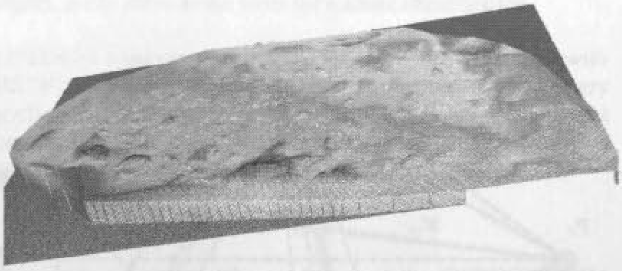


Figure 8: 3D perspective view of generated DSM

Figure 8 illustrates an example of the digital surface model generated based on the proposed method. The data set consists of 7 overlapping images acquired in 3 different flight lines, 1500m above the ground with 60% overlap in both cross and along track direction. It covers an area of 900mx700m on the terrain.

Due to the fact that a reliable ground truth information was not available at the time of writing this paper, we were unable to perform a true analysis on the quality of the resulting DSM. However, since the 3D positions of the points are derived based on a LS solution, the estimated variances (σ^2_x, σ^2_y, σ^2_z) of the terrain point co-ordinates are used as the quality measures to (currently) detect and remove the potential outliers from the result.

<table>
<tr><th colspan="7">Planimetric threshold $t_p = 0.5 \times GSD = 0.5m$
Altimetric threshold $t_h = 1.5 \times t_p = 0.75m$
Total # of 3D points = 9822; Total # of outliers = 1305</th></tr>
<tr><td># of intersection rays</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr>
<tr><td># of reconstructed 3D points</td><td>898</td><td>5580</td><td>2364</td><td>830</td><td>150</td><td>0</td></tr>
<tr><td>% of detected outliers</td><td>19.6</td><td>18.7</td><td>3.7</td><td>0</td><td>0</td><td>0</td></tr>
</table>

Table 1: Relative number of potential outliers vanishes with an increasing number of images.

Table 1 summarizes some of the statistical measures regarding the generated DSM. Note that the number of outliers is approaching zero as the number of images incorporating into the solution increases. This preliminary result is in agreement with the initial motivation of the FIFEDOM project that high information content and great wealth of information in highly overlapped image frames improves the resulting geo-spatial production significantly. The output of this component is in fact an input to other process such as Radiometric Balancing or an additional filtering process to generate the DEM product, which are outside the scope of this paper.

7. CONCLUSION

We have introduced a new method for the reliable and efficient reconstruction of a digital surface model through utilization of multiview, multiframe highly overlapped digital images. The results presented are intermediate results and thus some of the processes are only implemented in a simplified form such as terrain surface modelling and outliers detection. These processes will be fully implemented in the next stage of this study, which subsequently improves the result and efficiency of the proposed method.

8. REFERENCES

Ackerman, F. (1984), Digital image correlation: performance and potential application in photogrammetry, *Photogrammetric Record*, Vol. 11, No. 64, pp. 429-439.

Ameri, B. (2000), Feature Based Model Verification (FBMV): a new concept for hypothesis validation in building reconstruction, *IAPRS* Vol. 33, part B3, Amsterdam, The Netherlands

Baltsavias, E. (1991), *Multiphoto geometrically constrained matching*, Dissertation, Institut für Geodäsie und Photogrammetrie, Mitteilungen der ETH Zürich (49).

Ebner, H., B. Hoffmann-Wellenhof, P. Reiß, F. Steidler, (1980), *HIFI - A minicomputer program package for height interpolation by finite elements*, *IAPRS* Vol. 23, part B4, pp. 202-215.

Förstner, W. (1982), *On the geometric precision of digital correlation*, *IAPRS* Vol. 24, No. 3, pp. 176-189.

Förstner, W. (1986), A feature based correpondence algorithm for image matching, *IAPRS*, Vol. 26, No. 3/3, pp. 150-166.

Heipke, C. (1996), Overview of image matching techniques, *OEEPE Workshop on the Application of Digital Photogrammetric Workstations*, OEEPE Official Publications No. 33, 173-189

Helava, U. V. (1976), Digital correlation in photogrammetric instruments, *IAP*, Vol. 21, No. 3.

Schenk, T. (1999), *Digital Photogrammetry*, Vol. 1, TerraScience, OH, USA.

Tsingas, V. (1994), A graph-theoretical approach for multiple feature matching and its application on digital point transfer, *IAPRS* Vol. 30, No.3, pp. 865-871.

9. ACKNOWLEDGEMENTS

The authors express their gratitude to Precarn Incorporated for partial funding support of this project. Also Triathlon Ltd. for helping to define the DSM requirements and continued evaluation of the results. We acknowledge the Range and Bearing Environmental Resource Mapping Corporation for acquisition of the test data.

DESIGN AND IMPLEMENTATION OF A GIS BASED BICYCLE ROUTING SYSTEM FOR THE WORLD WIDE WEB (WWW)

Manfred Ehlers, Stefan Jung and Katrin Stroemer
Research Center for Geoinformatics and Remote Sensing (FZG)
University of Vechta
P.O. Box 1553, D-49364 Vechta, Germany
mehlers@fzg.uni-vechta.de

Commission IV, WG IV/6

KEY WORDS: GIS Based Bicycle Route Planner, Internet GIS, Interactive GIS

ABSTRACT:

Routing systems for motorized vehicles in the internet are well known and widely used. But until recently, there had been no possibility to completely design and individually plan a tourist bicycle tour. This paper describes the conceptual design and implementation of such a bicycle routing system which has been operational since July 2001. Design and implementation of the routing system was done by the Research Center for Geoinformatics and Remote Sensing (FZG) at the University of Vechta in coorporation with the organization 'Verbund Oldenburger Münsterland e.V.'. To achieve the desired goals, a server/client architecture was used which was based on commercial GIS, database system, and internet components. It is planned that in the near future the bicycle routing system 'Fahrradies' will be extended with the data of neighboring tourist regions as well as with new functions. It will also prove necessary to adapt the system to the new developments of the underlying GIS software and emerging internet standards. Other options include the possibility to integrate the routing system with online navigation and positioning data in the field using wireless communication technology.

1. BICYCLE TOURIST REGION 'OLDENBURGER MÜNSTERLAND'

The internet has become a very important medium for communication and provides several ways of using maps and geographic data. Today, maps and data created with Geographic Information systems (GIS) can easily be published on the WWW. Geographic data are a necessary database especially for routing systems which are well known and widely used for motorized vehicles. In contrast to these, the "Fahrradies" is a routing system, which allows the complete design and individual planning of a bicycle route through the tourist region "Oldenburger Münsterland" in Lower Saxony (Northwest of Germany). With little differences in altitude, the area is perfectly suited for people of all ages to go bicycling and has an extensive network of marked bike paths. The organization 'Verbund Oldenburger Münsterland e.V.' is in charge of the regional marketing (culture, tourism, and economy) for the region which is formed by the two counties of Vechta and Cloppenburg. Having used the internet for marketing for several years, the Verbund Oldenburger Münsterland has now expanded its homepage with the bicycle routing system 'Fahrradies' (mixture of the German words 'Paradies' and 'Fahrrad'; engl. paradise/bicycle) for bicyclists interested in touring the region.

2. ARCHITECTURE OF THE 'FAHRRADIES'

Design and implementation of the routing system was done by the Research Center for Geoinformatics and Remote Sensing (FZG) at the University of Vechta. To implement this, a server-side architecture was used which was based on commercial GIS, database system and internet components (Schneeweiss, 2001). The server provides GIS applications and a graphical web user interface based on HTML forms which are enhanced by JavaScript. The software basis for the application is formed by the desktop GIS ArcView 3.2 with the extension Network Analyst and the ArcView Internet Map Server (IMS). As the source for the necessary stops, the entries from the existing data sets (lists of hosts, list of sights and recreational facilities) in MS Access format are used. These entries are transformed into HTML format and incorporated into the server architecture. The client (= user) accesses the bicycle routing system through a standard internet browser such as Netscape or MS Internet Explorer. The communication is facilitated through HTML forms and JPEG graphics. The GIS based routing analysis, however, is done in vector mode, the results are displayed as raster graphics for standardized communication (Kraak, 2001). The bicycle paths network marked in the analog maps serve as the source for the digital lines. All calculations are performed on the server, the results are transformed into HTML pages and sent to the client. Consequently, there is no need for the internet user to have a GIS installed at his computer.

In contrast to the inflexible way of using an analog map, the user has the possibility to form his own route, with thematic routes and selectable stops (points) taken as a basis. An interactive map and extensive information on sights, recreational and accommodation facilities give the user the possibility to plan an individual bicycle tour. But the "Fahrradies" does not only provide an easy-to-use planning tool, it is also free of charge and available at all times through the World Wide Web. It can be accessed through the homepage of the 'Oldenburger Münsterland' (http://www.oldenburger-muensterland.de/) or directly through the homepage which is maintained by the FZG (http://www.fahrradies.net). To keep the system up-to-date, it is maintained at a central place using a dedicated server. For better performance, the 'Fahrradies' will be upgraded using a second dedicated server system for data storage.

This bicycle-routing-system is now being newly implemented from a previous version to improve its performance and graphical representation and to extend the functionality. The

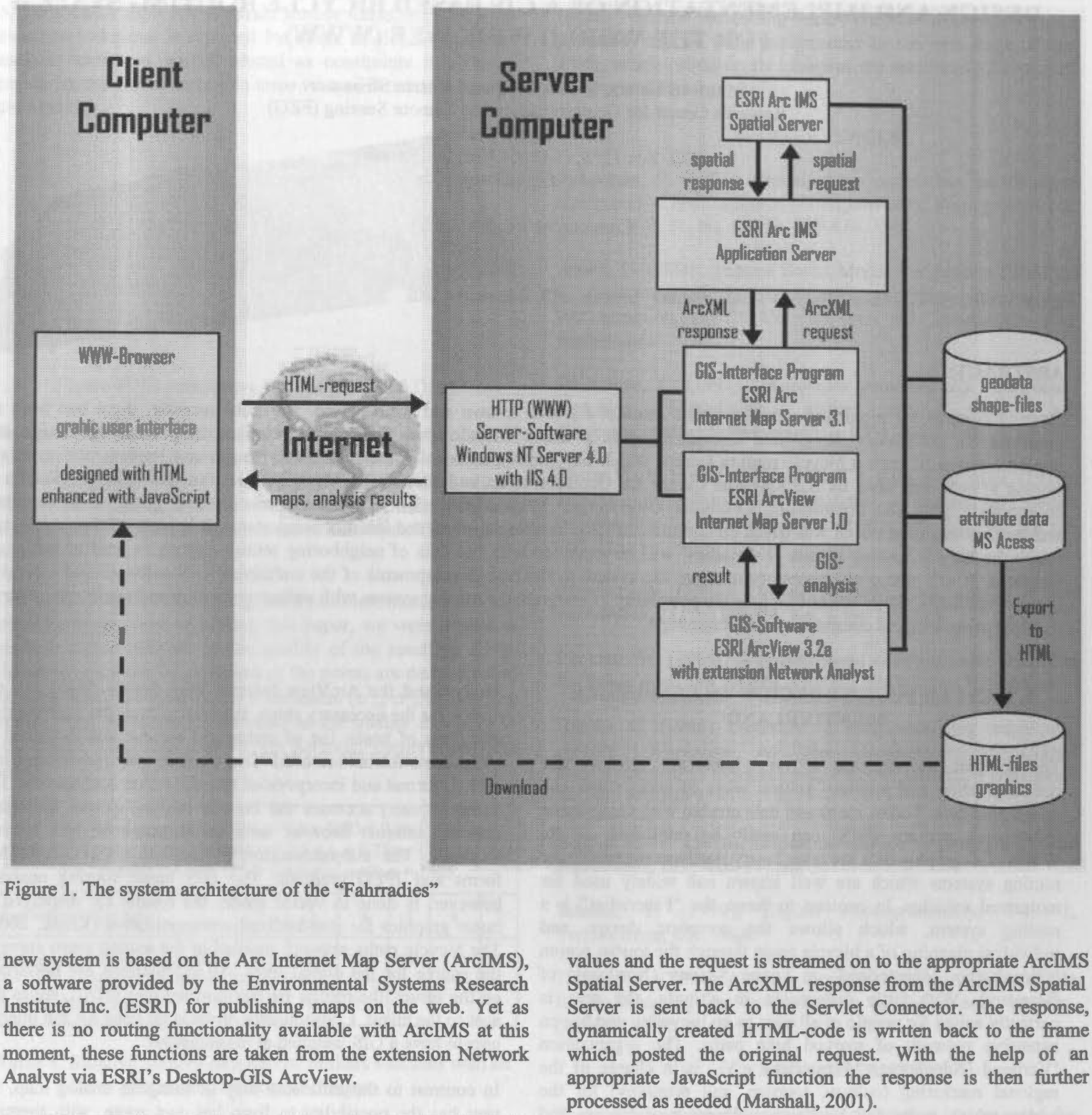

Client Computer
Server Computer
ESRI Arc IMS Spatial Server
spatial response
spatial request
ESRI Arc IMS Application Server
ArcXML response
ArcXML request
GIS-Interface Program ESRI Arc Internet Map Server 3.1
geodata shape-files
WWW-Browser
grahic user interface
designed with HTML enhanced with JavaScript
HTML-request
Internet
maps, analysis results
HTTP (WWW) Server-Software Windows NT Server 4.0 with IIS 4.0
GIS-Interface Program ESRI ArcView Internet Map Server 1.0
attribute data MS Acess
result
GIS-analysis
GIS-Software ESRI ArcView 3.2a with extension Network Analyst
Export to HTML
HTML-files graphics
Download

Figure 1. The system architecture of the "Fahrradies"

new system is based on the Arc Internet Map Server (ArcIMS), a software provided by the Environmental Systems Research Institute Inc. (ESRI) for publishing maps on the web. Yet as there is no routing functionality available with ArcIMS at this moment, these functions are taken from the extension Network Analyst via ESRI's Desktop-GIS ArcView.

The "Fahrradies" consists of a GIS-Application based on the above mentioned software on the server-side and a graphical web user interface based on HTML and JavaScript on the client-side (Fleming, 1998). The architecture of the GIS application consists of three main components: the WebServer, the Application Server and the Spatial Server. The Application Server handles the incoming requests and hands it off to the appropriate Spatial Server, which is the backbone of ArcIMS. It processes the requests for maps and related information, e.g. creates cartographic map image files or streams map features.

Communication between the client and the server is based on requests and responses written in ArcXML. A JavaScript function writes the appropriate ArcXML request which is then sent to the server. The request is then passed via JavaScript on to a frame where it updates a form input value. This form is submitted to the Application Server, which redirects it to the Servlet Connector. The Servlet Connector extracts the submitted values and the request is streamed on to the appropriate ArcIMS Spatial Server. The ArcXML response from the ArcIMS Spatial Server is sent back to the Servlet Connector. The response, dynamically created HTML-code is written back to the frame which posted the original request. With the help of an appropriate JavaScript function the response is then further processed as needed (Marshall, 2001).

The custom bicycle route is calculated with the help of ArcView Internet Map Server. In this case there are no ArcXML requests involved but the values of an HTML form posted to ArcView IMS (esrimap.dll). The Internet Map Server then sends the request to the GIS ArcView which executes the appropriate Avenue script. The result is streamed back to the client in form of HTML code. This consists of a form with information about the extent of the requested bicycle route and the name of the created route shapefile. An ArcXML request then adds this bicycle route shapefile as acetate layer to the main map. This way, only the calculation of the route is run by ArcView/Avenue scripts. Everything else, e.g. the graphic presentation of the route, is processed via ArcXML requests by ArcIMS.

3. FUNCTIONALITY

The key features of the „Fahrradies" include display and query functionality, search for route templates, and calculation of individual biking routes with different routing functions (Jung and Schneeweiss, 2001).

The "Fahrradies" includes a comprehensive set of map navigation tools (e.g. pan, zoom). It also contains a query tool which can be used to get information about a specific location or vice versa to get the location for a specific stop. An advantage of the newly implemented "Fahrradies" application in comparison with the previous version is the overview map, which helps the user to still orient himself while he zooms into the main map. In addition to that, there is an assistant designed for inexperienced users. This assistant guides the user through the application to help him and to avoid mistakes. The "Fahrradies" furthermore also offers comprehensive help which explains each button of the system. Moreover, the user has the opportunity to select one of many "ready-made" bicycle routes which are based on the marked bicycle network of the region. For every single of these so called route templates textual information as well as an overview map are provided.

The biggest difference to other routing systems is the ability for the user to design a completely individual route: It's possible to change the route templates by adding stops or to select stops for an individual route. For this, there are lists with possible stops, sorted by the categories hosts, sights, recreational facilities, and the towns of the region. For each stop, specific map and addresses or background information is available. Besides, the user can get information about one stop by clicking on its symbol in the main map. After choosing stops from the lists a starting point has to be set for the routing calculation. Then the following calculation process starts and the resulting bicycle route is shown in the main map. After the calculation of a route, more tools are made available to the user. These tools are associated with the route in such a way that he can get a route description which lists all chosen stops in a HTML page. Moreover, the distances between the stops and directions to them are listed and this HTML page can be printed out and taken to the field.

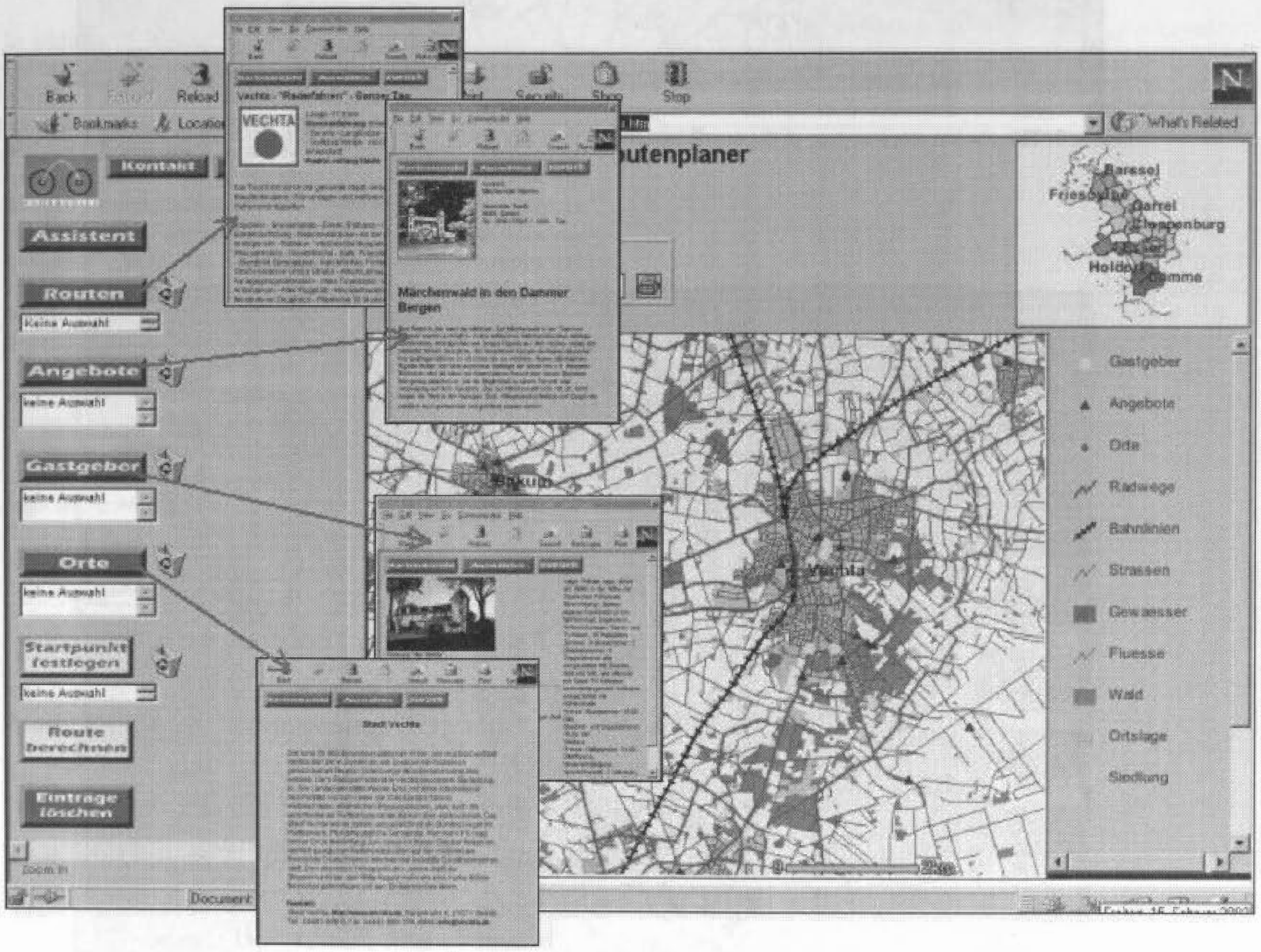

Back
Reload
Stop
Bookmarks
Location
What's Related
Kontakt
Assistent
Routen
Keine Auswahl
Angebote
keine Auswahl
Gastgeber
keine Auswahl
Orte
keine Auswahl
Startpunkt festlegen
keine Auswahl
Route berechnen
Einträge löschen
Document
utenplaner
VECHTA
Märchenwald in den Dammer Bergen
Stadt Vechta
Baresel
Friesoythe
Garrel
Cloppenburg
Holdorf
Damme
Vechta
Gastgeber
Angebote
Orte
Radwege
Bahnlinien
Strassen
Gewaesser
Fluesse
Wald
Ortslage
Siedlung

Figure 2. Information about the different stops, sorted by category.

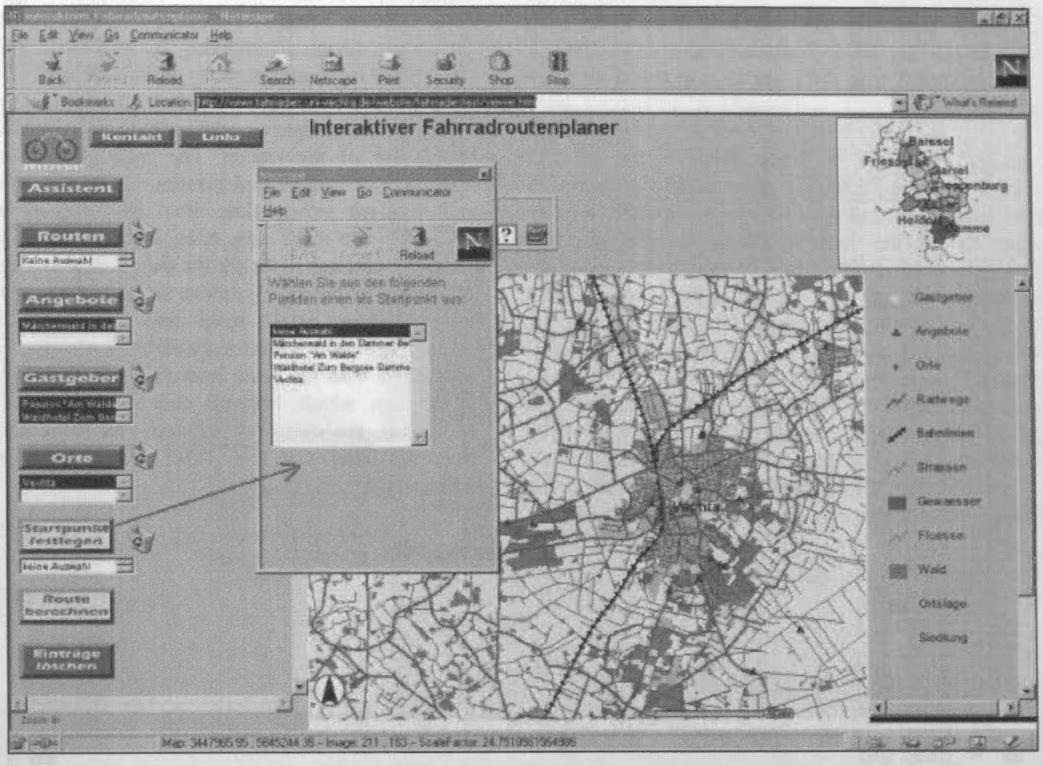

File Edit View Go Communicator Help
Back
Reload
Search
Netscape
Print
Security
Shop
Stop
Bookmarks
Location
What's Related
Interaktiver Fahrradroutenplaner
Kontakt
Links
Assistent
Routen
Keine Auswahl
Angebote
Gastgeber
Orte
Vechta
Startpunkt festlegen
keine Auswahl
Route berechnen
Einträge löschen
Wählen Sie aus den folgenden Punkten einen als Startpunkt aus
Pension "Am Walde"
Vechta
Barssel
Friesoythe
Garrel
Cloppenburg
Holdorf
Damme
Gastgeber
Angebote
Orte
Radwege
Bahnlinien
Strassen
Gewaesser
Fluesse
Wald
Ortslage
Siedlung

Figure 3. The chosen stops appear in the small lists in the user interface. One of them has to be set to starting point.

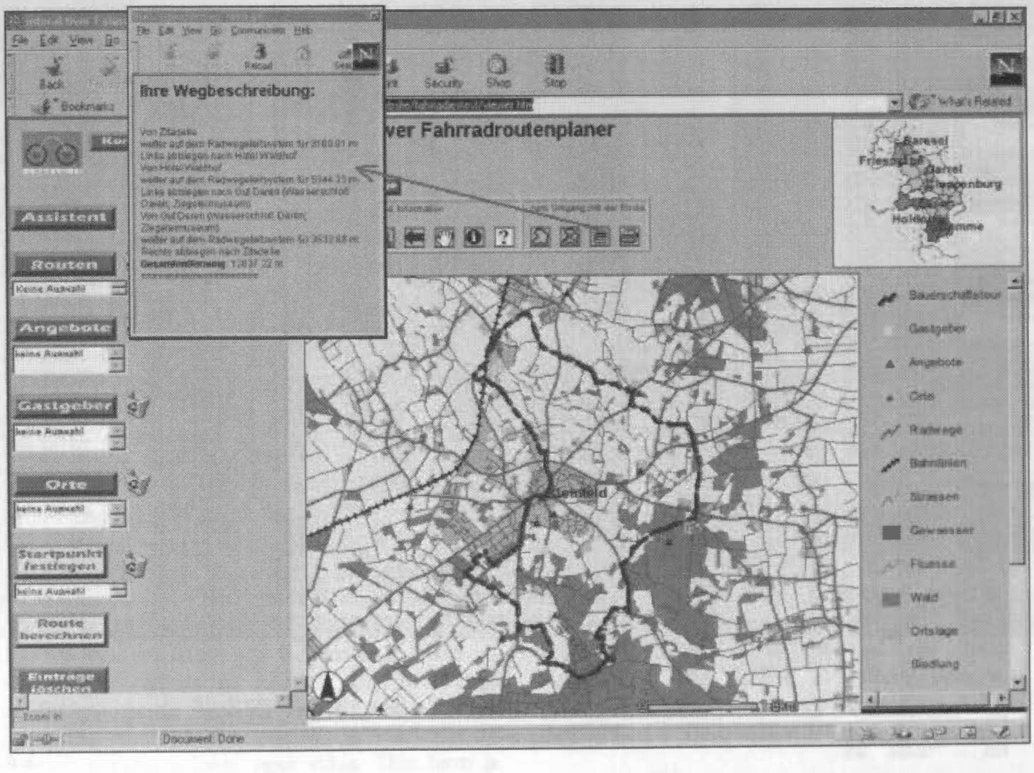

File Edit View Go Communicator Help
Reload
Security
Shop
Stop
Bookmarks
What's Related
Ihre Wegbeschreibung:
ver Fahrradroutenplaner
Assistent
Routen
Keine Auswahl
Angebote
keine Auswahl
Gastgeber
keine Auswahl
Orte
keine Auswahl
Startpunkt festlegen
keine Auswahl
Route berechnen
Einträge löschen
Document: Done
Barssel
Friesoythe
Garrel
Cloppenburg
Holdorf
Damme
Bauernschaftstour
Gastgeber
Angebote
Orte
Radwege
Bahnlinien
Strassen
Gewaesser
Fluesse
Wald
Ortslage
Siedlung

Figure 4. The individual bicycle route shown in the main map. A route description is available as well.

In addition, a print function for the main map exists. The user can choose between printing tiles of the map in the scale of 1:75000, printing the actual display or an overview map of the whole route.
Also, a function is provided for the user to get all stops along the route within a selected radius. It is possible to use these stops to customize a route as well.

4. FUTURE PERSPECTIVES

Currently, only one possible bicycle route can be calculated for the selected stops. The calculated route is always a roundtrip. However, more route functions are planned for the future.

A function will be implemented to let the user choose between a oneway and a roundtrip tour. Also, there will be a possibility to not only select the start but also the end point. Another function will let the user set the start and end point and choose one or more categories of sights, e.g. a recreational sight or a hotel. The system will calculate and suggest the possible routes and the user can select one of them. It is possible to extend this function by setting the duration in hours or the length of the tour in kilometres. For this an average speed has to be assumed.

In addition, it is planned to make the "Fahrradies" available for use with wireless devices. The user will be equipped with a GPS, a handheld computer with a built-in cellular phone for wireless internet access that make real time navigation and tour information possible. For many tourist areas, the necessary bicycle paths are already in digital format or can be easily annotated using analog bike maps. As a picture for the future the "Fahrradies" might be available as a "bikeheld navigator". With this, it resembles the ones for motorized vehicles which can be used as a wireless application and which display the chosen route on a handheld display depending on the actual location.

In conclusion, it would be desirable to extend the area covered by the "Fahrradies" application to the other parts of Germany, so that the bicycle routes do not stop at administrative boundaries. This means, however, that adjacent tourist regions have to work more closely together. This might be difficult due to competition between these regions.

5. REFERENCES

Fleming, J., 1998. Web Navigation – Designing the User Experience. O'Reilly. Beijing. Cambrigde. Cologne.

Jung, S., Schneeweiss, H., 2001. "Fahrradies" Radtouren im Internet. In: *arcaktuell*, Kranzberg, Germany Vol. II p. 27.

Kraak, M.-J., Brown, A.. [Eds.], 2001. Web Cartography – developments and prospects. Taylor & Francis Inc. London, New York.

Marshall, J., 2001. Developing Internet-Based GIS Applications. INDUS Corporation. Vienna (VA,USA). http://www.giscafe.com/TechPapers/Papers/paper058/.

Schneeweiss, H., 2001. Konzeption und Implementierung eines GIS-basierten Radroutenplaners für das World Wide Web (WWW) – Das "Fahrradies" für das Oldenburger Münsterland. Vechta. Diplomarbeit. Hochschule Vechta.

PHYSICAL MODELS OF GIS OBJECTS BY RAPID PROTOTYPING

Wolf-Dieter Rase

Federal Office for Building and Spatial Planning (BBR) , Deichmanns Aue 31-37, 53177 Bonn, Germany

rase@bbr.bund.de

KEY WORDS: cartography, visualization, GIS, model, planning, surface

ABSTRACT:

During the last years new technical solutions for the fast and inexpensive production of mechanical parts were developed. The prototypes are built by computer-controlled devices using a numerical description of the part. The prototypes serve mainly to judge on the form and appearance of the final parts. The available technical solutions for rapid prototyping were evaluated to build physical models of three-dimensional GIS objects. A recently introduced device, the 3D plotter Z406, is able to produce physical models in full color. A number of models were built using the 3D plotter, visualizing 3D choropleth maps and conceptual surfaces derived from demographic and economic data. The experiences and results in building 3D multicolored models of GIS objects are described, including the costs involved to build the models.

1. VIRTUAL REALITY AND PHYSICAL MODELS

Advanced visualization techniques have been used for some time in architecture and planning. The architect or planner is able to inspect buildings and landscapes which are not existent in reality (yet), but are stored as numerical models in a CAD or other information system. The technical possibilities stretch from photo-realistic images over stereoscopic views to walk-throughs or flights over a virtual landscape.

Despite the availability of virtual reality technology it is remarkable that physical models of buildings are still requested in architectural competitions. VR provides more visual information than the simple view on the building. For example the interior can be visualized under the varying light conditions and perspectives during a walk-through, or a view from inside through the windows can be generated. Why are the physical models still requested? The cost of model building is a marginal quantity in relation to the cost of the competition or even the real building. VR is still rather expensive when several persons have to view the building simultaneously. Probably architects have a special preference for hand-made objects. The haptic experience, the possibility to feel an object with their own hands, is certainly an important factor as well.

Physical models of three-dimensional GIS objects, however, are rarely realized. The missing tradition of model building is one explanation, the other is that the cost of building a model is not a negligible quantity in relation to the overall cost of spatial analysis. If the physical model of a building has advantages over VR in certain cases, it is possible that real models of GIS objects lead to a new quality in visualization, and subsequently to a better insight into the processes forming the environment.

During the last years new technical solutions were developed which enable the fast and inexpensive production of mechanical parts directly from a CAD system. The prototypes serve mainly to judge on the form and appearance, in certain cases to evaluate the mechanical function. Usually the prototypes are not exposed to the mechanical stress that their real counterparts have to withstand, for example the prototype of a steering wheel for a car. Thus it is not necessary in most cases to use the same material of which the final part will be made.

Rapid prototyping can also be applied to create physical models of three-dimensional GIS objects. In the following chapters the technical alternatives for rapid prototyping are described, with special attention to their application in cartography. Models of discrete and continuous surfaces derived from demographic and economic data serve as examples for rapid prototyping of GIS objects. The term *rapid prototyping* is also used in software technology, describing the fast and inexpensive development of computer programs.

2. TECHNICAL ALTERNATIVES

2.1 Main groups of rapid prototyping

The different ways to produce prototypes can be subsumed under three groups which have their equivalents in the arts:

- *Michelangelo*: A block of material is treated with a tool until the final form is reached. A block of marble is transformed into a sculpture.
- *Rodin*: The model is built by accumulating small amounts of material. A sculpture or a scaled model is aggregated from plaster or other plastic material.
- *Chillida*: The material is formed by strong forces, like a blacksmith transforms a rod of iron into a sickle.

An example of the Michelangelo group is a numerically controlled milling cutter. Although low-cost devices have been available for a few years, numerically controlled milling is usually not fast and mostly expensive. Thus milling is not applicable in most cases for the rapid prototyping of GIS objects. Methods of the Chillida group are rather rarely used or are secondary for the creation of GIS models.

2.2 Assembly from layers of material

The model is built from layers of material by a computer-controlled device. Principally only 2½D surfaces (only one z value at any point in the x-y plane) are possible because each layer needs a solid base. Concave parts must be reoriented in space to become a 2½D surface. If this is not possible, the part has to be assembled from convex subparts. Some prototyping methods solve the problem by providing a second material to fill hollow spaces. The material is removed after the model has been built. Some materials are also sturdy enough to bridge caverns.

Several alternatives have been developed for the assembly by layers of material (for detailed descriptions see Cooper 2001 or Dimov and Pham 2001):

- Paper. The layers are cut out in paper and assembled from bottom to top until the part is finished (laminated object manufacturing, LOM). The parts outside the model must be removed manually.
- Powder. The device deposits a thin layer of a powder, consisting of starch, plaster, photopolymer or metal. The parts of the layer belonging to the model are fixed by spraying an adhesive, by polymerization with UV light, or by local heating with a laser beam. When the first layer is fixed the second layer is deposited, and so on. The loose powder is removed. Certain devices allow the recycling of the unused material. The model can be infiltrated with a fluid after buildup to improve the mechanical and thermal stability.
- Thermoplastic material. The layers are deposited by a moveable print head similar to an inkjet print head. The material is melted in the head before deposition, or is hardened by UV light.
- Stereolithography. The energy of a moveable laser beam induces a phase change in a thin layer of a photosensitive fluid. The fluid becomes solid (polymerizes) where the beam hits the surface. A new layer is built on the solidified layer. The procedure is repeated until the last layer is finished.

The decision which alternative should be used depends on the purpose for which the prototype is intended. Stereolithography, for instance, results in very smooth surfaces, but is expensive. A combination of starch powder and wax can be used to build a model for investment casting. The model is burned by the hot molten metal with nearly no remainders. A part built of plaster powder infiltrated by cyanoacrylate can be finished like wood.

2.3 Color

The visual variable *color* is essential in cartography and for the visualization of GIS objects and maps. Until recently it was not possible to integrate the coloring into the assembling process using rapid prototyping. The raw models had to be colored manually, for example by using an airbrush. In 2001, ZCorporation introduced the 3D printer Z406 with integrated coloring. The device applies three different adhesives containing pigments in the subtractive base colors cyan, magenta and yellow to fuse the powder. The result is a multicolored part.

The Z406 and the new Z810 (with a larger build space) are until now the only devices on the market providing multicoloured parts. The capability to integrate color opens some new application fields for rapid prototyping besides cartography and GIS, such as multicolor models of molecules, buildings or plants, or the large field of medical visualization based on tomography data.

3. PRODUCTION OF THE MODELS

3.1 Outsourcing

Generally devices for rapid prototyping are rather expensive. For a small to midsize production volume the most economic way to build prototypes is outsourcing to a specialized firm (4Dconcepts in this case, http://www.4dconcepts.de/). The numerical description of the part is transmitted in a file, for example as an attachment to an e-mail. The part is assembled, infiltrated with a fluid, if required, and expedited to the customer by a parcel service. The Rapid Prototyping Homepage at the University of Utah contains numerous references, also for outsourcing (http://www.cc.utah.edu/~asn8200/rapid.html).

3.2 File format

To facilitate the transmission of the part description a general, non-proprietary file format should be used. The numerical representation in the system of the customer is converted into the commands and data formats that the specific prototyping device needs to build the model. In the STL format (the name is derived from its first application in stereolithography) the part, or better its surface, is described by a set of triangles (vertices and triangle normals in 3D). Two versions are in use, one with the triangle descriptions as plain text, the other in a binary format resulting in a smaller file size.

The STL file format is not able to transmit color information. The scene description language VRML (from *Virtual Reality Markup Language*) can be used instead to define the triangles and the associated color, either for each triangle or for the whole model or parts of the model. A description of the latest version of the VRML and related documents can been downloaded from the website http://www.vrml.org/.

The triangles in the model are defined by pointers to a set of vertices. The vertices are stored only once, whereas in the STL format the same vertices must be repeated for each triangle, resulting in larger files. The following VRML script defines a tetrahedron with different colors for each triangle.

```
#VRML V2.0 utf8
#Tetrahedron
Shape{
geometry IndexedFaceSet {
colorPerVertex FALSE
color Color { color[0 0 1,0 1 0,1 0 0,0 1 1]}
coord Coordinate { point[1 1 1,-1 1 -1,1 -1 -1,
-1 1 -1,-1 -1 1] }
coordIndex[2,1,0,-1,2,4,1,-1,4,2,0,-1,1,4,0]
colorIndex[0,1,2,3] } }
```

The VRML files of the examples in chapter 4 were produced by a software package for cartographic applications written by the author.

3.3 Limitations of the 3D plotter

The maximal build space for parts produced by the 3D plotter Z406 is 20 by 25 by 20 cm. The thickness of each layer can be varied from 0.076 to 0.254 cm, dependent on the powder material used and the resolution required in the z direction. The build time is proportional to the number of layers. In color mode two layers per minute can be produced, in monochrome mode six layers. The powder consists of either starch or plaster. It is recommend to infiltrate the raw model to enhance the mechanical and thermal stability, and to lower the sensitivity against humidity. The infiltrant can be wax (in combination with starch), or different resins based on cyanoacrylate. After infiltration with the latter the model can be sanded like wood, colored or electroplated to imitate the appearance of metal. The newer model Z810 uses the same technical principles as the Z406, but has a larger build volume of 50 by 60 by 40 cm.

The software of the 3D plotter accepts only triangles (VRML node IndexedFaceSet). No other geometric objects possible in VRML (nodes) are recognized, such as extrusion bodies, text strings or heightfields. Each object must be represented by a set of triangles, even lines, which can inflate the number of triangles and the size of the file considerably. The need for triangles and the resolution of the 3D plotter restricts the use of text objects. The letter size must be large enough to be recognized, which usually requires too much space on the model. The rear side of the model can be used to imprint, for example, a logo or a copyright notice.

4. EXAMPLES

4.1 3D choropleth map of land prices in Germany

Several models of 3D choropleth maps and conceptual surfaces were built using the 3D plotter Z406. The model in fig. 1 is a three-dimensional choropleth map visualizing the average cost of land in the counties and cities of Germany in the year 2000 (BBR 2001). The height of the prisms is proportional to the average price of land in Euro per m². In contrast to a 2D choropleth map depicting value classes, the absolute differences in height can be perceived immediately. To facilitate the comparison the tops and the walls of the prisms bear the colors associated with the value classes. The phone card (credit card size) in the photograph serves as a standard of comparison.

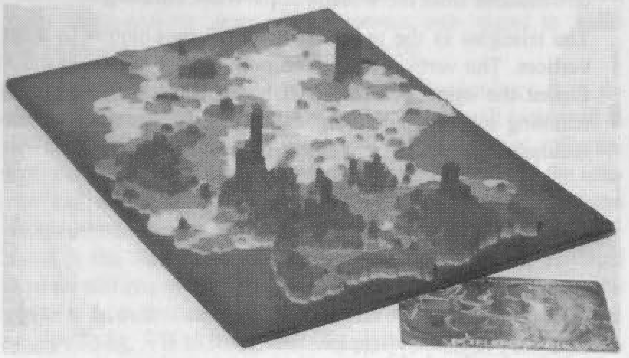


Fig. 1 3D choropleth map representing the price of land per km² in the year 2000

The powder material is plaster. The layers have a thickness of 0.01 cm. The production in the 3D printer took approximately 3.5 hours. After assembly the model was infiltrated with epoxy resin. The overall production cost including preparation of the files, infiltration and shipment were about 400 Euro.

4.2 Continuous surface of land prices

Fig. 2 depicts a continuous surface derived from the same data using the pycnophylactic interpolation procedure. A smooth surface is interpolated from polygon-based data ensuring the conservation of volume above the polygon (Tobler 1979, Rase 2001). The perception of height is facilitated by different colors representing the range between two adjacent contour levels. The state boundaries in Germany, the boundaries of the neighboring countries and the larger water surfaces (North Sea, Baltic Sea, Lake Constance) were added to provide better topographical cues for users less familiar with the locations.

The model was infiltrated with cyanoacrylate, which results in an enhanced mechanical stability. Due to the more expensive fluid the overall cost was slightly higher than for the previous model. At a closer look some dark spots in the low areas around Berlin can be identified. At these locations the distance between the surface (yellow) and the base (dark blue) did not suffice to deposit a layer of yellow material thick enough to cover the blue underground completely, as intended.

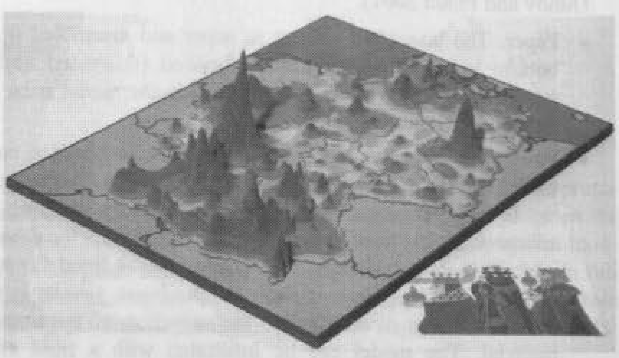


Fig. 2 Continuous surface interpolated from the data in fig. 1 by the pycnophylactic interpolation method

4.3 4D choropleth map with population data

Fig. 3 is a photograph of a 3D choropleth map representing the population in Germany. Units of reference are the so-called "planning regions" used for spatial monitoring in Germany. The volume of the prisms is proportional to the number of inhabitants in each unit. The height of the prisms is proportional to the population density expressed as inhabitants per km². To achieve a better perception, the colors associated with the value (height) classes and the class boundaries are depicted on the prism walls. The differences in population and density between the more rural regions and the densely populated cities of Berlin, Hamburg and Bremen are clearly visible. On the other hand the comparison of the less populated areas is more difficult due to the smaller differences in height.

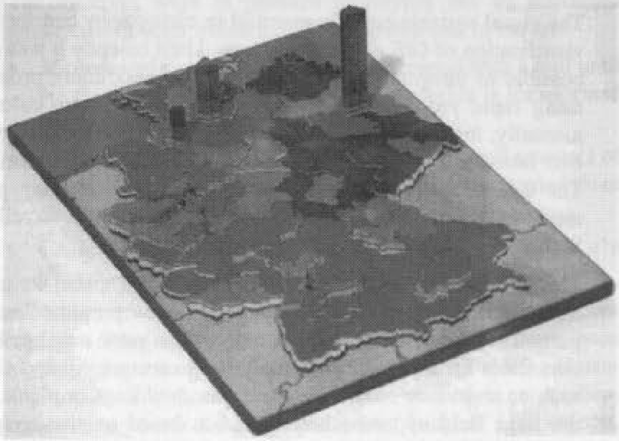


Fig. 3 3D choropleth map representing inhabitants and population density

A second variable, the change in population between 1990 and 2000, is represented by the color of the prism caps. Blue caps express two classes of population loss, red caps two classes of population gain. The 4D map – the two dimensions of the plane, the height proportional to the population density, and the cap colors representing classes of change rate – transmits additional information, but is more difficult to conceive than a 3D map with redundant visual variables. Theoretically additional visual variables could be used to encode even more dimensions, but the limits of usability seem to be reached with a 4D map.

4.4 Continuous surface of population density

The continuous surface in fig. 4 was interpolated from the population values and the polygons of the cities and counties, again with the pycnophylactic interpolation procedure. The black isolines separate the colored regions representing height classes. The state boundaries (green) are intended to provide topographical cues. The peculiar lines at the slope of the "mountain" in the east (left) are the boundary of the city of Berlin, projected onto the slope (also visible in fig. 2).

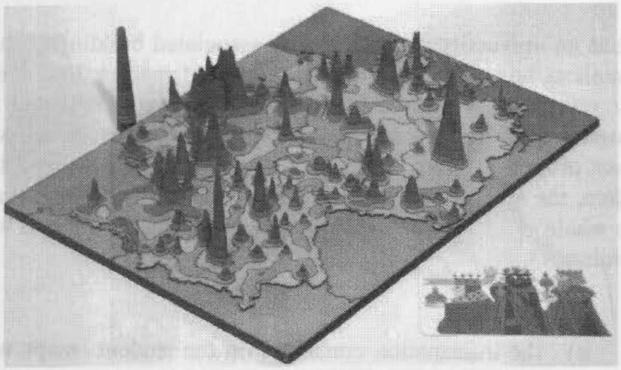


Fig. 4 Smooth surface depicting population density based on county data. The moveable cone at the left serves as a legend to estimate the height.

5. PROS AND CONS OF REAL MODELS

The few real models of GIS objects produced so far have revealed the technical possibilities and limitations in hardware and software, the approximate production costs, and the benefits and problems of outsourcing the production. But what are the advantages of physical models in comparison to conventional maps, perspective drawings, or VR techniques?

On "normal" maps the third dimension is encoded by visual or graphic variables (size, lightness, color, texture, shape, orientation). The reader of the map must be able to decode the visual variables to achieve a mental reconstruction of the third dimension, assisted by a legend. Some of the recipients have problems to decode the visual variables of maps because they do not have the necessary training or experience. In this case the message in the map fails to reach the addressee.

Humans have a life-long training to interpret 3D objects or pictures thereof and to construct a mental model in 3D (which, of course, can go wrong in certain cases, called optical illusions). Especially for decision-makers with limited experience in map reading the real 3D model transmits the cartographic message much better and faster than a two-dimensional map.

The 3D effect can be achieved by a perspective drawing of the object as well. Physical models have the advantage over 2D drawings that slight movements of the head or body suffice to look at parts of the model covered by obstacles in the line of sight, for example high prisms or mountains in the foreground. In this case a second or third drawing must be provided to present all parts of the surface.

Humans are able to estimate distances in a 3D scene by stereopsis and minor changes of the eye point due to lifelong training and experience in evaluating 3D scenes. Thus it is easier to compare the heights in a model than in a drawing. Optical irritations and illusions which might occur in perspective drawings are solved for real models by a slight change of the viewpoint.

Probably the most important disadvantage of the physical models shown here is the lack of textual information necessary for legends and topography. The 3D maps shown here are not self-explanatory. To understand the content of the models either a verbal description or additional legends on paper are required.

The nearest competitors for real models are the techniques of virtual reality (VR). As mentioned in the introduction with respect to buildings, VR can provide functions that are not possible with real models, such as immediate response to interactions of the user, visualization of the time dimension in animation sequences etc. Stereoscopic images can be generated with the help of shutter glasses, monitors with lenticular lenses, or head-mounted displays, for example. In situations where several persons have to look at a 3D scene simultaneously VR is still quite expensive. It depends on the specific purpose and situation whether VR or physical models are the appropriate solution in terms of perception and economics.

6. CONCLUSION

The first experiences concerning the reception of the models by researchers in my agency were very encouraging. And surprising: nearly everybody tried to touch the models with his hands, *to get a feeling* in the literal sense. The use of the haptic channel, to touch and feel something, is a basic sensual capability. Although it is less important than the optical and acoustical senses, it has been necessary for the development of mankind, to invent and use fine tools, for example. Besides the possibility to provide maps for blind and visually impaired persons, the haptic experience could be utilized as an additional stimulus for transmitting a cartographic message and to induce insight.

REFERENCES

BBR, 2001. Markets for building land and real estate 2000 (in German; Bauland- und Immobilien-Märkte, Ausgabe 2000). Berichte Band 9, Bundesamt für Bauwesen und Raumordnung, Bonn, Germany. http://www.bbr.bund.de/

Cooper, Kenneth G., 2001. Rapid prototyping technology. Marcel Dekker

Dimov, S. S., Pham, D. T., (2001). Rapid Manufacturing: The Technologies and Applications of Rapid Prototyping and Rapid Tooling. Springer

Rase, Wolf-Dieter, 2001. Volume-preserving interpolation of a smooth surface from polygon-related data. Journal of Geographical Systems (2001) 3, pp 199-213

Tobler, Waldo R., 1979. Smooth pycnophylactic interpolation for geographical regions. Journal of the American Statistical Association, Vol. 74, No. 357, pp 519-535

TRIDIMENSIONAL MODEL FOR A UNIVERSITY CAMPUS USING DTM

Cláudio Márcio Penna de Moraes
Luiz Alberto Vieira Dias

Paraiba Valley University - UNIVAP
São José dos Campos, SP, Brazil
mclaudio@univap.br and vdias@univap.br

Commission IV, WG IV/6

KEY WORDS: DTM, Tridimensional model, Computer Graphics, ArcView, Environment

ABSTRACT

This work presents a method for viewing in three dimensions an university campus and its associated buildings, when the cartographic data was not completely available. The authors had to rely on old and outdated maps to prepare a Digital Terrain Model of the campus. Another difficulty was that the data was divided into seven maps and an adjustment had to be made on the borders of them. Based on the recovered data a Delaunay Triangulation was performed. However, the model so acquired did not represent properly the terrain features, because of the shape of the triangles in the automatic triangulation. To solve this problem, the authors did generate a new isoline, with an altitude one meter below the minimum of the region, circulating the whole campus area. Once these models were created, it was possible to navigate and fly over the campus terrain thus obtained.

1. INTRODUCTION

Paraiba Valley University, UNIVAP is a young and fast growing university, with about ten thousand students on the undergraduate and graduate levels. When UNIVAP that, at the time, was located only at downtown São José dos Campos, São Paulo State, Brazil decided to expand and build another campus, the Urbanova Campus, on the mentioned city suburbs, the selected area was on a hilly terrain with no detailed geographical maps available.

2. DISCUSSION

The existing area maps were old and outdated, with contour lines spaced by 20 m. This was the cartographic material on hand when the construction of the new campus was begun, in 1993. In 1998, the Cartography Course students, under the guidance of Prof. Vicente Maia (MAIA, 1998), divided the area into seven sections, and carefully produced a map of each section of the area under consideration, not geo-referenced (about 4 million square meters), but with contour lines 1 m apart.

It was a huge progress, but what was needed was one digital map of the whole area. This task was then passed on to the Geo-processing Laboratory of the Institute for Research and Development, IP&D, Paraíba Valley University, UNIVAP at Urbanova Campus. In order to fulfill this task the work that is described in this paper was performed. It consisted in:

a) based on these seven student conventional maps, it was prepared a new digital map, from now on to be called MAP1, integrating all the seven previous student maps;

b) the information contained on the student maps was geo-referenced into UTM coordinates;

c) from MAP1, it was generated a DTM of the region, using ESRI's ArcView 3.0;

d) d) Finally a 3D visualization of the resulting DTM was prepared.

In order to prepare the new integrated map, MAP1, it was necessary, using a GPS, to identify three known terrain points, to geo-reference all the points on the student's original seven maps.

The next step was to assure the continuity of the isolines at the borders of the adjoining maps. The altitude correction for each isoline joint was then performed.

After that a first terrain model was generated (TIN = triangular irregular network), using Delaunay triangulation (MILITÃO,1996). However, the model so acquired did not represent properly the terrain features, because the triangles in the automatic triangulation program ArcView 3.0 (BAROS & MORAIN, 1996) produces lines, to join triangle vertices, that lay outside the real terrain. In order to solve this problem, the authors did generate a new isoline, with an altitude one meter below the minimum of the region, circulating the whole campus area. The level of this isoline, in this case, was 544 meters, since the minimum level of the campus was 545 meters. The creation of this new line, one meter below did not caused the generation of new features.

Figure 1 - Three-dimensional isolines

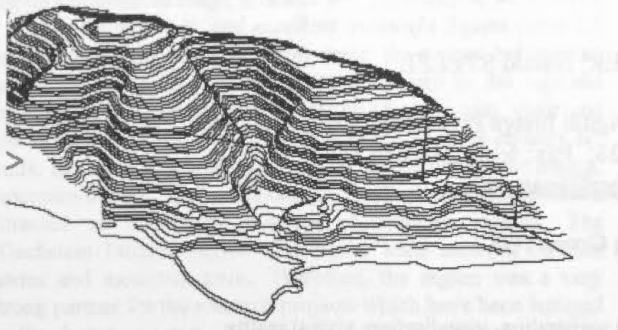
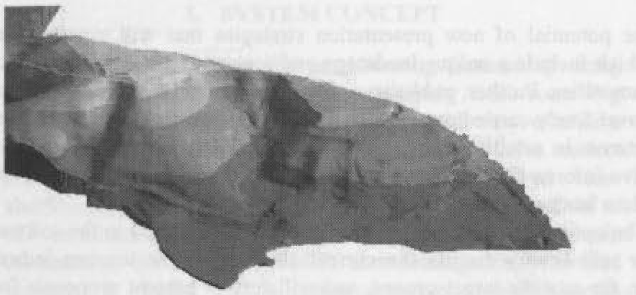


Figure 2 - Rendered DTM



3. RESULTS

Partial results are shown in Figures 1 and 2. Figure 1 presents part of the three-dimensional isolines obtained, while Figure 2 depicts the rendered digital terrain model (DTM) for the same part of the region under study. In both figures it is possible to calculate areas, volumes and distances. Figure 2 is drawn in such a way that different altitudes are shown in different colors (from a maximum of 630 meters, depicted in yellow, to minimum of 545, depicted in brown).

A new isoline at an altitude of 544 meters encircling the whole area under study, was artificially created, in order that the all the Delaunay triangles (MILITÃO, 1996) would fall within the real terrain. Once these models were created, it was possible to navigate and fly over the campus terrain thus obtained.

The campus DTM so created was then used with the VRML language (ENVI, 1997), and visualized on a internet browser. With the availability of the DTM it was possible to simulate the location of new buildings on the campus, navigate between the already constructed architectural structures, and assess its environmental impact.

4. CONCLUSIONS

Since 35% of the campus is located on an environment preservation area, defined by a specific law, the University has to follow strict construction and landscape limiting rules. Every new construction, earth move, or even the cutting of a single tree has to be authorized by the City Council on Environment, and consequentially, has to be carefully planned in order to reduce to a minimum the impact on the ambient. The DTM, as well as the digital MAP1, has been in use by the Civil Engineering Section of the University, on its planning of the campus expansion. The geographical region centered around São José dos Campos, São Paulo State, Brazil, has about one million inhabitants and is heavily industrialized, including chemical, pharmaceutical, automobile and aircraft industries, so the environment is a major concern of the local government.

This work is far from complete, since new studies on interpolation and three-dimensional plotting are under way, and will use as model the structures on the Urbanova Campus. Especially the plotting of the buildings on the terrain, and how to do it integrating DTM and Computer Graphics (FOLEY et al., 1992).

REFERENCES

Baros, S. V. L.; Morain, S. (1996) "Raster Imagery in Geographic Information Systems". Santa Fe, NM, USA, OnWord Press.

ENVI (1997) "Tutorials". Lafayette, Co, USA: Better Solutions Consulting (BSC).

Foley, J. D.; VanDam, A.; Feiner, S. K.; Hughes, J. F. (1992) "Computer Graphics - Principle and Practice". 2nd Edition. Addison-Wesley.

Maia, V.P.S. (1998) Private Communication.

Militão, J.G.O. (1996) "Uma metodologia para o posicionamento de prédios em terrenos acidentados ("A Methodology for Building Positioning in Hilly Terrain"). Master's Thesis. Catholic University, Rio de Janeiro, RJ, Brazil.

MULTIMEDIA VISUALISATION OF GEOINFORMATION FOR TOURISM REGIONS BASED ON REMOTE SENSING DATA

Alexander ALMER, Harald STELZL

Joanneum Research, Institute of Digital Image Processing, Graz, Austria
Phone: ++43 316 876 1738; Fax: ++43 316 876 1720
alexander.almer@joanneum.at

Working Group IV/6

KEY WORDS: remote sensing, multimedia, data fusion, image registration, visualisation, virtual reality

ABSTRACT:

The tourism and leisure industries are currently exploring the potential of new presentation strategies that will maximise the promotional appeal of tourism regions and their main assets, which include a unique landscape and scenery, an impressive cultural heritage, and a wide range of sports facilities and other tourist amenities. Further, publishers of tour guides and maps similarly have expressed a strong interest in the new technologies. The EU Travel Study carried out by Sloggett (1997) revealed that major travel and tourist companies and organisations take an increasing interest in satellite technology, and in the development of a Travel, Leisure and Tourism Information Service. The kind of innovative Information Service they envisage will provide information on a resort and the region, entertainment possibilities and will also allow business transactions.

Online and offline data visualisation for tourism topics have to be optimised for the different output devices related to the software, the hardware requirements and the display design. This paper will briefly discuss the current situation in the tourism industry regarding currently available modes of information presentation for specific target groups, and will further present proposals for a novel interactive multimedia information system focusing on the various components of such a system. The system will integrate satellite images, GPS data, GIS information and other information on leisure activities and tourist facilities. Aerial and satellite images will be used to implement the visualisation of a tourist region for multimedia CD-ROMs, display on the Internet and also on mobile devices.

1. INTRODUCTION

There are many different ways of presenting information for tourists. In the past, tourist regions largely relied on analogue media. Over the last few years, tourist offices and hotels have increasingly started to use the Internet, especially for online booking services. However, most websites are still largely modelled on the analogue products which were used for advertisements and brochures. Only a small number of companies within the travel, tourism and leisure industries make full use of the varied range of Internet services that are available today. In the travel and tourist industries it is important that potential customers can quickly find all the information they need. Everywhere there is strong competition between the different tourist regions, especially in the Alps. Therefore, the tourist boards are increasingly looking for new ways of presenting their main assets drawing on the full range of analogue and digital media. One of the main attractions that tourist regions in Alpine areas have to offer is their often spectacular landscape. Logically, therefore, tourist information about these regions will use digital media like the Internet, CD-ROMs, DVD, mobiles devices, video and TV, and include 3D visualisation techniques based on remote sensing data for maximum effect.

Electronic media can store not only more but also more accurate information, and they permit interactivity, which printed media are unable to provide. 2D and 3D maps and virtual flights on Internet and CD-Rom are seen by the tourism boards as an opportunity to upgrade and expand their business and also enhance their image.

The fast developments within information technologies and communication have in common a huge impact on product deliveries and consumption.

Today and increasingly more in the future mobile devices such as cellular phones, personal digital assistants (PDAs), palmtops and GPS devices are used to get mobile access to information. That is why in this concept mobile devices play a crucial role. PDAs in connection with GPS function allow to relate the information to the user's present position ("Location Based Services") so that unlike regular Internet users he is offered a geographically interesting piece of information via a mobile device (Fritsch, 2001).

Information systems are generally created for people who need directions that can guide them to the desired destination. Therefore, the most important customer groups for Alpine tourism comprise mountain bikers, hikers, high mountain skiers, and cross country skiers. Tourist information systems have to offer a broad range of orientation and information tools.

In order to guarantee a comprehensive planning of holiday activities 3D maps, virtual flights, guide books, height profiles, photographs, etc. are provided on the Internet for hiking or biking tours and it is also possible to download the route to a GPS device or print out a description of the desired tour.

2. REGION

In the framework of two research projects a close co-operation with the "Dachstein-Tauern" tourism region was suggested. (Almer et al., 2000). This region includes such well-known skiing resorts as Schladming (host to world skiing championship events) and Ramsau (world championship in nordic skiing 1999) and is one of the most attractive tourist regions in Austria (see http://www.dachstein-tauern.at). The region's impressive range of hotels and guesthouses, which offer a total of 22,300 beds, and excellent overnight figures (over 2.2 million overnight stays in 1998) make the tourist industry an important factor which contributes massively to the regional economy. The region boasts a total of 88 cable cars, chair and tow lifts, 170 km of ski runs, and 300 km of cross country ski trails. In recent years, the area's many opportunities for hiking, mountain-biking and relaxation (wellness centres) have attracted an increasing number of summer visitors. The "Dachstein-Tauern" region also has a wide network of bike routes and mountain trails. Therefore, the region was a very strong partner for the research projects which have been realised in the last two years. The results of these projects will be presented in the framework of this paper.

3. SYSTEM CONCEPT

In Austria as well as in the neighbouring German speaking regions (Bavaria, Switzerland) there already exist some big providers of touristic information systems with very extensive data bases. All systems tested within the scope of a market analysis (Almer et al., 2002), however, showed clear shortcomings concerning visualisation, above all in the field of 3D visualisation, as well as in the storing of spatial and thematic information. So as to avoid such shortcomings the presented concept is based on a comprehensive geo-multimedia information pool, out of which end-user applications can be provided on different output media such as world wide web, CD-ROM, info terminals and mobile terminals. In spite of the different technical possibilities and restrictions the fundamental navigation elements and possibilities are offered on all media by means of which a far reaching customisation of interactive contents is made available. The basic elements for the complete system pursued in this project are shown in the following figure.

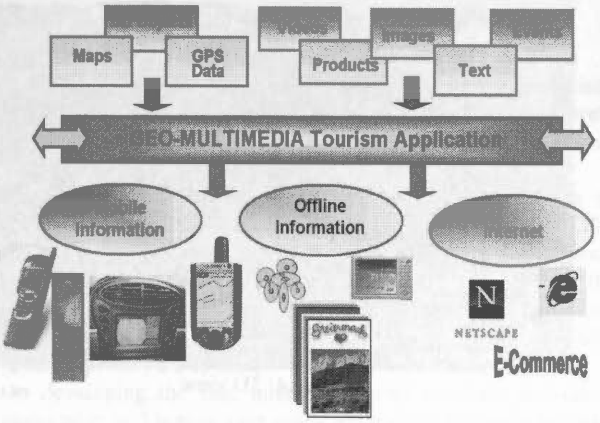

Maps
GPS Data
Videos
Images
Events
Products
Text
GEO-MULTIMEDIA Tourism Application
Mobile Information
Offline Information
Internet
N
NETSCAPE
E-Commerce

Figure 1: Product Definition

The development of a thematically and geographically related data base is integrated into an interactive multimedia information system and allows an optimised online and offline presentation of tourist information on a variety of digital media. The system will integrate satellite images, GPS data, GIS information, terrestrial images and will also include tourist information and offer multimedia, 2D and 3D visualisation technologies and interactive navigation. Tourism information is basically location based information and provides information on a tourism region on the Internet, on CD-ROMs as well as on mobile devices and is therefore of high interest for tourists. When developing an interactive multimedia information system for a tourism region, work has to proceed in several stages. For economic reasons the whole system has to be planned as a long-term concept.

4. PRODUCT DEFINITION

Multimedial presentation of tourist information signifies improved conveying of information by means of a well-aimed combination of different media, like text, sound, picture, video and animation. This presentation must not be restricted to one or two output media, but should work with every sort of existing technology like CD-ROM, Internet and mobile devices. Tourist information is modelled on the topographic conditions, so the application of remote sensing data for landscape visualisation is particularly appropriate for this purpose. The combination of digital elevation models (DEMs) with aerial and satellite images is the basis for the development of 3D views, virtual flights and panoramas. When combining this with tourist information in the form of texts, photos and videos a comprehensive information system emerges, which offers the tourist optimum access to the region of his choice. This access is marked by a high degree of interactivity, regarding the realization of multimedia CD-ROMs as well as Internet solutions.

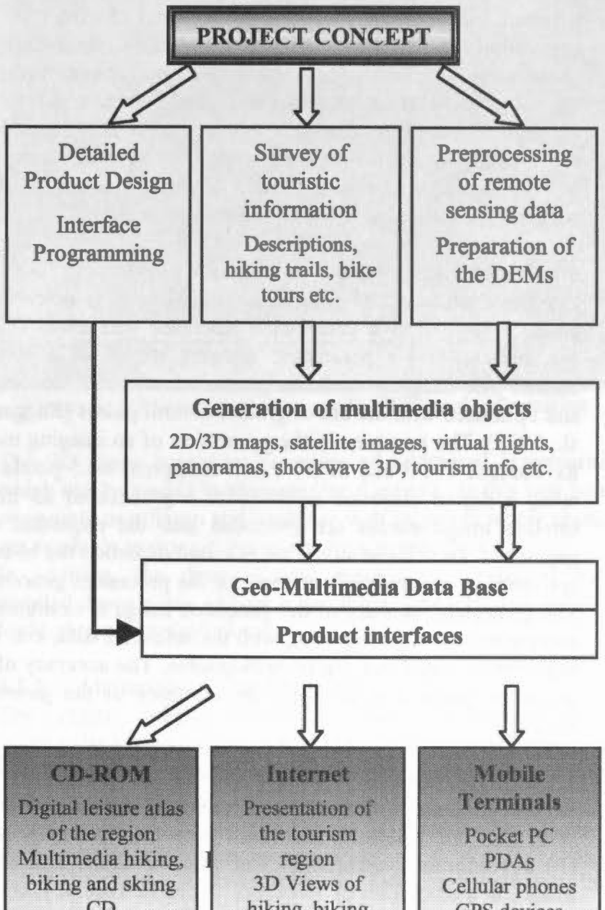

PROJECT CONCEPT
Detailed Product Design
Interface Programming
Survey of touristic information
Descriptions, hiking trails, bike tours etc.
Preprocessing of remote sensing data
Preparation of the DEMs
Generation of multimedia objects
2D/3D maps, satellite images, virtual flights, panoramas, shockwave 3D, tourism info etc.
Geo-Multimedia Data Base
Product interfaces
CD-ROM
Digital leisure atlas of the region
Multimedia hiking, biking and skiing CD
Internet
Presentation of the tourism region
3D Views of hiking, biking and skiing tours
Mobile Terminals
Pocket PC
PDAs
Cellular phones
GPS devices

Figure 2 gives an overview of the product development for a comprehensive multimedia presentation of a tourism region. The following chapter treats the particular products in detail except the implementation of the geo-multimedia data base.

5. MULTIMEDIA PRESENTATION OF LEISURE & TOURISM INFORMATION

5.1 Preprocessing of remote sensing data

For the visualisation of a tourism region three-dimensional landscape models are very suitable, as they give the user a particularly realistic impression. In order to design such 3D models as realistically as possible, satellite image data are used in the scope of the project. For the scale required within the framework of the project the resolutions of the sensors of Landsat TM (30 m multispectral) and SPOT PAN (10 m panchromatic) are most suitable.

5.1.1 Preparation of the DEMs

Digital elevation models are not only required for the geocoding of remote sensing data, as described in chapter 5.1.2., they also play a vital role in 3D visualisation. They are the core piece of every 3D model, regardless if 3D views, virtual flights or panoramas are calculated. The required gridsize of the DEMs depend on the desired scale of the depiction. A digital elevation model is a gray scale image which is assigned a determined resolution and a determined map projection with geodetic datum. For geocoding as well as for 3D visualisation the DEMs are to be converted into the required reference systems and the required resolution. This transformation consists of the conversion of a pixel's rectangular coordinates into geographic coordinates, with reference to the respective ellipsoid, further of the datum transformation into the desired reference datum and the subsequent projection into the requested new rectangular coordinate system. For this conversion the bilinear resampling is very useful, since it possesses a slight smoothing effect whereby the formation of edges is avoided.

5.1.2 Geocoding of satellite image data

For the application of remote sensing data it is essential to geocode them into a determined reference geometry. This is accomplished by a parametric imaging model of a satellite sensor. The imaging model is generated with the sensor data and optimised with the aid of ground control points (Raggam et al., 1999). The accuracy in the generation of an imaging model for Landsat TM 5 and SPOT PAN is approx. ±0.5 pixels. On using a digital elevation model with a gridsize of 25 m the satellite image scenes are geocoded into the requested map geometry. To achieve an optimum visual depiction the bi-cubic convolution resampling is applied for the process of geocoding. The geometric accuracy of the geocoded image is examined by comparing and superimposing with the reference data, i. e. with topographic maps and digital orthophotos. The accuracy of the geocoded image has to equal the accuracy of the generated model.

5.1.3 Data fusion

In the context of the works a Landsat TM scene (30 m resolution multispectral) was fused with a SPOT PAN scene (10 m resolution panchromatic) in order to obtain true colour images at a resolution of 10 m (see figure 3). Precise geocoded remote sensing data are a precondition for such a data fusion, so as to receive an accurate fusion result.

The applied method is based on the principal component transformation and supplies high resolution true colour images.

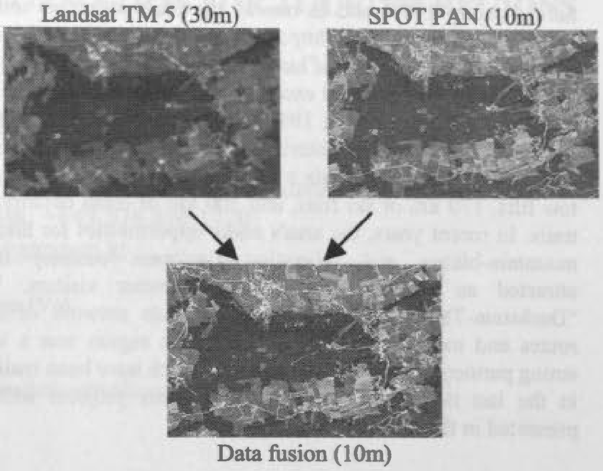

Landsat TM 5 (30m)
SPOT PAN (10m)
Data fusion (10m)

Figure 3: Data fusion

5.2 Generation of multimedia objects

5.2.1 Generation of 3D views

3D views are calculated from the combination of the fused 10 m RGB images and the DEMs. A precondition for this data generation is the exact geometric concordance of geocoded satellite images and the DEMs (Patterson, 1999).

The predefinition of the contemplator's position results from the definition of the image acquisition parameters, i. e. camera position and view direction. These parameters may be modified interactively in order to get the best visual 3D view of a landscape. After predefining the parameters the 3D views are rendered with a defined resolution (see figure 4). For visualisation on CD-ROM the resolution of e. g. 1200 x 800 pixels proves useful (Questar Productions, 1998).

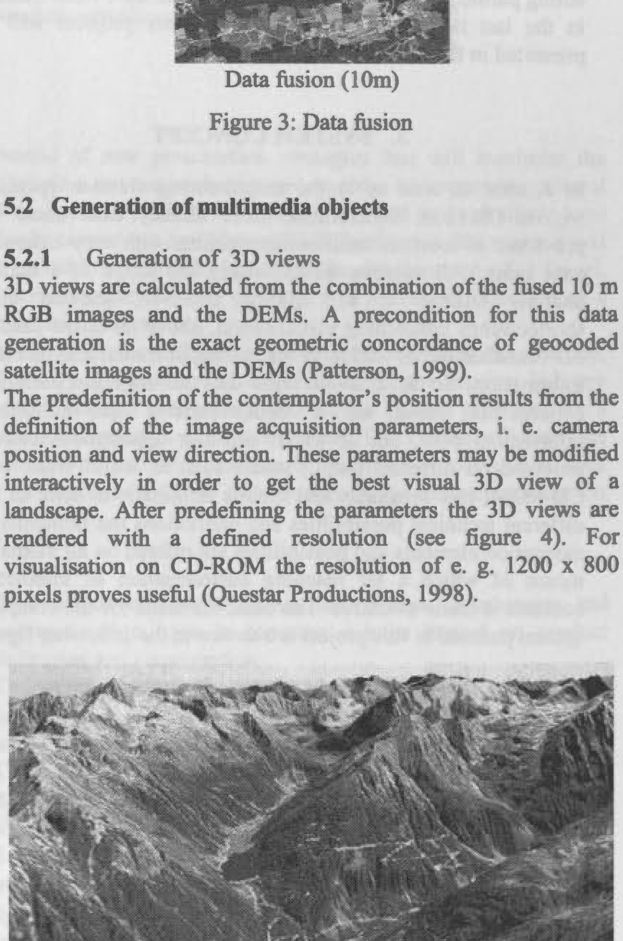


Figure 4: 3D view

The generation of 3D views as well as the following description of flights and virtual panoramas are treated in the animation software "World Construction Set". This software is highly suitable for the animation of geo data, in particular for the procession of large DEMs and satellite image textures.

5.2.2 Virtual flights

For the generation of a virtual flight some thousand seperate images are rendered and combined to an animation file (mpg-file), depending on the projected flight duration. For a flight animation it is vital to define a flight path and to determine the variable view direction along this flight path and the equally variable flight height. Starting from a standardised frame rate of 30 images per second 1800 separate images have to be calculated for a one minute animation. In order to avoid flickering in an animation, antialiasing-techniques are applied. For the generation of an animation file separate images are coded into a compressed format. For this, the mpeg-1-format has proved useful, because it possesses a high compression degree and yet a high quality.

5.2.3 Virtual panoramas

The generation of 180° or 360° panoramas offers an impressive possibility to visualise a region. These panoramas can be generated both from separate images and from virtual 3D views of satellite images (see figure 5). Thus, the user is able to navigate and zoom the panoramas as well as to use links to other panoramas and to receive further touristically relevant information.

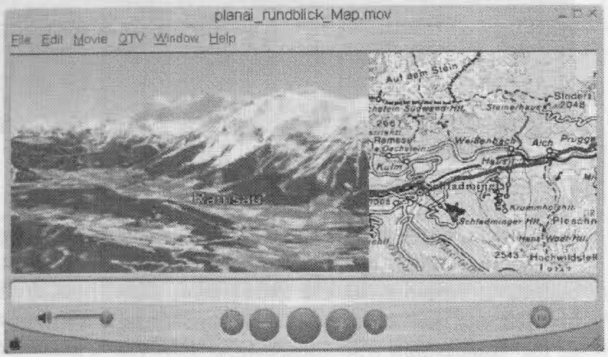


Figure 5: Virtual panorama with map

The combination of different panoramas with maps permits the realisation of a "virtual stroll" through the region. The user may move interactively through villages and towns, enter buildings, such as hotels and shops, and on the general map he may always see his position and perspective.

5.3 Design and programming

The integration of the above described multimedia objects into different applications is realised by the definition of concrete projects like multimedia CD-ROMs, Internet performances or applications on mobile terminals.

The design for the user interfaces is realised by means of graphic programmes such as Photoshop or Coreldraw. The programming of the multimedia surroundings is performed using Director, Shockwave, Flash and other products of the software company Macromedia. Further, data base solutions (Microsoft Access, SQL etc.) and data base interfaces (DataGrip, etc.) are applied.

On developing the user interface special attention is paid to uncomplex and logical user guidance and the system's software and hardware requirements are adjusted to the target group.

5.4 Data integration

In the framework of the visualisation of tourist information large data from existing archives and newly surveyed data flow in. For multimedia hiking or biking guides these are data of tours (GPS routes), photos and descriptions, as well as general tourist information. For the visualisation of tourism regions the whole tourist infrastructure (hotels, info points etc.) plays a crucial role. Partly, these data are already available within existing Internet solutions and need to be enhanced and completed.

The above described multimedia objects emerge as novel sources of information, visualising outstandingly a region's topographic conditions. Above all the high degree of interactivity, which distinguishes the products described below permits the user a completely new access to touristic and topographic data.

5.5 Products

5.5.1 Multimedia mountainbike guide

The bike CD-ROM "Salzburg-Steiermark" ("Salzburg-Styria") is a multimedia mountainbike guide, which describes 42 bike tours in the regions of "Dachstein-Tauern", "Lungau" and "Sportwelt Amadé", using 3D views and 3D animations. For each tour the CD comprises exact descriptions and road books with all waypoints, height profiles and photos. The tour description allows the user to select bike tours according to individual demands and convenience. It is possible to enter fitness and technical parameters for an optimum choice. 3D views and virtual overflights show the tours in the most impressive way giving a true impression of the terrain's features and the changes in elevation along the bike tour.

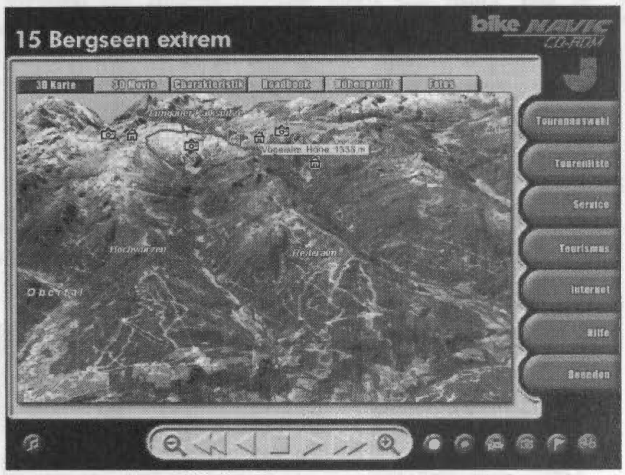


Figure 6: 3D tour visualisation with additional information

The 3D maps comprise a variety of additional information which can be loaded by the user (see figure 6). In this way huts, waypoints or villages and towns as well as points where photos are available appear on the screen by a mouse click.

In addition, the three tourism regions are presented with detailed descriptions of how to get there, where to stay, cultural highlights, sports facilities etc.

5.5.2 Multimedia hiking CD-ROM

The digital hiking CD-ROM "Dachstein-Tauern" is a multimedia hiking and tourism guide. It offers the description of 43 hiking tours in the "Dachstein-Tauern" region with detailed texts, maps and unique photos, completed with multimedia objects such as virtual overflights and 3D maps for each tour.

Each tour is visualised by means of a 30 to 120 seconds overflight. During the overflight the user may manipulate the visual display of the film. These overflights give the user a particularly realistic and exciting impression of the hiking tours (see figure 7).

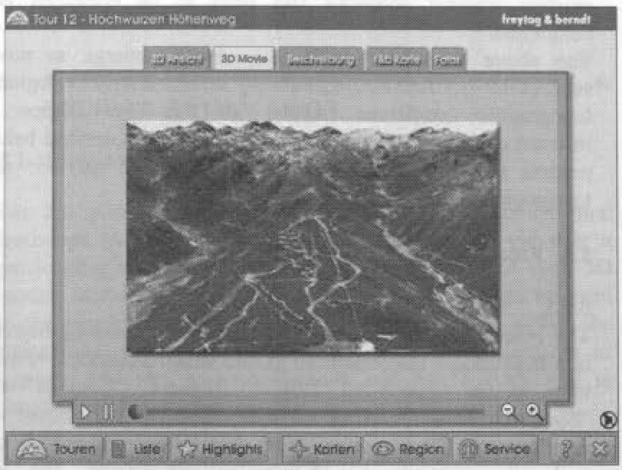


Figure 7: Virtual flight over a hiking tour

The structure of the hiking tours is adjusted to the requirements of different target groups. The CD offers many various kinds of tours ranging from extreme hiking tours to adventure trails for children, in addition comprehensive information on history, myths, legends, geology and pasture economy. The combination of text, video and audio material permits the observer to contemplate geography and history of the mountains with new eyes. In order to offer a navigation aid for mountaineers all the hiking routes are also provided to the Garmin GPS devices, the user being able to select between guided routes and detailed tracklogs.

5.5.3 Global presentation of the tourism region

The global presentation of the tourism region on Internet and CD-ROM is a multimedia tourism guide for the "Dachstein-Tauern" region.

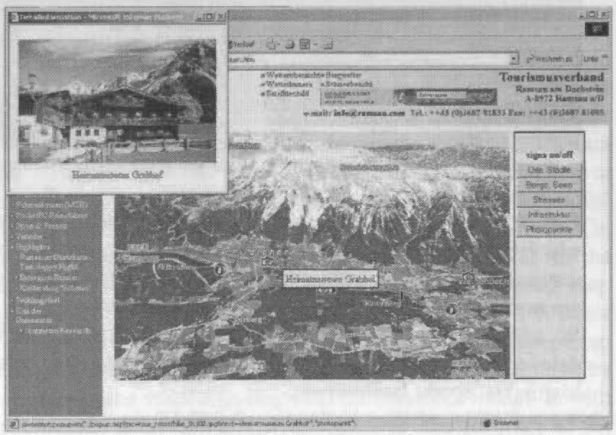


Figure 8: Touristic infrastructure in a 3D view

For visualisation 3D models, virtual panoramas, digital photos and satellite images are used. The digital tourism guide contains up-to-date information on infrastructure (see figure 8), geographic position, sport and events, as well as detailed infos on accommodations with the corresponding photos and a virtual stroll through buildings (see figure 9).

This geo-multimedia information system was combined with the region's existing information structures, that is to say, that e. g. the existing Internet presentations were enhanced and extended. Geo-multimedia data complete the comprehensive and effective existing tourism information systems. The geo-coded gathering and administration of all infrastructure objects (hotels, sports facilities, info points, etc.) permits position and place referenced processing. This in turn allows the tourist optimum orientation and provides regional information. In addition, download facilities are provided for mobile systems in order to download tourism information, route information etc. on mobile systems such as organizers or GPS devices.

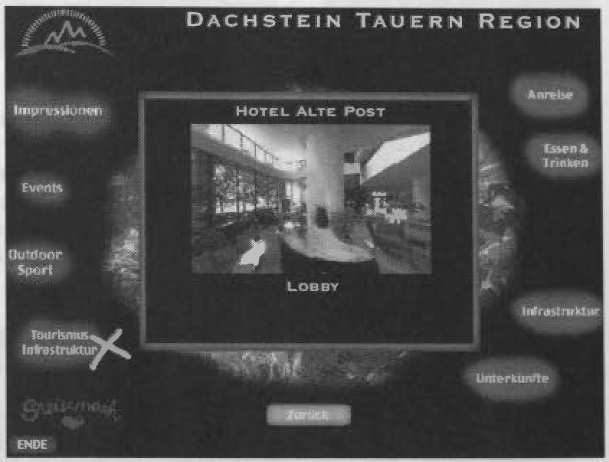
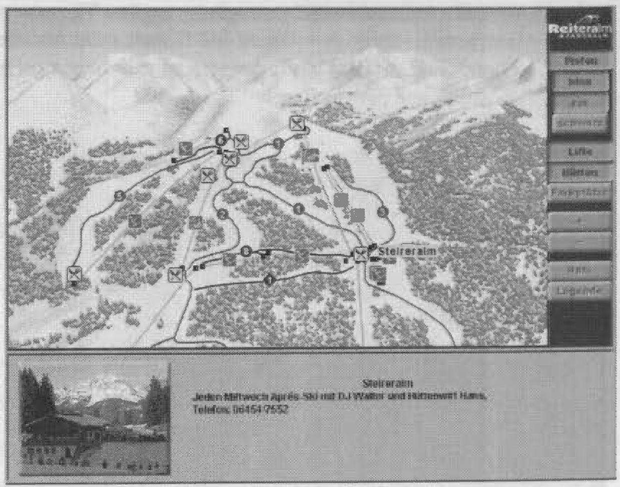


Figure 9: Virtual stroll through a hotel

5.5.4 Interactive Alpine Skiing Visualisation

On the basis of DEMs and forest classification results emerges a virtual winter landscape with modelled trees, houses and other objects. Information on the skiing areas (lifts, runs, huts) is inserted into the landscape and animated. The implementation of the visualisation with Macromedia Shockwave guarantees a high degree of interactivity when viewing the information (see figure 10).



Steireralm
Jeden Mittwoch Après Ski mit DJ Walter und Hüttenwirt Hans,
Telefon: 06454/7552

Figure 10: Virtual skiing area

The user may opt for the display of his individual information level, e. g. downhill runs, arranged according to their degree of difficulty, lifts, huts, guesthouses and parking facilities. Every symbol appearing on the screen provides additional information, which can be downloaded by means of a mouse over effect, and detailed descriptions and photos are available for huts, guesthouses and lifts.

5.5.5 Mobile information systems

For the potential client of a tourism region the region related information is particularly interesting. It is the combination of mobile devices with GPS devices that allows the user to relate the information individually to his present position ("Location

Based Services") and to obtain unlike a regular Internet user a geographically interesting piece of information via a mobile device. Nearly always, tourism related information shows to a large extent geographic references. Thus, it is useful for a tourist to receive recommendations concerning restaurants and leisure facilities within close range of a hotel and not of all the establishments in the whole region. Considering the main focuses of tourism in the "Dachstein-Tauern" region a PDA of the company Compaq (iPAQ) and a Garmin GPS device were applied for the presentation of particular topics.

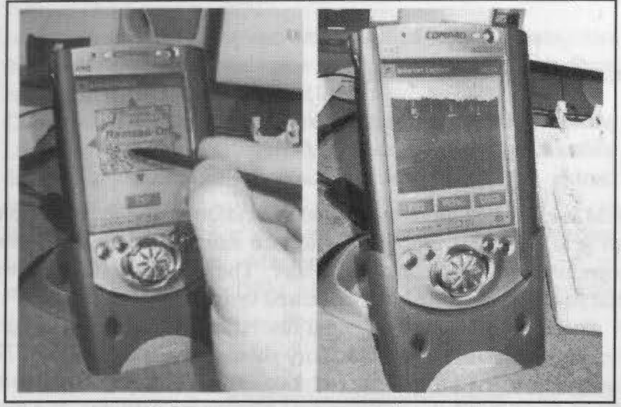


Figure 11: Tourism information on an iPAQ

For a multimedia visualisation of tourism information, picture, sound, text and video elements were integrated into the PDA in order to record the effectiveness of the mobile device. Further information is obtained by cartographic illustration (see figure 11) of the site and its surroundings. The user may move on the map and different areas can be seen. Particular info points demonstrate that for an object depicted on the digital map a different information level with multimedia objects is available. For these info points descriptions, spoken texts or short video clips are available. The device's storage capacity is adequate for the realisation of a comprehensive and smooth presentation of a whole region.

In the sector of outdoor sports the use of Handheld GPS receivers is gaining more and more momentum. In particular for hiking and mountain biking the use of GPS devices in combination with routes and waypoints is highly suitable. Previous to the hiking tour points are downloaded to the device in order to orient oneself. Since the deactivation of "Selective Availability" the accuracy of the GPS positioning is approx. ±10 m (Hofmann-Wellenhof et al., 1997) and is therefore more than sufficient for outdoor navigation. So as to provide tourists with an optimum orientation guide it is possible to download GPS coordinates for hiking tours to Garmin GPS receivers. These routes offer a suitable orientation guide, above all when using the route mode, where the device calculates and displays the respective direction and distance to the nearest point of the route.

6. CONCLUSIONS

The existing competition in tourism demands new and innovative strategies for the presentation of the tourism regions. In order to reach important tourist target groups it is vital to prepare information in a geo-multimedia information system and to present these data on different digital media. The implementation of a geo-multimedia information system in a tourism region requires a close co-operation with the tourism boards within the framework of a working concept which is established for various years.

The user-related presentation ("Customisation") of information alongside the possibility to link this information to a geographic position ("Location Based Services") will play a crucial role in the acceptance of the digital tourism information services. Because of the rapid technological progress in the telecommunications industry the integration of the mobile information systems into the multimedia tourism information concept is also of prime importance.

7. BIBLIOGRAPHY

Almer A., Nischelwitzer A. K., 2000. 3D Visualisation of Leisure & Tourism Information based on Remote Sensing Data. ISPRS - Technical Commission V/5, ISPRS Congress Amsterdam, 16-23 July 2000.

Almer A. et al., 2000. Interactive Multimedia Information System for Alpine Travel & Tourism Based on Satellite and GIS Data. European Commission, Joint Research Centre, Space Applications Institute, Project N0 15741-2000-02 F1SC ISP AT. Project-report and project CD-Rom. Ispra/Graz. Dec. 2000.

Almer A. et al., 2002. Vernetzte Multimediale Informationssysteme für räumlich, zeitlich und thematisch strukturierte Daten. Founded by "Bundesministerium für Bildung, Wissenschaft und Kultur". Project-report and project CD-Rom. Graz. January 2002.

Fritsch D., 2001. Positionsbezogene Dienste: Mit Mehrwert angereicherte Geodaten. *Geo-Informationssysteme*, 9/2001.

Hofmann-Wellenhof et al., 1997. *GPS. Theory and Practice*. 4. Edition. Vienna / New York: Springer Verlag.

Patterson, T., 1999: *Designung 3D Landscapes. In: Multimedia Cartography*. Berlin / Heidelberg / New York: Springer-Verlag.

Questar Productions, LLC, 1998. *World Construction Set 4. Using WCS 4.* Questar Productions, LLC, 1058 WCR 23.5, Brighton, Colorado 80601, USA.

Raggam, J. et al., 1999. *RSG in ERDAS IMAGINE. Remote Sensing Software Package Graz. Field Guide.* RSG Release 3.23. JOANNEUM RESEARCH Forschungsgesellschaft mbH., Graz, Austria.

Sloggett D., 1997. Earth Observation Data Utilisation. CEO Travel Study – Final Report. Project Reference: RGC9B/96.

OPTIMUM SPATIAL RESOLUTION OF DIGITAL ELEVATION MODEL FOR TOPOGRAPHICAL ANALYSIS

Masataka TAKAGI, Hiroshi ASANO and Yuki KIKUCHI

Dept. of Infrastructure Systems Engineering, Kochi Univ. of Tech.,
takagi.masataka@kochi-tech.ac.jp

Commission IV, WG IV/6

KEY WORDS: Digital Elevation Model, Spatial Resolution, Accuracy, Landform Classification

ABSTRACT:

An influence of spatial resolution for topographic analysis using DEM is evaluated. In classification of landform, number of category and spatial resolution was compared. And the possibility of 50m grid DEM for topographic feature extraction was mentioned. Moreover, very high resolution DEM was prepared using photogrammetry and laser range finder. The accuracy of DEM was intended to generate 1m in spatial resolution. Finally, required spatial resolution of DEM was discussed using auto-correlation.

In landform classification, the results showed number of category must set depending on spatial resolution. In case of 400m grid DEM, number of category should be less than 3 to 6. In case of 50m grid DEM, number of category should be less than 6 to 8. The optimum number of category for landform classification was existence depending on spatial resolution. Therefore, very high spatial resolution was required for detail analysis. Nowadays, 50m grid DEM is very popular and published by many institutes. We must recognise limitation of analysis by using 50m grid DEM.

In this study, 1m grid DEM was generated. Digital photogrammetry was used in mountainous area and laser range finder was used in glass land. An auto-correlation function of generated DEM was calculated. The results showed 8m spatial resolution was required to keep 0.9 of auto-correlation coefficient in mountainous area, 16m spatial resolution was required in case of hill area. In the present situation, meaning of auto-correlation coefficient in DEM is not understood clearly. However the auto-correlation coefficient would make one kind of index for complexity of landform.

1. BACKGROUND

Digital Elevation Model (DEM) is indispensable for many analyses such as topographic feature extraction, runoff analysis, slope stability analysis and so on. Those analyses require generally a high accurate DEM. The accuracy of DEM is usually represented by spatial resolution and height accuracy. The accuracy influences results in those analyses. Therefore, optimum accuracy of DEM for each analysis must be found.

In Japan, 50m grid DEM was published throughout the country. The DEM has been widely applied in many GIS users. However, there have not been many discussions about influence of DEM accuracy for each analysis. Now one question arose "Does 50m grid base DEM have enough accuracy for topographic analysis?"

On the other hand, high accurate DEM can be generated by digital photogrammetry, airborne SAR, laser scanner and so on. These methods have an ability to measure in 1cm accuracy. Therefore, optimum spatial resolution can be evaluated in the current condition. And finding optimum spatial resolution will make improve efficiency for measurement, computing time and data storage.

2. OBJECTIVES

In this study, an influence of spatial resolution for topographic analysis using DEM was evaluated. In classification of landform, number of category and spatial resolution was evaluated. And the possibility of 50m grid DEM for topographic feature extraction was mentioned. Next, very high resolution DEM was prepared using photogrammetry and laser range finder. The accuracy of DEM was intended to generate 1m in spatial resolution. Finally, required spatial resolution in some topographic conditions was discussed.

3. SPACIAL RESOLUTION AND TOPOGRAPHIC FEATURE EXTRACTION

3.1 Materials

In this chapter, 50m grid DEM was used, which provided by geological survey institute, Japan. From the DEM, many low spatial resolution DEM were generated which are 100m, 200m and 400m resolution by the nearest neighbor resampling. And, DEM in various topographic conditions was prepared for evaluation. Because, the topographic analysis will make unique results in each topographic condition.

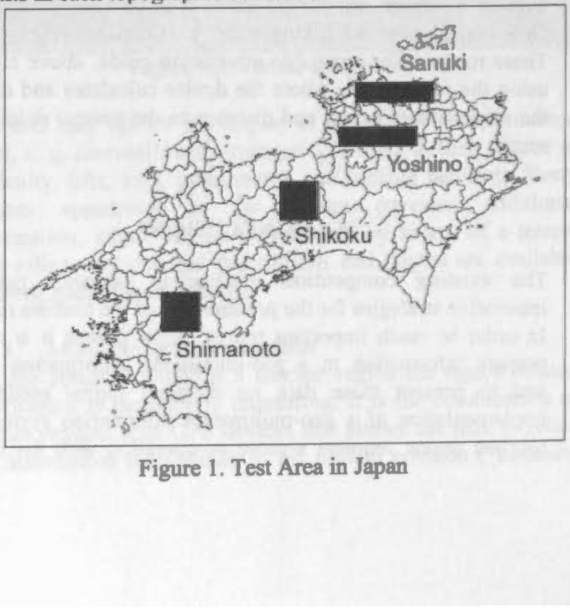


Figure 1. Test Area in Japan

As test area, three typical topographies were selected which are mountainous area pan (Shikoku and Shimanto), hill area (Sanuki) and terraced area (Yoshino). Figure 1 shows location of selected test area.

3.2 Landform Classification

Using those DEM, landform was extracted by classification with image processing. Three items of classification are an inclination, a concave-convex and a ridge-valley. The classification method was carried out following a paper by Iwahasi and Kamiya(1995)[1].

Inclination: The inclination was classified into two categories those are gentle slope and steep slope. The threshold value was calculated mean value in the test area.

Concave and Convex: For extracting concave and convex, Laplacian filtering was applied. Laplacian filter can calculate differential coefficient of second order, which can extract concave and convex.

Ridge and Valley: For extracting ridge and valley, Median filtering was applied. Median filter means one kind of smoothing. When the DEM after median filtering subtracted from original DEM, ridge indicates positive value, flat land indicates zero, valley indicates negative value.

The classified images have many noises come from detail undulation of ground surface. So, uniform linear transformation in 10 pixels radius was applied in all classified images.

The generated three images of classification were overlaid each other. Totally 12 categories are generated, which are consisted with steep or gentle, concave or convex and ridge, valley or flat. The classification results were compared each other according to spatial resolution.

3.3 Results

A relative frequency of each 12 categories according to spatial resolution was shown in Figure 2. The spatial resolution influenced seriously on concave-convex extraction. The concave area increased remarkably depending on increasing grid size.

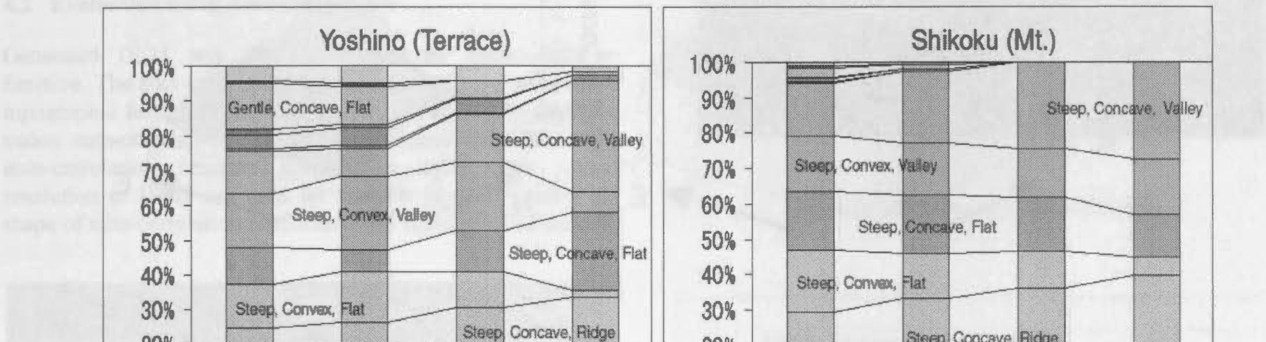
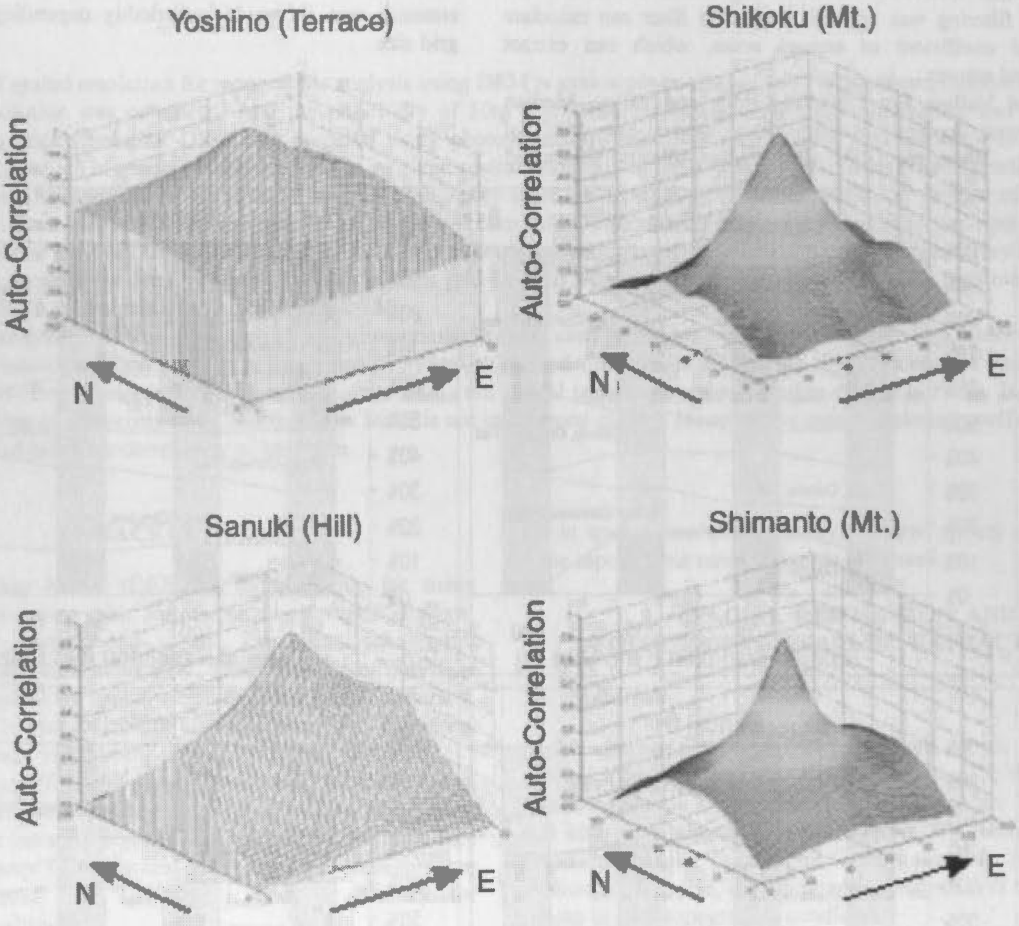


Figure 2. Classification Results in each Test Area

In the terraced area and the hill area, number of category decreased remarkably depending on increasing grid size. These areas have so many undulations on the ground surface that wide spatial resolution erased the small undulations.

The results showed optimum number of category was depending on spatial resolution. In case of 400m grid DEM, number of category should be less than 3 to 6. In case of 50m grid DEM, number of category should be less than 6 to 8. Those results meant that very high spatial resolution was required for detail topographic analysis. In this study, 12 categories were used for landform classification. However, even 50m grid DEM didn't have enough resolution.

Figure 3 showed auto-correlation function of 50m grid DEM in each test area. The auto-correlation is generally used in signal processing for extracting periodicity. When the auto-correlation is applied to DEM, complexity of landform might be extracted. The results showed auto-correlation coefficients fall remarkably in mountainous area. Terraced area kept high auto-correlation coefficients because of continuous same size of terrace. The hill area had anisotropic shape of auto-correlation function depending on direction because one big lineament crosses the test area. Anyway, landform in pointed shape of the auto-correlation function should need high resolution DEM. Therefore, the auto-correlation function of DEM could be represented topographic feature and optimum resolution of DEM might be shown.


Yoshino (Terrace)

Shikoku (Mt.)

Sanuki (Hill)

Shimanto (Mt.)

Figure 3. Auto-Correlation of Each DEM

4. REQUIRED RESOLUTION OF DEM

4.1 Data Acquisition

In previous chapter, possibility of 50m grid DEM was mentioned. 50m grid DEM didn't have enough resolution for landform classification in 12 categories. So, higher spatial resolution must be used.

Firstly, very high accurate DEM was prepared for detail evaluation on accuracy. Digital photogrammetry and laser range finder was used for DEM generation in this study. Selected test area was mountainous area in Shikoku, Japan as high undulated mountain. For comparison, Mongolian grassland also selected as gentle hill area. In the mountainous area, digital photogrammetry can be applied because of easy for stereo matching. In the hill area, laser range finder must be applied because of difficult for stereo matching. The accuracy of DEM was intended to generate 1m in spatial resolution.

4.1.1 DEM generation in mountainous area

For generating mountainous DEM, general digital camera with 4 million pixels was used. This digital camera has an ability to generate DEM in about 4cm accuracy in case of 30m distance from object points.

The test area was landslide prevention zone in Shikoku, Japan where is very steep area. Ten ground control points in the test area equipped, which were surveyed ground coordination by using GPS in static measurement. In the test area, 3 scenes were taken on the ground to look out over whole test area and almost ground control points. And stereo matching was succeeded with getting 1,116 corresponding points. By using these corresponding points, DEM was generated in 1m accuracy. Figure 4 shows generated DEM by shading. The mean value of inclination was indicated 18.9 degree.

4.1.2 DEM generation in grassland area

For generating DEM in grassland, laser range finder with 10cm accuracy was used. The laser range finder was mounted on telescope that can control attitude in 1 minute accuracy.

Test area was grassland in Mongolia. The laser range finder was set on ground control points where surveyed by GPS. And 511 points were measured ground coordination. By using these measurement points, DEM was generated in 1m accuracy. Figure 5 shows generated DEM. The mean value of inclination was indicated 3.2 degree.

4.2 Evaluation using Auto-correlation

Generated DEM was firstly evaluated by auto-correlation function. The auto-correlation function was efficient to evaluate topographic feature in previous chapter. Complicated landform makes auto-correlation coefficients fall remarkably. When the auto-correlation function keeps high value, low spatial resolution of DEM can used for analyse. However, when the shape of auto-correlation coefficients fall rapidly, DEM must be generated in high spatial resolution.

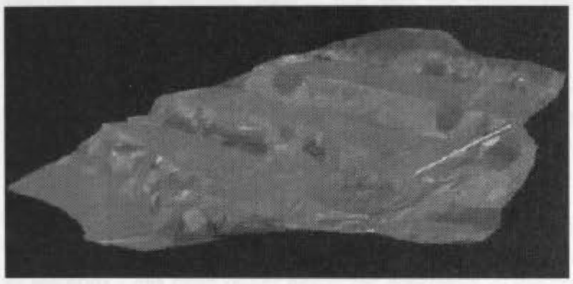


Figure 4. Generated DEM in Mountainous Area

Figure 6 shows each auto-correlation function in one profile. The results showed inside of 4m from origin point indicate to keep 0.9 of auto-correlation coefficient in mountainous area. On the other hand, inside of 8m from origin point can keep 0.9 of auto-correlation coefficient in hill area. The distance can be expressed 8m and 16m resolution each in spatial resolution.

5. CONCLUSIONS

In landform classification, influence of spatial resolution was mentioned. The results showed number of category must set depending on spatial resolution. In case of 400m grid DEM, number of category should be less than 3 to 6. In case of 50m grid DEM, number of category should be less than 6 to 8. The optimum number of category for landform classification was existence depending on spatial resolution. The results meant that very high spatial resolution was required for detail analysis. Nowadays, 50m grid DEM is very popular and published by many institutes. We must recognise limitation of analysis by using 50m grid DEM.

Next, 1m grid DEM was generated. Digital photogrammetry was used in mountainous area and laser range finder was used in glass land. And an auto-correlation of generated DEM was calculated. The results showed at least 8m spatial resolution is required to keep 0.9 of auto-correlation coefficient in steep mountainous area, 16m spatial resolution was required in case of gentle hill area. In the present situation, meaning of auto-correlation coefficient in DEM is not understood clearly. However the auto-correlation coefficient would make one kind of index for complexity of landform. The shape of auto-correlation function in each test area showed unique. It means there is possibility for classification of landform by using auto-correlation function. For definition of topography by using auto-correlation function, many experiments should carry out.

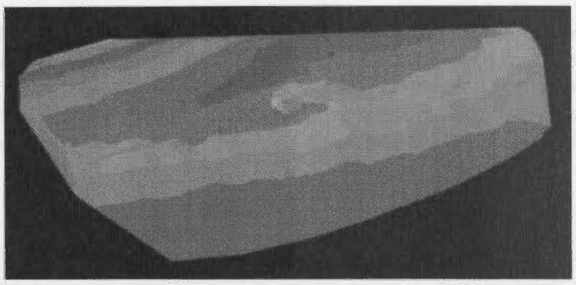


Figure 5. Generated DEM in Hill Area

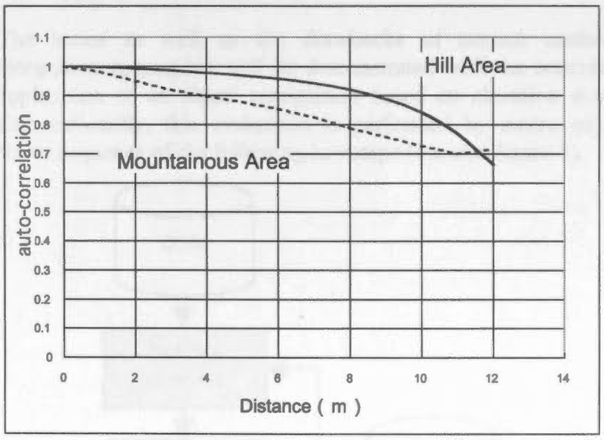


Figure 6. Auto-correlation Function

References:

[1] Junko Iwahashi and Izumi Kamiya, (1995), "Landform Classification using Digital Elevation Model by the Skills of Image Processing (In Japanese)", Geoinformatics, Vol.6 No.2, pp.97-108

[2] Masataka Takagi, (1998), "*Accuracy of Digital Elevation Model according to Spatial Resolution*", International Archives of Photogrammetry and Remote Sensing, Vol.32 Part 4, pp.613-617.

[3] Masataka Takagi (2000), "*Relationship between Spatial Resolution and Terrain Feature in DEM Accuracy*", International Archives of Photogrammetry and Remote Sensing, Vol.33 Part B4, pp.1038-1045

[4] F. Ackermann (1994), "*Digital Elevation Models - Techniques and Application, Quality Standards, Development*", International Archives of Photogrammetry and Remote Sensing, Vol.30 Part 4, pp.421 - 432

[5] P. A. Burrough (1986), "Principles of Geographical Information Systems for Land Resources Assessment", Oxford University Press.

DEVELOPMENT OF A METHOD NETWORK FOR OBJECT RECOGNITION USING DIGITAL SURFACE MODELS

G. Bohmann, J. Schiewe

University of Vechta, Research Center for Geoinformatics and Remote Sensing, PO Box 1553, 49364 Vechta, Germany, {gbohmann, jschiewe}@fzg.uni-vechta.de

Commission IV, WG IV/6

KEY WORDS: method integration, system architecture, object recognition, DEM/DTM/DSM, blunder analysis, normalization

ABSTRACT:

While the integration of data has been performed on various levels for quite a time, the logical consequence - the integration of methods - has been rather neglected especially in the context of the analysis of remotely sensed data. On the other hand the present way of combining methods is partially responsible for unsatisfying results which have to be noticed for instance for object recognition processes. Hence, goal of the paper is to demonstrate the respective drawbacks of currently applied linear sequence approaches, to present general design concepts for alternative method networks, and to describe a corresponding implementation for the task of object recognition based on data of Digital Surface Models.

1. MOTIVATION

It is well known that progresses in the fields of data acquisition and processing are acting as catalysts for the development of integrated evaluation approaches between but also within disciplines (e.g., Ehlers, 1993). With respect to the remote sensing domain we can observe not only the development of single sensors showing better geometrical, spectral and radiometrical properties, but in particular the trend to *data integration* which is driven by multi-sensor systems that acquire not only spectral but also elevation data (e.g., by laserscanning) and orientation information (e.g, by GPS/IMU) in a sequential or simultaneous mode.

While the integration of data has been performed on various levels for quite a time, the logical consequence - the *integration of methods* - has been rather neglected especially in the context of the analysis of remotely sensed data. Conventionally this task is performed by means of a linear sequence of the single, special-purpose processes. Beside the facts that a couple of these processes are far away from maturity and intermediate errors are propagated from one step to the other, especially the way of method integration is responsible for unsatisfying results which have to be noticed in the field of object recognition.

General goal of this paper is to verify the mentioned drawbacks of such a linear approach on one hand (chapter 2), and to present the concept (chapter 3) as well as an implementation example (chapter 4) of an alternative *method network*.

For these purposes we will concentrate on an object recognition based on information from Digital Surface Models, which have become a very important source for this task due to their improved operational features (e.g., availability) and technical characteristics (i.e, horizontal resolutions and vertical accuracies). The specific goal will be to demonstrate that an intelligent integration of the involved key processing steps - blunder analysis, terrain surface estimation and object recognition - leads to more reliable classification results in a network configuration instead of using a sequential approach.

2. CURRENT METHOD INTEGRATION APPROACHES

2.1 Application description

The status as well as the drawbacks of current method integration approaches will be demonstrated with the concrete application of an object recognition based on elevation data. Conventionally, this evaluation is performed by means of a linear sequence of the following key steps (see also figure 1).

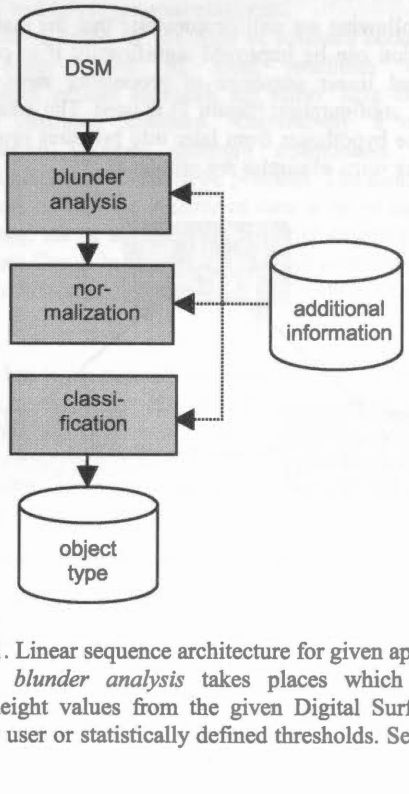


Figure 1. Linear sequence architecture for given application

Firstly, a *blunder analysis* takes places which eliminates extreme height values from the given Digital Surface Model (DSM) by user or statistically defined thresholds. Secondly, the derivation of object heights takes place by subtracting a given or an estimated Digital Terrain Model (DTM) from the DSM (*normalization*). In the case of the non-trivial estimation process, morphological filtering (e.g., see Vosselman, 2000), stochastical procedures (e.g., see Kraus, 1997) or region-based approaches (Schiewe, 2001) can be applied. Finally, these object heights (and eventually other parameters) are introduced into the *classification* step which is generally based upon probabilistic or fuzzy logic approaches. For a more detailed description of the algorithms which are used within our study we refer to the implementation example in section 4.2.

2.2 General problems

Applying such a typical evaluation process as outlined above some typical and commonly known problems occur. First of all the practical realization is done not only by one but by several software packages. These partially monolithic systems are heterogeneous with respect to their data structures and import and export functionalities so that a couple of data conversion processes have to take place (e.g., the elevation model is needed not only in the original point-wise ASCII-, but also in one or two raster image formats).

Furthermore, all methods including the transformations have to be invoked interactively by the user. Very often a batch processing is not possible due to the limited or not available batch functionality of one single component. A couple of information which are needed for the call of one function have to be repeated for another call.

The rather high efforts to invoke a component are one major reason for their single use within a linear sequence. The resulting drawbacks will be elaborated within the next section.

2.3 Problems related to sequential approach

In the following we will demonstrate that the quality of object recognition can be improved significantly if in contrast to the traditional linear sequence of processing steps (figure 1) a network configuration (figure 2) is used. The general idea is to backtrace hypotheses from later into previous processes. In the following some examples are presented.

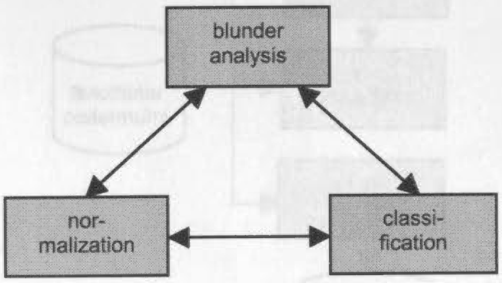


Figure 2. Network architecture for given application

It is obvious, that the **blunder analysis** can be improved by introducing information derived from the **classification** process: Based on a object type hypothesis one can predict its relative height behaviour and detect blunders by comparing this with actual data. For some regions one can assume constant height values (e.g., for waters), while for others height gradients will be very low in all directions (e.g., for airport runways, greenland) or at least in one direction (e.g., for roads).

After detecting blunders or data gaps their meaningful removal becomes necessary: Considering the object type associated to such points or areas one can optimize the method and window size for a reasonable interpolation of surrounding values. For instance, in figure 3 data gaps with the laserscanning data set occured due weak laser beam reflections. With the knowledge of the associated object being a building, the interpolation will take place only within the limits of the building in order to get a sharp transition to the surrounded terrain surface.

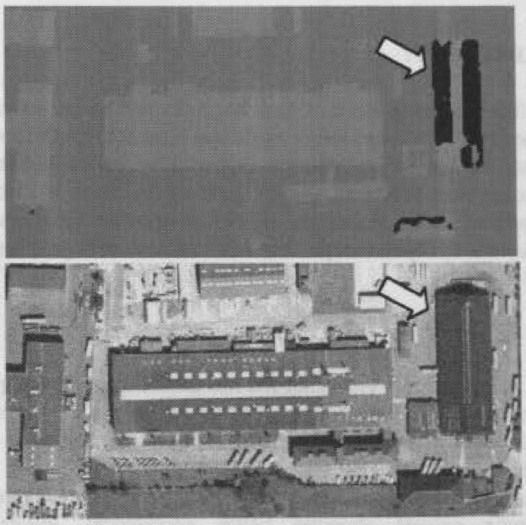


Figure 3. Extraction of interpolation information for data gaps in DSM (top) from semantical information (bottom)
- data courtesy of TopoSys GmbH -

But also the **normalization** process can be improved by introducing **classification** results. If for example morphological filter algorithms are applied for the detection and removal of regions within the DSM that do not belong to the terrain surface (in particular buildings and wooded areas), the critical filter window size can be derived from the actual object extent, or the filtering can be avoided at all if no such region was detected.

Some normalization algorithms separate the detection and the removal of objects, that stand clearly above the terrain surface, from their substitution (i.e., interpolation) which ends up with the so-called estimated Digital Terrain Model (eDTM). For some applications (e.g., hydrological modeling) it is necessary that only some of these regions under consideration will be interpolated (e.g., wooded areas) while others (e.g., buildings) have to be marked as blocking area because no water will actual flow here. Schiewe (2001) describes a respective region-based methodology for the separation of such draining and blocking areas.

Another important example for a meaningful DTM estimation is given in the case of removed buildings where the assumption of a horizontal plane instead of an interpolation within the surrounded, eventually inclined terrain represents a more suitable substitution. As figure 4 points out, the latter approach may lead to inconsistent and wrong object heights.

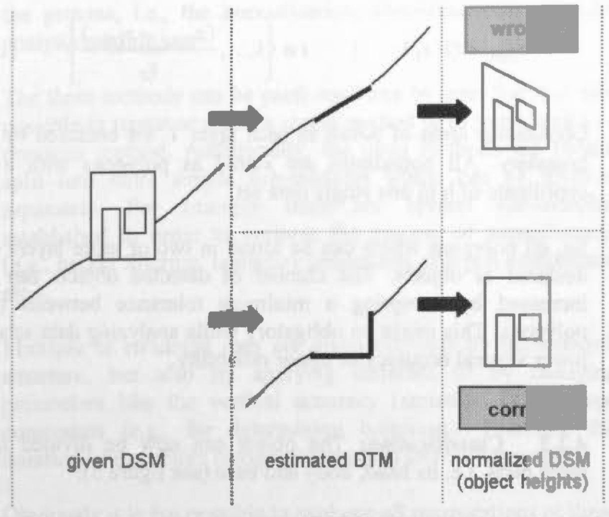


Figure 4. Choice of estimated DTM influences derivation of object heights

Finally, also the **blunder analysis** can be further improved by the results of the **normalization** process by introducing the extent of regions that have been reduced to the terrain surface and that are generally characterized by sharp rather than by ramp height edges.

In summary, the presented examples have pointed out that a significant improvement can be achieved by using a method network instead of a linear sequence architecture enabling the use of all hypotheses and information for all components.

3. CONCEPTUAL DESIGN OF METHOD NETWORKS

Central aim of the conceptual design of a method network is the optimized linkage of the involved components that allows for an objective control of the process with a minimum of user interaction. We will not concentrate on data modeling topics here but will focus on system architecture aspects by testing models coming from the software engineering domain on their applicability to the above mentioned problems. Based on general design criteria (section 3.1) we will present configuration solutions for current as well as for future systems (3.2 and 3.3, resp.), considering the experience that system integration is a evolutionary rather than a revolutionary process (e.g., Ehlers et al., 1989).

3.1 General design criteria

Central aspect of the design of a evaluation architecture is the definition of their connecting elements. With respect to their functionality we have to take into account (Abel et al., 1994)

- *transformation operations* for the exchange of data between the components,
- *constructor operations* for the (automatically or user-driven) generation of control commands, and
- *accessor operations* for the actual execution of these commands.

Designing these interface elements the general principles of continuity and safety have to be considered. Hence, as less information as possible should be exchanged (principle of *loose coupling*) and the number of interfaces should be kept to a minimum. With respect to the latter aspect a complete network between all n components (ending up with a number of interfaces of order n^2) would lead to a too costly and error-prone system.

3.2 Current configurations

For the design of current configurations we have to consider an integration of existing *closed components*. This assumption is based on various experiences that have shown that it is hardly possible to interfere with or to modify existing programs. It has to be noted that with this also an optimization of data modeling and handling will remain a difficult task.

In this context, we see a configuration using a *common interface module* as the best solution. The central component summarizes the user interface but in particular all connecting operations (transformation, constructor, accessor). The number of interfaces is reduced to a minimum (maximum of n interfaces for n linked components). Finally, the desired non-linear processing sequence can be controlled by this central component.

It should be noted that contrast to the field of Geographical Information Systems (GIS) where the general topic has already been discussed for a long time (e.g., within the Open GIS Consortium) and a couple of such architectures have been designed and implemented (e.g., see Abel et al., 1994; Waugh & Healey, 1986), for remote sensing evaluation systems no similar concepts have been presented so far.

3.3 Future configurations

General aim of future developments should be the possibility of an open usage of data and methods for a variety of users from distributed and heterogeneous platforms.
One realization could be the copy of software code (e.g., Java applets) from server to local machines (mirroring). Disadvantages of this approach are rather long downloading times and licensing problems. Alternatively, a standardized communication between software components placed on distributed platforms seems to be possible. The disadvantage of this approach is that the transfer of data to be processed could take too long. As an example for the latter structure the Object Management Group has presented the Common Object Request Broker Architecture (CORBA; OMG, 1998) for the GIS domain.

Finally, it has to be pointed out again that the proposed client-server-architectures for the integration of remote sensing software components are not yet to realize, because we still struggle with heterogeneous, not object-oriented data structures, too large software components and no standards that enable the connection to common interface modules.

4. IMPLEMENTATION EXAMPLE

The implementation of such a method network shown in figure 2 has been initially realized based on the concepts we introduced in sections 2.3 to 3.2. The desktop GIS ArcView® was chosen for implementation (section 4.1). For the single processing steps - blunder analysis, normalization and classification - case specific and not general purpose components have been applied (section 4.2) and connected to a desired network (section 4.3). Tests were performed on two different data sets (section 4.4). from which a couple of conclusions could be drawn (sections 4.5, 4.6).

4.1 Choice of ArcView®

Being aware of the variety of problems we decided to implement that network with only one software package - in our case the desktop GIS ArcView®. In contrast to the previous section's final conclusion a homogeneous and object-oriented data structure was supposed to fit best.

Originally, ArcView® is a vector based GIS which can properly model the object representation by continuos areas of elevation points either through its boundary or by characteristic relations (i.e. trends) between these points.

Additionally there are extensions available which process and store raster data, too. Therefore grids can be analyzed and those results can be stored either as raster or as vector data. Furthermore, ArcView® provides the capabilities to implement user-specific functions. This can be done by using Avenue™, an object-oriented script language. In summary, all necessary processing steps can be done within one software environment.

4.2 Description of components

Referring to figure 2 three closed components for the main tasks have to be taken into account. These methods are linked with each other by a common interface module which guarantees for the general design criteria posted in section 3.1 as well as for a free navigation between the components.

4.2.1 Blunder analysis: The blunder analysis has been reduced to gap detecting and gap filling procedures, because there is no guarantee that detected extrema that for instance have been returned by a bias analysis are real blunders and not real objects like flag poles or even wells instead.

Although bias analyses appear to be unsuitable, information about multiple biases and their dispersion can be taken into account for improving the classification of objects. Forest areas or tree groups may not be dense enough to cover all terrain points with its leaf area. Therefore, objects containing widely spread minima with small extents may be interpreted as vegetation.

It has been found that in contrast to figure 1 blunder analyses (i.e. gap filling) performed in a method network are best placed after normalization or classification.

4.2.2 Normalization: The task of normalization is to differentiate between the surfaces of the terrain and of outstanding objects. Considering the approaches mentioned in section 2.1 and based on the experiences that in particular morphological filtering might have negative effects on data quality (like loss of information) we prefer region-based approaches. The algorithm realized in this example is as follows: Depending on the vertical accuracy (s_Z) of the input DSM multiple selections have to be made. Each selection contains the points which heights (h) are greater than

$$h = z_{\min} + (i \cdot s_Z) \quad | \quad i \in \left\{1,\ldots,\frac{(z_{\max} - z_{\min})}{s_Z}\right\} \qquad (1)$$

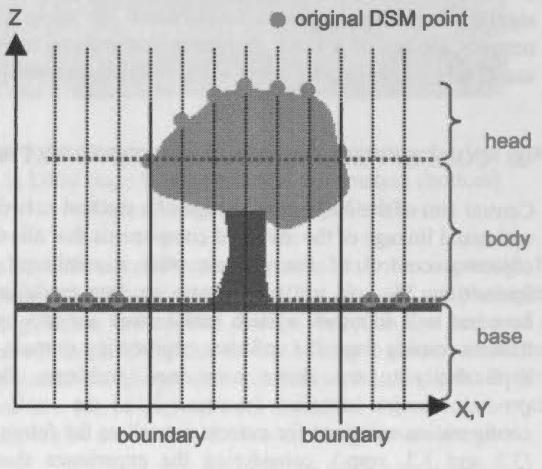
Continuous areas of points in each layer i are enclosed by its boundary. All boundaries are stored as polygons with a z-coordinate of h in one single data set.

So, all polygons which can be found in two or more layers are declared as objects. The number of detected objects can be increased by accepting a minimum tolerance between two polygons. This might be obligatory while analyzing data sets of lower vertical accuracy or minor reliability.

4.2.3 Classification: The object can now be divided into three parts, i.e. its head, body and base (see figure 5):

1) all points inside the boundary belong to the object's surface and can be considered as its head shape;
2) the minimum height of all DSM points inside the boundary gives an idea of its body height;
3) the maximum height of points outside next to the polygon can be declared as terrain height and as object base.

Figure 5. Object represented by DSM heights and its boundary



Furthermore the object can be described by analyzing the corresponding polygon within the detected boundaries considering the following parameters:

1) area, perimeter and volume;
2) compactness (2D, 3D);
3) rectangularity or parallelism of the boundary (after dividing it into line segments);
4) texture (e.g. standard deviation, variance) of the head points.

4.3 Fusion of methods

4.3.1 Concepts: As discussed in section 3.2 a common interface module has been established in order to control the application, to evaluate the current progress and, if necessary, also to stop it. It is also possible that the central module can switch back to a prior state if the classification has become worse.

Neglecting some pre-processing modules generating point data sets to work with, the network might be entered at any step of the process, i.e., the normalization, classification or blunder analysis module.

The three methods can be performed one by one. But it is also possible to repeat or recall a single method or to jump back to a previous method. Additionally, the main methods are further split into more simple sub-modules which can be invoked separately. For example there are several sub-modules established in order to perform the process of normalization (e.g., height filtering, boundary generation, polygon comparison etc.).

Multiple or recurring runs are given not only by the network structure, but also by applying different or by changing parameters like the vertical accuracy (equation 1) or shape constraints (e.g., for determining buildings). Therefore the duration of the object recognition process can vary strongly.

Obviously it is not possible to evaluate all permutations of these methods. Hence, we have tested only one which will be described in the following sections.

4.3.2 Status quo of implementation: Presently the implemented modules perform the operations of data conversion, height selection, boundary generation, polygon comparison, gap detection and filling, bias analyses, texture analyses, calculation and analyses of shape parameters.

Significant modules not realized yet are concerned with the detection and analysis of linear or planar trends. Furthermore the full potential of the control module is not implemented yet so that a couple of its duties are still performed by a human operator.

We prefer the normalization as starting point. Assuming that significant objects like buildings and trees show larger height values we start with an elevation interval of 1 m as s_Z (equation 1).

After the differentiation between objects and terrain an analysis of the object's boundary takes place. Area, perimeter and volume are calculated and compared to predefined values. Furthermore the boundary is simplified and the single line segments are compared with each other in order to search for rectangular or parallel sections. This boundary gives a first representation of the object. Nevertheless a blunder analysis is necessary in order to detect neighbouring gaps which can be adjacent to an object or belong to the object, respectively. Filling and joining the gap's and the object's area might lead to a better classification (compare figure 3).

Finally, a boundary-based classification completes the object recognition process. If the results are less satisfactory, the process is repeated taking a lower elevation interval, higher tolerances or both of them into account.

4.4 First results

Tests with the implemented method network have been performed with DSMs from two sensors: Two sites have been obtained with the TopoSys laser scanner (www.toposys.com) which produces first and last pulse elevation data with a height accuracy of about ± 0.2 m delivered as point data in ASCII. The other two DSMs have been derived by multiple matching from stereo imagery of the High Resolution Stereo Camera – Airborne (HRSC-A, http://solarsystem.dlr.de/FE/) given with a horizontal resolution of 0.5 m and a proposed vertical accuracy of ± 0.2 m.

Defining buildings as test objects the number of detected buildings was compared to the actual existing number (table 1).

Sensor		TopoSys				HRSC-A			
# Test site		1		2		1		2	
# Buildings		38		28		27		33	
Elev.	Tol.	Detected objects							
		#	%	#	%	#	%	#	%
1 m	0 m²	1	3	2	7	0	0	0	0
	2 m²	16	42	14	50	1	4	2	6
0,5 m	0 m²	8	21	5	18	0	0	1	3
	2 m²	29	76	25	89	4	15	11	33
0,25 m	0 m²	15	39	18	64	1	4	5	15
	2 m²	34	89	28	100	25	93	25	76

Table 1. Results of normalization depending on different elevation intervals ("Elev.") and tolerances ("Tol.")

It can be concluded that the minor the elevation interval and the higher the applied tolerance is, the larger the number of detected objects becomes. But the higher the elevation interval and the higher the tolerance is, the less reliable the corresponding results will be.

Exemplary the detected objects with an extent of more than 10 m^2 were classified. Differing only between buildings and vegetation, and assuming that buildings show certain parameters (area = 75 m^2, compactness = 0.4, standard deviation of elevation = ± 1.7 m) the tests already led to satisfying results (table 2). Buildings classified as trees show higher elevation standard deviations (i.e. up to ± 2.1 m) compared to the predefined parameters.

Objects	Classified as		Correctly classified
	Buildings	Trees	
Buildings	15	3	83 %
Trees	-	46	100 %

Table 2. Exemplary object classification

4.5 Gain of the network approach

Although the results of the methods of normalization (table 1) and classification (table 2) are not satisfactory yet the main advantage of the network approach already becomes obvious: Due to the recursive architecture valuable information can be exploited (in terms of data mining) for all modules while this is not possible using a linear sequence of methods.

As an example, classification parameters can be adapted with respect to the shape parameters of not correctly classified buildings. Referring to table 2 a classification adapting a modified standard deviation of ± 2.1 m led to further improved results (table 3).

Objects	Classified as		Correctly classified
	Buildings	Trees	
Buildings	18	-	100 %
Trees	2	44	95 %

Table 3. Exemplary object classification with adapted parameters

Furthermore the number or percentage of detected objects during the normalization process can determine acceptable tolerances and / or elevation intervals, respectively.

Hence, multiple runs can be evaluated by adapting parameters resulting from previous classifications and by comparing the new results with previous ones. If the classification is getting worse within a single run the currently applied parameters are to be neglected for future runs.

4.6 Problems and limitations

In fact the above described implementation example appears just as another linear sequence of methods. Due to the yet incomplete implementation the gain of the network approach could only be outlined in this chapter.

Beside the incomplete implementation also the single modules have to be further developed. For example, the algorithm failed to detect objects at elevation intervals of 1 m (table 1). This can be explained partially by the lower quality of the HRSC-A data sets derived by stereo matching (Bohmann, 2001). As a consequence, the normalization should be performed with intervals related to the vertical accuracy. However, this leads to longer computation times and requires a lot more disk space to store the data sets and its derivatives.

5. SUMMARY

The unsatisfying quality of object recognition procedures is partially due to the fact that no intelligent integration of the involved processing components is applied. Using various examples it has been shown, that in contrast to a linear sequence of methods a network architecture is able to improve the results of all inherent modules.

To realize this general idea we have presented general design concepts adopted from software engineering. While for current realizations a common interface module seems to be the best approach, for future developments an open usage from distributed platforms based on software code mirroring or on exchanging data and commands between distributed components should be taken into account.

We have presented an implementation example that aims for an object recognition based on information from Digital Surface Models. It is based on a common interface module which has been implemented under the ArcView® software environment. First experiences have proven the general applicability and gain of the network solution (in particular, the advantage of the recursive nature) but also the costs in terms of time and disk space. Further developments within the single modules as well as the networking elements have to be made in order to come a satisfying and operational solution.

REFERENCES

Abel, D.J., Kilby, P.J. & Davis, J.R., 1994. The systems integration problem. *International Journal of Geographical Information Systems*. 8(1): pp. 1-12.

Bohmann, G., 2001. Entwicklung eines Methoden-Netzwerks zur Integration Digitaler Oberflächenmodellen (DOM) in den Prozeß der Objektextraktion. Diploma thesis at the University of Vechta, Institute for Environmental Sciences.

Ehlers, M., Edwards, G. & Bedard, Y., 1989. Integration of remote sensing with geographic information systems: a necessary evolution. *Photogrammetric Engineering and Remote Sensing*. 55(11), pp.1619-1627.

Ehlers, M., 1993. Integration of GIS, remote sensing, photogrammetry and cartography: the geoinformatics approach. *GIS - Geo-Informationssysteme*. 6(5), pp.18-23.

Kraus, K., 1997. Eine neue Methode zur Interpolation und Filterung von Daten mit schiefer Fehlerverteilung. *Österr. Zeitschrift für Vermessung und Geoinformation*, (1), pp. 25-30.

OMG, 1998.: CORBA 2.0/IIOP Specification. Object Management Group formal document 98-07-01.

Schiewe, J., 2001. Ein regionen-basiertes Verfahren zur Extraktion der Geländeoberfläche aus Digitalen Oberflächen-Modellen. Photogrammetrie - Fernerkundung - Geoinformation, 2, pp. 81-90.

Vosselman, G., 2000: Slope based filtering of laser altimetry data. In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, NL, 33 (B3), pp. 935-942.

Waugh, T.C. & Healey, R.G., 1986. The Geolink system, interfacing large systems. In: *Proceedings of AutoCarto London*, pp. 76-85.

NATURAL COLOUR URBAN 3D MODELLING
A STEREOSCOPIC APPROACH WITH IKONOS MULTISPECTRAL AND PANCHROMATIC DATA

Yun Zhang
Department of Geodesy and Geomatics Engineering
University of New Brunswick
Fredericton, New Brunswick, Canada E3B 5A3
Email: yunzhang@unb.ca

Working Group IV/6

KEY WORDS: colour image map, 3D, urban modelling, IKONOS

ABSTRACT:
This paper presents an approach for generating urban 3D colour models using a stereoscopic technique and the commercial high-resolution satellite image data. The concept and result of the approach is presented. The paper demonstrates that it is possible to model an urban environment in natural colour 3D using the new commercial high-resolution satellite image data. The approach is fast and effective. Urban objects, including cars, can be viewed clearly in colour 3D.

1. INTRODUCTION

Three-dimensional (3D) information of the Earth's surface is important for a variety of geo-related applications. Geo-scientists, environmental planners, engineers, mapping experts, and military officers need 3D information of interested areas for better interpreting their spatial details and understanding their environments. Such 3D information is, some time, determinant for the success of the missions.

Stereoscopic viewing of images has been the most common method for obtaining 3D information of the Earth's surface for the last 50 years (Toutin, 2001). Different optical-mechanical, analytical, and, in the last years, digital 3D systems have been developed to meet the needs for 3D information perception and extraction with aerial and satellite images.

Black and white as well as colour or multispectral image pairs from different airborne and spaceborne sensors have been utilized to view and extract 3D information for different applications. They include, for example, analogue, digital, frame and linear images, as well as satellite images from different orbit viewing angles – adjacent track, across track and alone track.

However, results of a stereoscopy based modeling of 3D urban environment in natural colour has rarely been reported, especially results from the new commercial high-resolution satellite image data.

2. EXISTING TECHNIQUES

In the last years, studies on 3D information generation with stereoscopic images have mostly concentrated on the generation of digital surface models (DSMs) (Al-Roussan et al. 1997, Petrie 1999, Toutin 2001). Existing techniques for urban 3D modelling are, in most cases, based on the concept of draping a 2D image over a DSM. Such techniques are good for dynamically displaying and analyzing a 3D environment on a computer screen. It is, however, very costly and time consuming to generate an accurate DSM which can show details of individual buildings.

On the other hand, many 3D systems have utilized the stereoscopic technique to directly visualize 3D information of ground objects. The most common way is to display a pair of black/white images on a monitor or print them on a paper in two of the tricolours, and then to view the image through a pair of complementary stereo glasses. However, the 3D image viewed is monocolour.

A pair of colour stereo images can also be displayed on a computer screen for colour 3D viewing. However, expensive polarized screen and polarized stereo glasses are needed to view the colour 3D. Further, when the colour image pair is printed on a paper, no 3D can be viewed because the image cannot be polarized. Very few systems have utilized stereoscopic technique and inexpensive complementary stereo glasses for colour 3D viewing.

3. POTENTIAL OF HIGH-RESOLUTION SATELLITE IMAGES FOR URBAN 3D

The new commercial high-resolution satellites, such as IKONOS and QuickBird, have started a new era for experts in all geo-related areas to perform large scale and detailed 3D visualization and 3D mapping of the Earth's surface, especially for urban image mapping. Different high-resolution images (panchromatic: 1m and 0.61m, multispectral: 4m and 2.4m) in different stereo models (across track, adjacent track and along track) can be provided for different 3D viewing and extractions (Fritz 1996). The geometric accuracy has been met or exceeded the designed requirements (Cook et al. 2001) and meets (US) national mapping standards for generating digital elevation models and ortho-images at a map scale of 1:24,000 (Li 1998).

To date, examples of stereoscopic viewing of the high-resolution panchromatic images have been presented by some 3D systems. However, reports on generation and visualization of high-resolution 3D colour images by combining IKONOS or QuickBird panchromatic and multispectral image data have not been seen by the author.

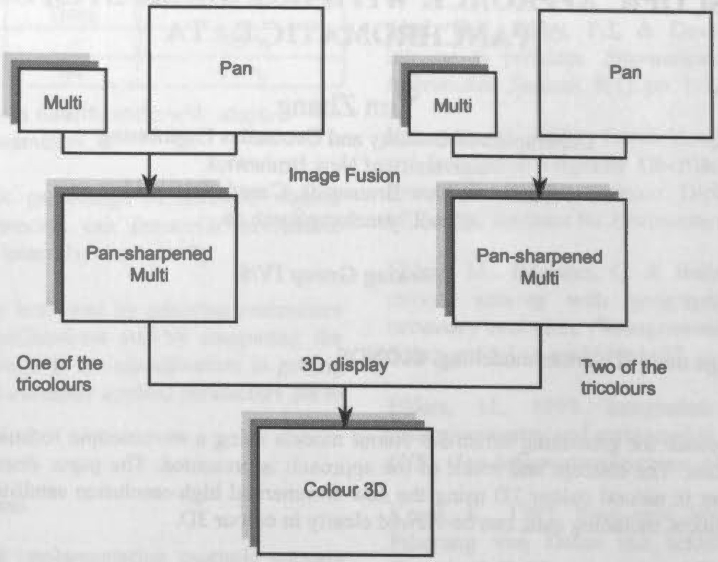


Figure 1. Flowchart of the generation of colour 3D anaglyph image

4. A NEW APPROACH

To better utilize the available high-resolution multispectral and panchromatic satellite images and to explore a more effective way for detailed colour 3D visualization at a lower cost, this paper presents a new approach to the generation of natural colour 3D stereoscopic images. The general concept of this approach is shown in Figure 1. The determinant step for this concept is the generation of well fused pan-sharpened natural colour images.

5. RESULT

Figure 2 shows a natural colour anaglyph image generated with IKONOS panchromatic and multispectral data. Different from existing 3D images, the 3D colour anaglyph image can be displayed on a screen and also printed on a paper. 3D colour information can be viewed with a pair of inexpensive complementary glasses. Moreover, the image generation process is simple and fast, so that near-real-time urban 3D colour modelling is possible.

6. CONCLUSIONS

This study has demonstrated that it is possible to model an urban environment in colour 3D using the new commercial high-resolution satellite image data. Urban objects, including cars, can be viewed clearly in colour 3D. The process speed of such colour 3D images is fast and the colour 3D viewing does not require expensive polarized screens and glasses. The colour 3D image can be printed on a piece of paper. The application potential is large.

7. ACKNOWLEDGEMENTS

The author thanks Mr. Rob Lunn, GIS supervisor of the City of Fredericton, NB, Canada for providing the raw IKONOS image data.

REFERENCES

Al-Roussan, N., Chen, P., Petrie, G., Toutin, Th., and Valadan Zoej, M.J., 1997. Automated DEM extraction and ortho-image generation from SPOT level-1B imagery. *Photogrammetric Engineering and Remote Sensing*, Vol. 63, pp. 965-974.

Cook, M.K., Peterson, B.A., Dial, G.; Gibson, L., Gerlach, F., Hutchins, K., Kudola, R., Bowen, H.S., 2001. IKONOS technical performance assessment. *Proceedings of SPIE - The International Society for Optical Engineering*, v 4381, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII, Apr 16-19 2001, Orlando, FL, p 94-108.

Fritz, L.W., 1996. The Era of Commercial Earth Observation Satellites. *Photogrammetric Engineering and Remote Sensing*, Vol. 62, No. 1, pp. 39-45.

Li, R., 1998. Potential of High-Resolution Satellite Imagery for National Mapping Products. *Photogrammetric Engineering and Remote Sensing*, Vol. 64, No. 12, pp1165-1169.

Pertrie, G., 1999. Automatic DEMs and orthoimages from SPOT stereo-pairs: SPOT DEMs of Jordan. GIM International, pp. 76-79.

Toutin, Th., 2001. Review article: Elevation modeling from satellite sivible and infrared (VIR) data. *International Journal of Remote Sensing*, Vol. 22, No. 6, pp. 1097-1125.

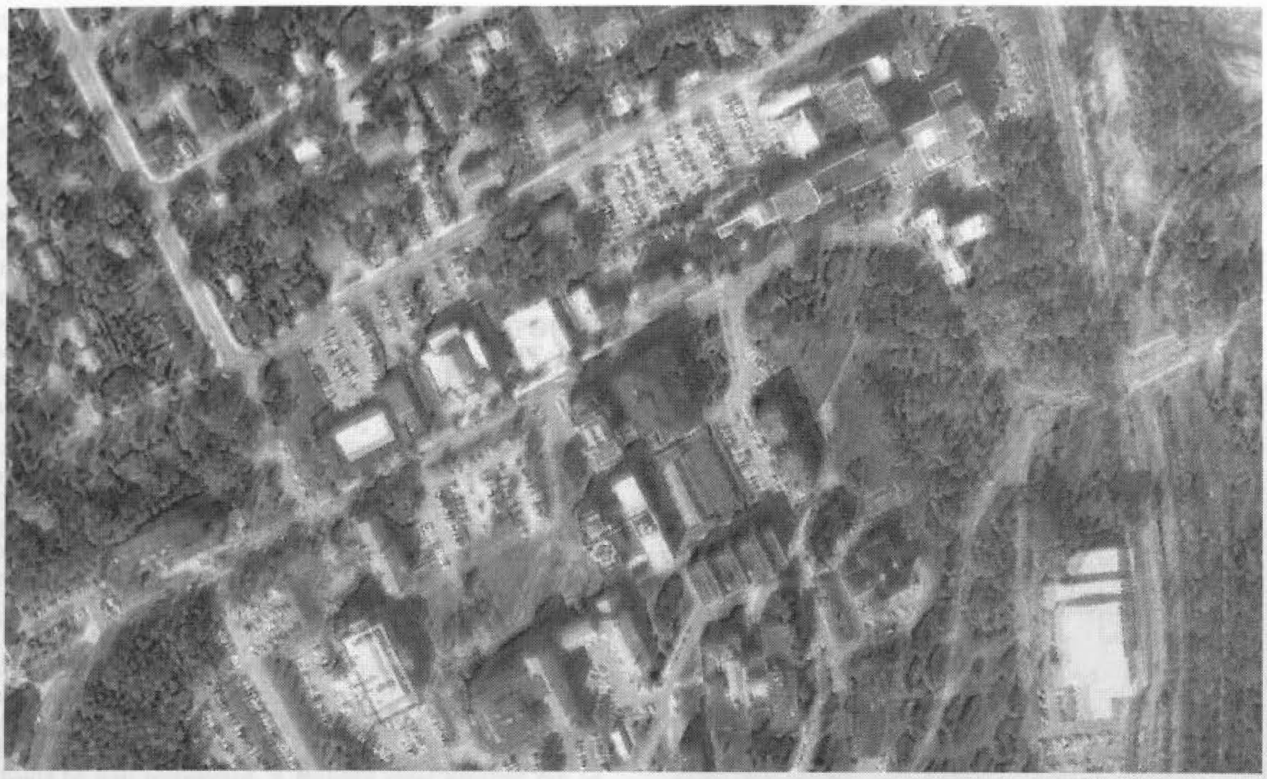


Figure 2. Natural colour anaglyph image generated with IKONOS panchromatic and multispectral data. The image shows a part of the campus of the University of New Brunswick, Canada.

3D-CITY MODEL SUPPORTING FOR CCTV MONITORING SYSTEM

Ming Ying [a], Jiang Jingjue [b*], Bian Fulin [a]

[a] The centre for spatial information and digital engineering research, Wuhan University, 129 Luoyu Road, China, 430079-brightm@public.wh.hb.cn
[b] School of Computer Science, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079-bigeyes@public.wh.hb.cn, whucindy@263.net

Commission IV, WG IV/6

Key words: 3D-city model, CCTV, security system, DEM, GIS, OpenGL

ABSTRACT: Closed circuit television (CCTV) monitoring system is widely used in many important fields to monitor fixed objectives or special areas due to its reliable capability of visual remote surveillance and control. In tradition, the CCTV system design is a complex process including field survey, plan and selecting proposed cameras sites. The visual design method is introduced in CCTV system design and maintenance. This paper suggests that the visual design method should be introduced and advantages of spatial analysis and information query of GIS and 3D city model should be fully used to aid CCTV monitoring system in design and maintenance.

1. INTRODUCTION

Due to its reliable capability of visual remote surveillance and control, CCTV monitoring system is widely used in many fields. It is not only an essential element of visual surveillance for Intelligent Transportation System (ITS), but also a useful monitor tool for detecting potential criminal activities in public areas or other special areas. It is also particularly suited to the protection of government or commercial buildings. CCTV monitoring system consists of various components including a CCTV camera unit, a controller cabinet housing the control equipment and a communication system, which connects the camera to a control center. Once a CCTV camera is mounted in a field, it is expensive to relocate. Therefore, thorough consideration needs to be given in the design to ensure proper sitting and mounting. In tradition, the monitoring system design is an abstract and complicated process, which consists of not only endlessly survey and calculation but also iteratively experimentation and adjustment. Furthermore, a lot of information about the proposed monitored areas must be looked up. When all devices are installed, the system can not be run immediately until those devices have been adjusted several times to provide a reliable monitoring. All of these will lengthen the design cycle time and increase system cost.

Today, with the development of three-dimensional Geographic Information System (3D GIS), quick reconstruction methods of 3D-city model make great progress. 3D city model is an accurate and visualized description for all fixed objects (building, vegetation, traffic way and waterways, etc) and their topological relations in a city (Monika, Ranzinger and Gleixner, 1997).

OpenGL is premier environment for developing portable, interactive 2D and 3D graphics applications. The representation of a 3D city model with terrain realistic scene can be implemented easily by powerful functions of OpenGL (Xue An, Ma Ai-nai, Li Tian-hong, 2001). In OpenGL, the transformation and representation of a view port in 3D coordinate space can be gotten conveniently. All these can provide a good tool for designers of a CCTV monitoring system to verify the condition of each camera viewing range in a 3D city model.

Spatial analysis, information management and inquiry, map and display are powerful functions of 3D GIS and 3D city model. It may be an effective means for both decreasing the design time and obtaining the better system performance by taking full advantages of them to plan the monitored areas, select cameras installing position and make maps of installing position for cameras, communication cable and other facilities. Also, all the data of locations, status and model of system facilities can be stored and managed in a database based on GIS, which can help operators use and maintain CCTV monitoring system.

2. RECONSTRUCTION OF 3D CITY MODEL

2.1 Reconstruction of outdoor 3D scene

Both the GIS and 3D city model of the surveillance areas must be created before a CCTV monitoring system is designed. To be able to view the results from the view port of the sitting and mounting CCTV cameras, one needs the 3D city model of the proposed monitored areas and the whole cityscape.

When the outdoor objects or areas are monitored, immobile objects in the monitored area may occlude part of view. Blind spots can be formed in back of objects. Improper blind area will threat the security of the system. Geometry size of objects, distance and height between objects and camera, and the undulate terrain surface are major factors to determine the size and distribution of blind spots. For this reason, when a 3D-city model is generated, terrain must be considered in order to represent the truth scene and to calculate blind spots. It is known that a building in a city always has the same height on its bottom boundary. Hence, the Digital Elevation Model (DEM) of this area, where a reconstructed building is superposed, is modified by the height of the building's bottom boundary. In order to get or extract the shape and size of a building, vegetation and other fixed objects for the reconstruction of city scenes, there are several methods can be used. One method (Bailard, Dissard and et al, 1998) is developed to extract aboveground areas from aerial stereo pairs. Another (Haala and Brenner, 1999) combines multispectral image and laser altimeter data in an integrated classification for the extraction of buildings, trees and grass-covered areas. Automatic reconstruction of buildings obtained by using airborne laser data and 2D ground plan information is also an efficient idea (Brenner, 1998). Frequently, these ground plans can be obtained either in analog form by maps and plans or in digital form by 2D GIS. Of course, DEM can be acquired from airborne images like stereo images or active range data like laser range scanner. Airborne laser scanners can usually provide more accurate DEM of high and homogeneous quality.

---

* Corresponding author. Tel.:86-027-87885503;
E-mail: bigeyes@public.wh.hb.cn

In order to reconstruct the 3D scene of a range around monitored area, we also can apply the method (Pollefeys and Koch, 2000) that reconstructs 3D scenes from sequences of images taken with of-the-shelf consumer cameras.

2.2 Reconstruction of indoor 3D scene

The generation of indoor scenes is a relevant and challenging task, since there are so many objects need to be reconstructed. However, The major factors affecting indoor camera view are the building inner structure and facilities or furnishings, which should be well represented. A building inner structure can be reconstructed according to its construction map, while furnishings can be reconstructed by assembling components (Kiyoshi and Kiyokawa, 2000), which are generated in advance and stored in component base. When reconstructing the indoor scene, these components, whose sizes can be adjusted, are put on their scene position. All texture can be obtained from their surface images, taken by still video cameras.

Indoor scenes also can be reconstructed by other methods. One method (Seqveria, Ng, et al, 1999) is to reconstruct indoor scenes by combining spatial data acquired by a Laser Range Finder (LRF), and visual data, acquired by a video or digital camera. Another method (FrankA, 1998) is to use a single image and geometric constraints for reconstruction.

3. CCTV SYSTEM DESIGN

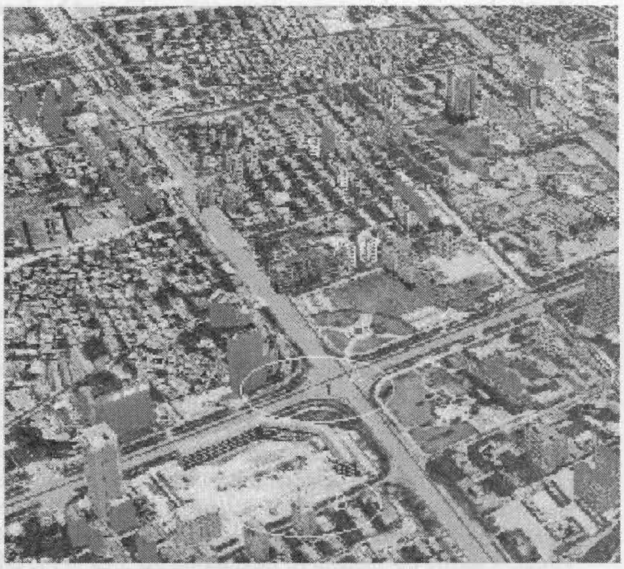
3.1 Planning monitored areas

To plan monitored areas is to propose all the areas that need to be monitored by CCTV cameras according to the objective of CCTV monitoring system. Since full coverage of all areas around a monitored target in an urban area would be cost-prohibitive, a plan of monitored areas must be determined strategically based on the applied objective of every CCTV monitoring system.

A system's planning solution of proposed monitored areas is different from others due to their difference objective. In an ITS, the primary objective of a CCTV monitoring system is to provide surveillance of freeway/highway segments or intersections and obtain a visual confirmation of incidents (Lee, Zachariah and Evarett, 1995). The CCTV monitoring system in a security system usually needs to provide surveillance in many different situations from parking areas and commercial districts to housing, recreational and transport facilities (Thiel, 2000). Besides providing a visual confirmation of incidents, the CCTV monitoring system in the security system supervises all activities in surveillance areas.

At this stage, the proposed monitored areas are marked on the 3D city scene. Also, an approximate number of CCTV cameras can be acquired according to the camera viewing range. Because of its fixed resolution and angle of view, the installing position of a camera can determine its view area. Typically, the CCTV camera viewing range is half a mile (for 10 to 1 zoom). Figure 1 shows the part scene of a 3D city model. Where, the white ellipse A and the white ellipse B are two proposed monitored areas.

Figure 1. The part scene of a 3D city model



3.2 Selection of cameras position

Selection of CCTV cameras position is a key work in a CCTV monitoring system design. Both power supply and communication cable network designs are based on them. The process is presented as following.

3.2.1 Proposing preliminary camera locations

In the 3D city scene, the best proposed camera location is selected as a preliminary camera location, which can provide the best monitoring and coverage of all the proposed areas, in terms of the CCTV camera view.

3.2.2 Checking blind spots

In the monitored area, parts of camera view may be occluded due to geometric constraints, other physical structures, fixed objects or existing planting. Improper or excessive blind spots will limit the camera view and degrade the system dependability. Considering the position of the supervised objects on the 3D-city model, the height and direction of cameras can be selected and adjusted by mouse, while the blind spots can be automatically calculated and displayed by h a spatial analysis algorithm (Zhou Yang, Tan Bing and Xu Qin, 2001). Then one can review all areas supervised by monitoring system.

If the proposed preliminary camera location needs to be adjusted in order to eliminate or minimize blind spots, the camera at the new location can still provide a similar viewing range.

3.2.3 Checking safety and repellence

All the underground and aerial existing facilities, such as irrigation facilities, high or low-risk utilities, drainage and existing planting, which will be impacted by construction activities for installing an camera, are displayed on the 3D scene where cameras are located around. It may be possible to move the camera location slightly to avoid impacting existing facilities, and still provides a better viewing range. At this step, the preliminary cameras locations are picked.

3.2.4 Checking available power source

Once the preliminary cameras locations have been picked, a check for available power source (utility company power pole, electrical service cabinet or underground vault) must be done by displaying all information of power sources around every preliminary camera location in the 3D scene. If the available power is too far to provide power, one should then consider moving the proposed cameras to another location that provides a similar viewing range. So, the process of selecting camera location from the sentence (3.2.1) to the sentence (3.3.4) needs to be done again.

3.2.5 Determining final cameras position

This phase is the video survey, which is a more detailed survey of all the proposed cameras viewing range at the various locations. The main purpose of the final site review is to verify cameras viewing range, select the best correct mounting height, and eliminate blind spots and significant overlaps of viewing range via displaying the 3D scene of each camera's viewing range. Finally, the best scheme for all camera positions can be acquired by the interactive way. Figure 2 shows the camera A, monitoring the area in the ellipse A shown in Figure 1. Similarly, Figure 3 shows the camera B, monitoring the area in the ellipse B shown in Figure 1.

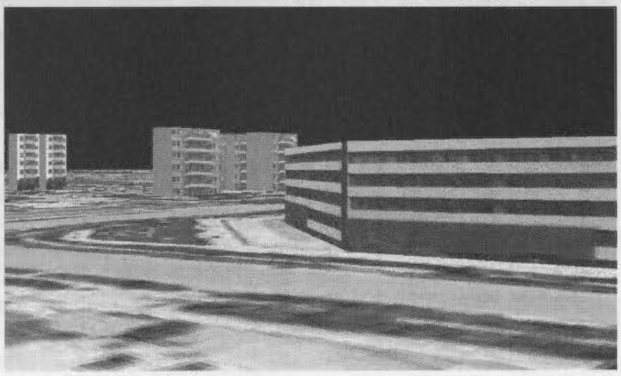


Figure2. The 3D scene of the A camera's viewing range

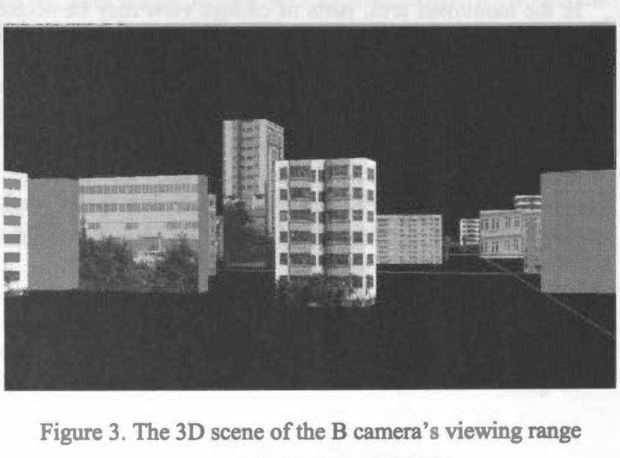


Figure 3. The 3D scene of the B camera's viewing range

3.3 Power supply and communication network

Because final camera positions are determined after checking available power source, the power supply solution can be obtained easily. On the other hand, according to final cameras positions, the installing plan of communication cable network from every camera to control center or controller cabinet can be made out in the same way as cameras positions are done. Then, the cable's length can be acquired by GIS network analysis function. Both number and installed position of the signal amplifiers can be calculated in terms of the cable's length and attenuation rate and amplifier's actual magnification (Gu Youshi, Qi Shijian, et al, 1995). In a word, the construction plan can be easily made out by the interactive way.

4. AIDING IN OPERATION AND MAINTENANCE

When the CCTV monitoring system is running, the 3D city model and CCTV cameras location map can be used to indicate the position corresponding to every image from cameras. At the same time, operator can watch a specified area or object conveniently on the 3D city model or 3D city scene when all of them are integrated with the control system of the CCTV monitoring system.

Both the cameras location map and the network map of power supply and communication cable may provide great help to find fault position. The model data about all the system's facilities can be used as referenced information for their replacing and repairing.

5. CONCLUSIONS

GIS and 3D city model can conveniently provide a lot of abundant and accurate and visual background information for CCTV monitoring system design. The visual design method may reduce cycle time and labour load of design, together with cost for a large CCTV monitoring system.

REFERENCE

Monika, Ranzinger and Günther Gleixner, 1997. GIS Datasets for 3D urban planning. Comput. Environ. And urban systems. Vol 21, No 2, pp 159~173.

Xue An, Ma Ai-nai, Li Tian-hong, 2001. A study on the Representation of Terrain Realistic Image Based on OpenGL. Journal of Image and Graphics. Vol. 6 (A), No. 8, Pp. 800~805.

Norbert Haala, Claus Brenner, 1999. Extraction of buildings and trees in urban environments. ISPRS Journal of photogrammetry & Remote Sensing, Vol. 54 (1999), Pp. 130~137.

C. Baillard, O. Dissard, O. Jamet, H. Maitre, 1998, Extraction and textural characterization of above-ground areas from aerial stereo pairs: a quality assessment. ISPRS Journal of Photogrammetry &Remote Sensing, Vol. 53 (1998), Pp.130~141.

Claus Brenner, Norbert Haala, 1998. Rapid acquisition of virtual reality city models from multiple data sources.

International Archives of Photogrammetry and Remote Sensing, Vol.XXXII, part 5, Pp. 323~330.

M. Pollefeys, R. Koch, M. Vergan Wen, L. Van Gool, 2000. Automated reconstruction of 3D scenes from sequences of images. ISPRS Journal of photogrammetry & Remote sensing, Vol. 55(4), Pp. 251~267.

V. Sequeira, K. Ng, E. Wolfart, J. G. M. Goncalves, D. Hogg, 1998. ISPRS Journal of photogrammetry & Remote sensing, Vol53 (1998), Pp.354~368.

Kiyoshi, KiyoKawa, 2000. Seamless-Design for 3D object creation. IEEE Multimedia, Vol. 7 (1), Pp.1~9.

Frank A. Van den Heuvel, 1998. 3D reconstruction from a single image using geometric constrains. ISPRS Journal of photogrammetry & Remote sensing, Vol53 (6), Pp. 354~368.

Lee, L. K., Zachariah, M., Evarett, P., 1995, CCTV Camera Site Selection: A Field Experience. Vehicle navigation and information systems conference, 1995. Proceedings. In conjunction with the Pacific Rim Trans Tech conference, 6th International VNIS, ' A Ride into the future' 1995, PP21~27.

Thiel, G., 2000. Automatic CCTV surveillance-towards: the VIRTUAL GUARD. IEEE Aerospace and Electronics Systems Magazine, Vol. 15 (7), Pp. 3~9.

Zhou Yang, Tan Bing and Xu Qin, 2001, A Spatial Analysis Algorithm Based on 3D Topographic Map, Journal of Institute of Surveying and Mapping, China, Vol.18 (1), Pp. 39~43.

Gu Youshi, Qi Shijian, Ye Qiu, Zhang Wenyi, 1995. CATV system design, installation, adjustment and maintenance. The Republic of postal and telegraphic publishing company, Beijing, Pp.199, Pp. 328~344.

USE OF GIS AND MULTITEMPORAL IMAGING SPECTROMETER DATA FOR MODELLING AND MAPPING ENVIRONMENTAL CHANGES IN MINING AREAS

Christian Fischer
Clausthal Technical University
Institute for Geotechnical Engineering and Mine Surveying, IGMC
Erzstrasse 18, D-38678 Clausthal-Zellerfeld, Germany
christian.fischer@tu-clausthal.de

KEY WORDS: Environment, GIS, Vegetation, Change Detection, Monitoring, Hyper spectral, Multitemporal

ABSTRACT:

Deep hard coal mining activities lead to land subsidence, resulting in infrastructural influences and hydrological changes. In the case of underground mining, a dislocation zone which transforms in the upward direction into flexure of the strata occurs directly above the mined area. A depression is formed on the surface. The shape of this depression depends on the structure of the deposit, mining parameters and the overlaying geology. Spatial and temporal influences of the environment result from the amount of subsidence, the local ground-water situation and actual land-cover and land-use. An important factor is the intensity of impairment regulated by properties of the local soil and vegetation data. To estimate changes of the dynamic system "subsidence-ground-water-vegetation", ground-water recharge rate and isobars of the ground-water table were calculated in detail for a small-parcelled area of 110 sqkm using a GIS. Because of the large extent of the area to be monitored, airborne surveys were done in August 1998 and in August 2000, using the HYMAP™ imaging spectrometer to detect affected vegetation stands.

KURZFASSUNG:

Untertägiger Steinkohlenbergbau führt zu Bodenbewegungen an der Tagesoberfläche. Diese führem zu Beeinflussungen der Infrastruktur und zu Veränderungen der hydrologischen Situation. Durch die großen Teufen, in denen der Abbau stattfindet, erstrecken sich die Gebirgsbewegungen innerhalb bestimmter Grenzwinkel trogförmig über weite Bereiche an der Tagesoberfläche oberhalb des Abbaus und es bildet sich eine Senkungsmulde. Die Ausprägung ist abhängig von der Lagerstätte, Abbauparametern und der lokalen Geologie. Eine räumliche und zeitliche Beeinflussung der natürlichen Umwelt erfolgt aufgrund der auftretenden Senkungsbeträge, der lokalen Grundwassersituation und der aktuellen Landnutzung und Landbedeckung. Die Intensität der auftretenden Beeinflussung an einem Standort wird durch Parameter des Bodens und der Vegetation gesteuert. Um bergbaubedingte Veränderungen in dem dynamischen System „Senkung-Grundwasser-Vegetation" erfassen zu können, wurden für ein 110 km² großes Untersuchungsgebiet GIS-gestützt die Grundwasserneubildungsrate und die Grundwasserflurabstände berechnet. Aufgrund der Größe des Untersuchungsgebietes wurden im August 1998 und 2000 Befliegungen mit dem bildgebenden Spektrometer HYMAP™ durchgeführt, um Veränderungen der Vegetation zu erfassen.

1. INTRODUCTION

Underground mining activities lead to subsidence and ground movements at the topographic surface. The present and future modifications of the topographic situation lead impairments of the infrastructure and to a permanent influence of the technical infrastructure as well as changes of hydrological and ecological parameters.

The management of mining activities has to be done in a dynamic way, due to be bound to deposit structures and geological structures. Present procedures for environmental monitoring are based by predeterminations of rock mechanics of the overlaying strata, topographic measurements done with photogrammetric measurement techniques and upon calculations of three-dimensional stationary ground-water models. Since these different data layer are present only to discrete times and in different scales, a consistent description of the effects and the prognosis of future modifications are severely limited.

Therefore space and time-variant models are necessary, which can describe measured variables and resulting influences related to the considered geo-system. This is the prerequisite to detect relevant ecologically effects and to control and simulate administrative measures. Because of the large extent of the affected area and the small-parcelled landscape structure, only a detailed monitoring concept together with GIS modelling techniques and high resolution imaging spectrometer data permit an appropriate way to update the large GIS database and to fulfill legal demands.

2. MINING IMPACTS

In the case of underground mining, a dislocation zone which transforms in the upward direction into flexure of the strata occurs directly above the mined area. A depression is formed on the surface. The shape of this depression depends on the structure of the deposit, mining parameters and the overlaying geology. Depending on the thickness of the overburden and the extent of the mined area, the surface impacts start to occur after around six months. At the end of approximately two years, around 80 % of the maximum anticipated depression has occurred, and the final condition is achieved after more or less five years.

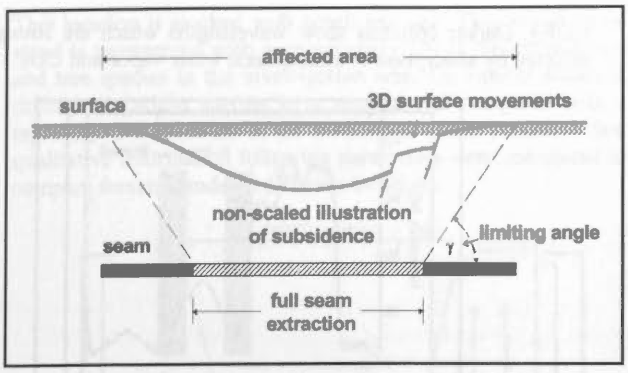


Figure 1: 3D-movements caused by mining activities

2.1 Surface Movements and Hydrological Parameter

The consequences of the extensive depression are changes in hydrological and associated ecological parameters. Such changes include alterations in the surface and underground catchment areas of individual watercourses and also in their geometry. The resultant changes affect the runoff, sedimentation and erosion conditions in watercourses. The impacts on the groundwater produce an extensive alteration in the spatial location and deformation of the groundwater-conducting strata ("aquifers"), and thus changes in the groundwater divides, as well as in directions and rates of flow. Locally, new isobars of the ground-water tables can occur. Changes to watercourses can be adjusted locally by means of constructional provisions, whereas the registration and evaluation of extensive changes in the groundwater situation is possible only by GIS-related modelling techniques based on accurate digital terrain models and ground-water models in particular. The intensity of the influence of a regarded surface is steered by its position in the relief and by the local characteristics of soil and vegetation (Drecker, et. al., 1995).

2.2 Monitoring

Depending on the mine site situation these investigations may cover an area of nearly 100 sqkm. Together with calculated and predicted surface movement data, resting upon current extraction plans, this information is of fundamental need for the individual monitoring concept for each mine site. During the mining process, extraction plan data and predicted surface movements have to be updated at several times. Therefore even the calculations of the resulting influences have to be updated. Considering the character of the effect forecast as a whole, the complete assessment area of the environmental compatibility survey must be monitored over a period of at least 20 years in order to verify determined forecasts.

The monitoring of terrestrial areas normally takes place at 2- to 5-year intervals. For this characteristic test sites were chosen, which describe the natural diversity of the area under investigation. Based on parameters derived from soil samples and different vegetation parameter an individual code is defined, which describe the ecological situation found on different areas. On these long-term monitored test sites, the biotope will be mapped regularly to determine the extent of the areas with undergoing change.

3. GIS Modelling

To monitor predicted movements high resolution digital terrain models (DTM) are generated at regular intervals based on photogrammetric surveys. A two-stage procedure was developed in respect to the accuracy of the DTM. The first step is to measure a regular grid, whereby the mesh size depends on the exactness of the approximation of the terrain. This DTM, initially roughly structured, is suitable for a first calibration of the ground water model. In order to reproduce the morphology of the terrain surface, the next stage involved the measurement of additional stations along breaklines. This results in a detailed DTM covering the whole area.

The description of the ground-water flow for theses areas can only be described with the help of numeric ground water models (GWM). To consider the pre-calculated amounts of subsidence in the GWM, the DTM updated by the subsidence values is used. This procedure leads to a direct entry of the influences caused by subsidence on the hydrologic parameters and enables the calculation of modifications of the ground-water situation. An important factor is the intensity of impairment regulated by properties of the local soil and vegetation data. To estimate changes of the dynamic system "subsidence-ground-water-vegetation", the ground-water recharge rate and isobars of the ground-water table are calculated in detail for a small-parcelled area of 110 sqkm. This was done for 1993, referring as a starting point of influence and for 2004 as a prediction. Additional data based on a current excavation plan with resulting subsidence data for 1998 and for 2000 were considered.

4. HYPERSPECTRAL SURVEY

Because of the large extent of the area which has to be monitored, actual research focuses on digital image processing techniques to derive information of the actual land-use and land-cover. Therefore, airborne data recording was done in August 1998 using the HYMAP™ imaging spectrometer. A second airborne survey was done in 2000 again with the HYMAP™. At the same time, detailed field surveys were carried out. Field measurements included spectroradiometrical data needed for the atmospheric correction of the hyperspectral data. Additional measurements of the leaf-area-index of specific plants were done in several subareas. The main emphasis of these investigations is the connection of calculated changes in periled areas and hyperspectral data features.

4.1 HYMAP™ Technical Parameter

The imaging spectrometer HYMAP™ has been developed for geological remote sensing applications. The system is built up in a modular manner which can be used in the most different fields of application. During the two flight campaigns in 1998 a subset of approx. 56 sqkm and in 2000 the whole area of 120 sqkm was recorded. The ground instaneous field of view (GIFOV) was 7m in 1998 and a 5m in 2000 respectively. The Table 1 below shows the general features of the HYMAP™ system (Cocks, et. al., 1999, Integrated Spectronics, Pty. Ltd. 2002).

spectral range	visible, near infrared, middle infrared
number of channels	100 - 200
spectral resolution	10 – 20 nm
GIFOV	2 – 10 m
aperture angle	60 - 70°
signal to noise ratio	> 500:1
operation altitude (above sea level)	6.000 – 15.000 ft.

Table 1: HYMAP™ technical features

4.2 Ground Reflectance Measurements

During the HYMAP™ data acquisition field spectra from selected representative targets were taken. Therefore a FieldSpec Pro FR from ASD Inc., operated by the GeoForschungsZentrum Potsdam (GFZ) was used. Additionally a Mark V from GER Inc., operated by the Federal Institute for Geosciences and Natural Resources (BGR) was used during the flight campaign in 2000.

	ASD FieldSpec FR	GER Mark V
Specral range	350 – 2500 nm	385 – 2548 nm
Spectral resolution	1,4 nm@350–1050nm 2 nm@1000–2500 nm	4 nm@385–1050nm 14 nm@1050–1785nm 23 nm@1785–2548 nm
scan time	0.1 seconds	60 seconds.

Table 2: Technical parameter of the ASD and GER portable field spectrometer

5. IMAGE PROCESSING

5.1 Preprocessing

The signal recorded by an airborne optical sensor is influenced by several components during the radiation's way from the sun to the ground and back to the sensor. The main influences include varying atmospheric conditions and variations of the sun-sensor-geometry. This modifies the reflectance values of the observed ground features. Especially for quantitative analyses and change detection applications with images from different times and sensors the atmospheric correction is an essential part of preprocessing and a prerequisite for the derivation of products needed for the subsequent image processing and analysis steps.

Atmospheric correction was done using the ATCOR-A and the newer version ATCOR-4 software package (Richter, 1996; Richter, 1997), which is commercially available. To model different atmospheric conditions it uses the look-up tables from MODTRAN (Berk, et. al., 2000). The atmosphere is modeled thereby a number of layers, whose layer boundaries can be parameterized. Standard atmosphere profiles describe different atmospheric conditions for different climate zones and seasons. The given values for a parameterization of different atmosphere layer (pressure, temperature, water vapor content) can be replaced by measured values. Additionally different aerosol models can be selected. Apart from a consideration of the spatial variability of atmospheric conditions differences in illumination within a scan line can be considered. Figure 2 shows the result of corrected spectra form the data flown in 2000 using ATCOR4. Darker columns show wavelengths which are strongly affected by absorption of atmospheric water vapor and CO_2.

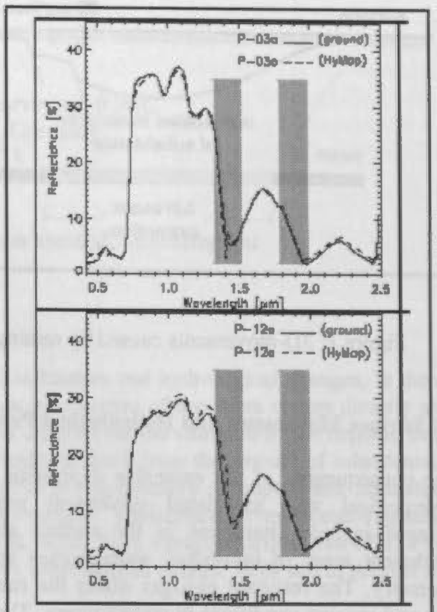


Figure 2: Comparison of ground reflectance spectra and corrected HyMap™ spectra.

The deviation of ground measured reflectance and retrieved reflectance is less then 3% in our case (Figure 2), except for the few HYMAP™ channels with noise problems and the regions of strong water vapour absorption.

5.2 Vegetation Analysis

First image processing steps of the HYMAP™ data lead to the detection of open water pits at different locations within the flown area. These locations occurred at places where groundwater table changes were calculated (Figure 3 at point 6.)

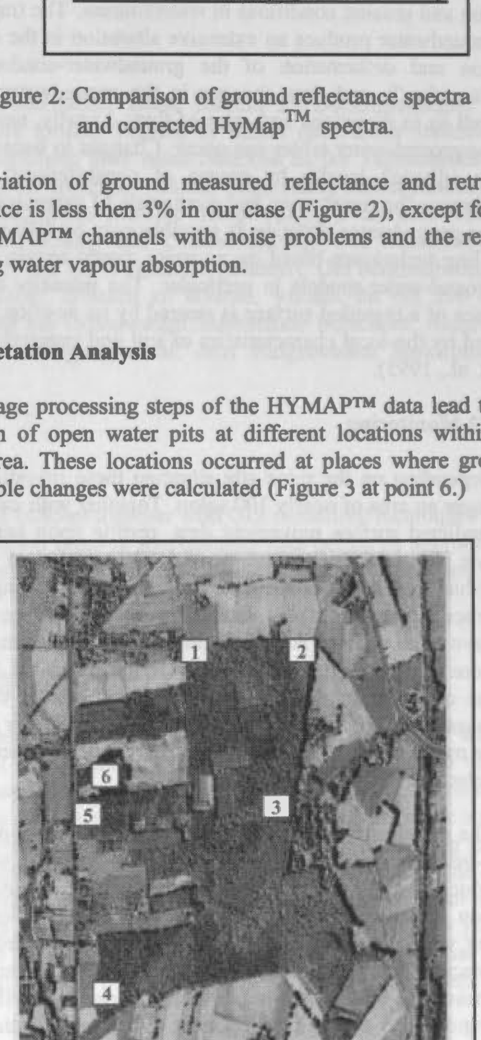


Figure 3: Subset with open water logging, point 6

This location is stocked with beech and oak. This mixed forest stand is represented with approximately 3% at the entire forest and tree species in the investigation area. On similar stands at different locations spectra were entered and compared with a representative spectrum of the influenced stand. To derive first qualitative information following parameters were calculated to compare forests stands on different locations.

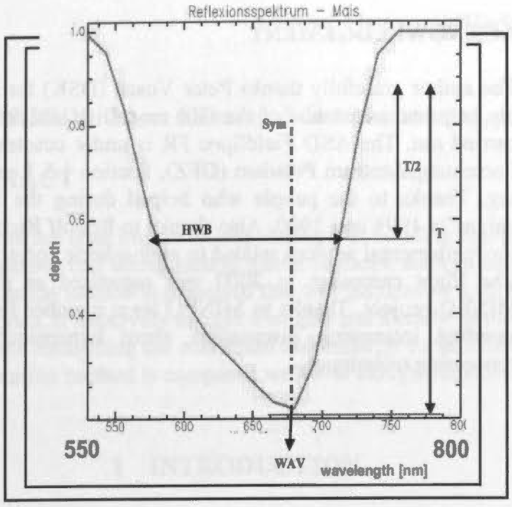


Figure 4: Schematic illustration of different parameter from normalized reflection spectra in the range between 550 and 800 nm.

This procedure was applied by Singhroy & Kruse (1991) to analyses reflection curves of different vegetation stocks and local mineralogical conditions. On the basis of laboratory analyses a connection between chlorophyll content of the plants and different concentration of different metal ions at the locations could be calculated. Further investigations (Clark, et .al., 1995; Koklay, et. al., 1999) improved these investigations and show that these analyses allow the discrimination of individual stands in terms of vitality. Table 3 shows the results for the affected location no. 6.

	loc 5	loc 6	Change
HWB	167.0	136.88	20 %
T	0.98	0.89	10 %
WAV	679.88	665.0	3 %
Sym	0.38	0.27	30 %

Table 3: Calculated Reflectance parameter from different locations

In a further step, it was verified if effects could also be detect on locations affected by a modification of the local soil water situation. The analyses are based upon different corn parcels mapped in the area under investigation, for which reference data were raised. This procedure was selected, since this type of vegetation was regarded as homogeneous, so that occurring modifications are not due to the natural variability, as for example with mixed woodland. Therefore a spectral library of different endmember associated with different stages of vitality of the corn fields was set up. Spectral unmixing procedures in conjunction with a subpixel analysis led to a significant improvement of the correlation between features of the hyperspectral data and the raised reference data. On the base of the correlation between derived endmember and measured leaf-area-index values (LAI) determinations of the LAI values for all corn fields were possible. The following Figures 5 and 6 show the calculated correlation between endmember values and LAI values measured within the field survey.

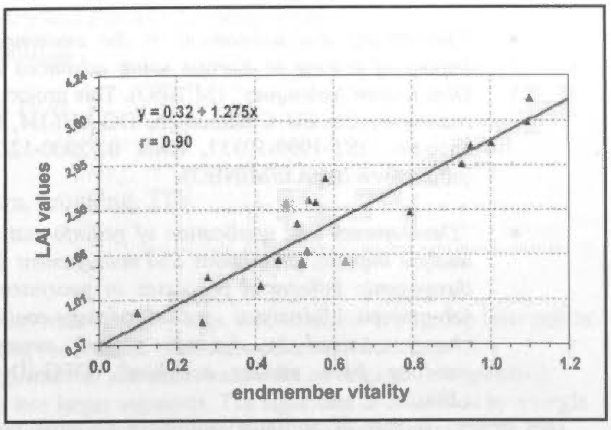


Figure 5: Relationship between measured LAI and derived endmember describing vital vegetation

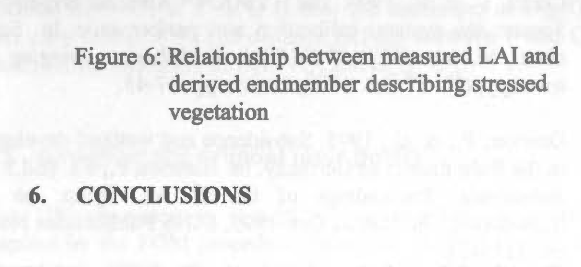


Figure 6: Relationship between measured LAI and derived endmember describing stressed vegetation

6. CONCLUSIONS

The high geometrical and spectral resolution of the data allows the detection of locations with mining affected changes of hydrological parameter and corresponding influences of different vegetations stands. At present there these investigations seem to be a methodological alternative to large scale field mapping procedures, which has a comparably potential of automation. Since the presented procedures for the analysis are based on derived parameters of the reflection signatures, the development of change detection routines for time series analyses are plausibly feasible. This is a crucial step to develop methods for efficient long-term monitoring procedures that become part of the GIS.

7. RESEARCH

Ongoing research is actually done within two projects:

- *"Monitoring and assessment in the environmental Impact of mining in Europe using advanced Earth Observation techniques"* (MINEO). This project is financed by the EU Commission, DG XIII-B4, reference-no.: IST-1999-10337, term: 01/2000-12/2002, (http://www.brgm.fr/MINEO).

- *"Development and application of geoinformation to analyse impacts, predictions and management of anthropogenic influenced processes in geosystems"* – sub-project: *"Analyses of landscape-ecological changes caused by dynamic surface movements resulting from mining activities"*, DFG-ID: BU 1283/3.

This project is part of an interdisciplinary research project, carried out together with the Institute of Computer Science of the University of Hamburg and is supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). The focus of this project is to model the dynamic impacts of the hydrological balance caused by subsidence within an object-oriented database. This database will be the core of a time- and scale-dependent simulation tool. The aim is the development of a 4D software environment.

REFERENCES

Berk, A, et. al., 2000. Modtran 4 User Manual, Air Force Research Laboratory, Hanscom, MA, USA.

Clark, R., et .al., 1995. Initial vegetation species and senescence/stress mapping in the San Louis Valley, Colorado using imaging spectrometry data. In: Posey, H. H.; e.a. (Ed.): Proceedings of the Summitville Forum '95. Colorado Geological Survey, Special Publication No. 38. Boulder, pp. 59-63.

Cocks, T., et .al., 1999. The HYMAP™ Airborne Hyperspectral Sensor: the system, calibration and performance. In: Schaepmann, M., e.a (Ed.): *1st EARSeL Workshop on Imaging Spectroscopy*, 6th –8th Oct. 1998, Zürich, pp. 37-45.

Drecker, P., et .al., 1995. Subsidence and wetland development in the Ruhr district of Germany. In: Barends, F.; e.a. (Ed.): *Land Subsidence*. Proceedings of the 5th Int. Symp. on Land Subsidence, The Hague, Oct. 1995, IAHS Publications No. 234, pp. 413–421.

HYMAP™ 2002: Integrated Spectronics Pty. Ltd., http://www.intspec.com (accessed 12. March 2002).

Koklay, R.; Clark, N., 1999. Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Linear Regression. In: *Remote Sensing of Environment*, 67(1), pp. 267–287.

Richter, R., 1996. A Spatially-Adaptive Fast Atmospheric Correction Algorithm. In: *International Journal of Remote Sensing*, 17(6), pp. 1201-1214.

Richter, R., 1997. Correction of atmospheric and topographic effects for high spatial resolution satellite imagery. In: *International Journal of Remote Sensing*, 18(5), pp. 1099- 1111.

Singhroy, V. H.; Kruse, F., 1991: Detection of metal stress in boreal forest species using the 0,67 µm chlorophyll absorption Band. In: *Proc. 8th Thematic Conference on Geologic Remote Sensing, Exploration, Engineering and Environment*, Denver, 29 April – 2 May 1991, pp. 361-372.

ACKNOWLEDGEMENT

The author gratefully thanks Peter Vosen (DSK) for considerably help and assistance of the GIS modelling which have been carried out. The ASD FieldSpec FR is under courtesy of GeoForschungsZentrum Potsdam (GFZ), Section 1.5 Remote Sensing. Thanks to the people who helped during the field campaigns in 1998 and 2000. Also thanks to Rudolf Richter (DLR) for fundamental advices related to atmospheric correction tasks. The flight campaign in 2000 was organized as part of the MINEO project. Thanks to MINEO team member fort the outstanding interesting discussions about hyperspectral image processing techniques.

SEGMENTATION OF TIN-STRUCTURED SURFACE MODELS

Ben Gorte
Technical University Delft, Photogrammetry and Remote Sensing Section
b.gorte@geo.tudelft.nl

Working Group IV/6

KEY WORDS: Laser scanning, 3-D reconstruction, segmentation, grouping, TIN

ABSTRACT

Segmentation is an important step during 3-D building reconstruction from laser altimetry data. The objective is to group laser points into segments that correspond to planar surfaces, such as facets of building roofs or the (flat) terrain between buildings. A segmentation method is presented that was inspired by a raster-based algorithm in literature, but works on original (triangulated) laser points. It iteratively merges triangles and already formed segments into larger segments. The algorithm is controlled by a single parameter controlling the maximum dissimilarity for adjacent segments such that merging them is still allowed. The resulting TIN segmentation method is compared with 3-D Hough transform.

1 INTRODUCTION

Laser altimetry is continuously gaining popularity. In the Netherlands, for example, a DTM of the entire country, called AHN (Actueel Hoogtebestand Nederland) has been produced by laser altimetry. Due to advancements in technology, the point density of laser altimetry data increases.

Besides converting laser data into 2.5-dimensional DSMs (Digital Surface Models) and DTMs (Digital Terrain Models) it is of particular interest to derive 3-D object models from high-density laser data. This involves *segmentation* or grouping: identifying groups of points that correspond to objects in the terrain, rather than treating points as independent (x,y,z) measurements. Also the location of the objects in 3-D has to be established.

Object modeling is best applied to man-made objects, such as buildings. In that case, the segmentation step in the object modeling process attempts to find planar regions (faces) in the point data set, corresponding to roof facets of buildings. Note that faces may be tilted in 3-D space.

Segmentation, for example by region growing or split-and-merge, is known from image analysis literature. These algorithms attempt to form segments of adjacent pixels that are similar, for example in having similar gray value or texture. When applied to the detection of planar regions in laser altimetry data (or other kinds of surface data), the goal is to create segments with constant slope, rather than with constant elevation. Therefore, when using `image processing' segmentation approaches in this case, not the surface data themselves would have to be segmented, but data sets derived by gradient filtering.

1.1 Laser data segmentation

In addition to image analysis segmentation methods, algorithms have been developed particularly for the purpose of segmenting laser data. [Geibel and Stilla, 2000] implemented four methods from literature and applied these to the same laser data set. Moreover, the article of [Geibel and Stilla, 2000] presents a new method termed *FOM procedure*, which is shown to perform better than the others. The algorithms have in common that they work in the raster domain. The original *(x,y,z)* coordinate measurements first have to be transformed into a raster, where a *x* and *y* coordinates defines a (row, column) position of a pixel, in which the corresponding *z* coordinate is stored as the pixel value.

The FOM procedure is based on iterative growing of similar areas from single pixels into large segments. In each step of the iteration, the procedure finds those two adjacent segments that are most similar to each other, and merges them. The process stops when no pair of sufficiently similar adjacent segments can be found anymore. Similarity is expressed in a distance measure D(P,Q) between adjacent segments P and Q (small distance means large similarity). Its definition is based on the Euclidean distance D(p, Q) in 3-D between a pixel p and a plane fitted through all pixels of segment Q. The directed distance D(P→Q) of P to Q is now the maximum of D(p,Q) over all pixels p in P. The distance D(P,Q) **between** P and Q, finally, is the minimum of D(P→Q) and D(Q→P).

1.2 Grouping the original laser points

The TIN segmentation algorithm presented here is greatly inspired by the FOM procedure. However, in the footsteps of Vosselman, 2000], who applies, for example, filtering on the original laser points, TIN segmentation directly uses triangulated laser points, instead of a rasterized version. Thereby, the data do not have to be discretised.

[Maas and Vosselman, 1999] describe laser altimetry segmentation, also based on the original points. They use 3-dimensional Hough transform.

Hough transforms were first developed to find lines and other parameterized shapes in point sets in 2-D images [Duda and Hart, 1972]. For each point the sets of possible parameter values are accumulated in a discretized parameter space. Consequently, parameter sets that obtain high accumulated values represent shapes passing through many of the input points. [Maas en Vosselman, 1999] apply the same principle to

3-D point sets, such as laser data sets. For each point p, the parameters a, b and c of planes $z=ax+by+c$ containing p are accumulated in a discretized (a,b,c) parameter space. Combinations of a, b and c that occur many times represent planes that pass through many points. [Goos, 2001] uses a (ρ,θ,λ) parameterization for a plane:

$$\rho = x \cos(\theta) \cos(\lambda) + y \cos(\theta) \sin(\lambda) + z \sin(\theta)$$

where ρ is the distance from the plane to the origin; θ and λ are the angles describing slope steepness and direction. The advantage of a (ρ,θ,λ) over an (a,b,c) parameter space is that the former is finite.

A drawback of the above-described Hough transforms is that they do not take adjacency or closeness (or any spatial relationship other than coplanarity) of the original points into account. A plane can be constructed through points that are very far away from each other and, therefore, belong to completely different objects.

This paper presents TIN segmentation as a method to segment point clouds, which does not suffer from this drawback. A comparison between the Hough transform segmentation and the algorithm proposed here will be given at the end of the paper.

2 TIN SEGMENTATION

The goal of TIN segmentation is to identify planar regions in a Digital Surface Model (DSM) that is represented in a TIN (triangular irregular network) data structure. The algorithm creates segments that are supposed to coincide with these regions. Each segment consists of a set of adjacent triangles in the original TIN, and each triangle is going to belong to one segment (complete partitioning). The program yields a list of segments with certain attributes including parameters (a,b,c) of the plane-equation in 3-D, as well as a segment id. It assigns to each triangle the id of the segment it belongs to.

Boundaries between segments are composed of arcs in the original TIN. Of course, not every arc in the TIN is going to belong to a segment boundary. Likewise, the breakpoints in the boundaries are original data points, but many data points are going to be inside a segment and not on a boundary.

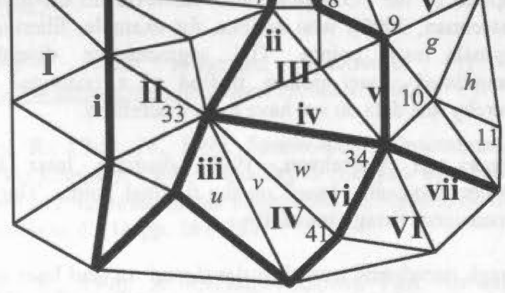


Figure 1: Segmented TIN (intermediate result).
1,2,3 : points; *a, b, c*: triangles; **I, II, III**: segments;
i. ii, iii: boundaries.

2.1 Data structure

The data structure that governs the algorithm can be described as a relational schema with four tables, for points (PNT), triangles (TRI), segments (SEG) and boundaries (BND). The point table has an ID-column and three columns x, y and z for the coordinates of each point. Each triangle in the triangle table contains de ID's of three points and a segment ID, denoting the segment that the triangle belongs to. The point table and the triangle table (with empty segment-ID's) are the input for the TIN segmentation algorithm. After the program finishes, the segment-ID column of the triangle table is filled, as are the segment and boundary tables. The segment table stores for each segment the number of participating points and triangles, the area of the segment and the parameters (a,b,c) of the plane through all points according to a least squares fit. The boundary table stores segment topology: the ID of the two segments to either side of the boundary. Also the distance between the two segments is stored, as well as the number of points on the boundary and its length.

The complete data structure consists of 4 relations (tables), for points (PNT), triangles (TRI), segments (SEG) and boundaries (BND). An example of these tables, corresponding to the situation in Figure 1., is shown in Figure 2.

2.2 Algorithm

While the program executes it iteratively merges triangles into segments, and segments into larger segments. During the initialization phase one segment is created for each triangle. The three vertices of the triangle uniquely define the three parameters of the plane containing the segment. Subsequently, all pairs of adjacent segments are analyzed. Between the two segments of each pair similarity (to be defined) is calculated. The most similar pair is selected and the two segments are merged into a single one. The new segment contains more than three points (the number can get arbitrarily large), which probably are not exactly in one plane. The parameters of the plane equation are obtained by a least square fit. The process is repeated until no pair of segments with sufficient (as specified by the user) similarity can be found anymore.

The algorithm needs to measure similarity between adjacent sub-segments. Clearly, similarity has to encompass coplanarity of sub-segments. Moreover, it must be ordered, to be able to select a maximum and compare with a user-specified threshold. Two similarity measures were considered: angle and distance. The angle between two segments is defined as the angle at which the two planes (whose parameters are stored with each segment) intersect. The distance between two segments is defined similarly to the definition in the FOM procedure. The distance from a point *p* to a segment *Q* is the (vertical or orthogonal) distance from *p* to the plane defined by the parameters in *Q*. The (directed) distance from a segment *P* to a segment *Q* is the maximum of the distances from each point in *P* to *Q*. Finally, the distance between two segments *P* and Q is defined as the minimum of the distances from *P* to *Q* and from *Q* to *P*. Of course, the relationship between *distance* and *similarity* is an inverse one: a small distance means that similarity is high. So far, the algorithm was applied to laser data with high point density (8 points per m^2) and standard deviation of 20-25 cm. In such a situation the orientations of triangles in one planar region may vary considerably, making 'angle' not a suitable similarity measure. Good results were obtained, however, using 'distance', where the threshold value can be easily related to the standard deviation in the point measurements.

PNT

ID	x	y	z
1			
2			
3			
4			
5			
6			
7			
	…	…	…
33			
34			
41			
	…	…	…

TRI

ID	p1	p2	p3	seg
a	1	6	5	I
b	1	2	6	I
c	2	7	6	II
d	2	8	7	II
e	3	9	8	V
f	3	4	9	V
g	4	10	9	V
	…	…	…	…
u				IV→III
v	33	34	41	IV→III
w				IV→III
	…	…	…	…

SEG

ID	#npnt	#tri	area	a	b	c	
I	9	7					
II	9	8					
III	7→10	6→9	!	!	!	!	
IV	5	3					removed
V	9	8					
VI	5	3					
	…	…	…	…	…	…	

BND

ID	seg1	seg2	#pnt	len	dist	
I	I	II	5		!	
ii	II	III	3→4	!	!	
iii	II	IV	2			removed
iv	III	IV	2			removed
v	III	V	4		!	
vi	IV→III	VI	3		!	
vii	V	VI	2			
	…	…	…		…	

Figure 2: Data structure and content corresponding to Fig. 1, and updates when removing boundary **iv** between segments **III** and **IV**. Attributes with ! are recalculated.

2.3 Post processing

On the basis of attributes in the segment-table, selections can be made as to keep only the relevant ones. Selection can be based on segment size (using number of point, number of triangles or area), or on segment slope (derived for the plane parameters).

A segmented TIN (either containing the complete set of segments or a selection) can be easily transformed into a VRML model, as is show in the next section. In the Figures shown there, the laser points were projected into the planes they belong to. Note that points on segment boundaries will be projected into two different planes, and the intersecting line of these planes usually does not pass through the boundary points.

3 EXAMPLES

Three examples will be shown. The first two contain buildings from the village of Wijhe in the East of the Netherlands. They were recorded with the helicopter-mounted FLI-MAP system of Fugro-Inpark. It measures 12,000 points per second. From an altitude of around 60m, a point density of 5 to 10 points per square meter can be reached.

The first example (Figures 3, 4 an 5) is a relatively simple house with a gable roof, an extension (also with a gable roof) and a dormer. Note that the house has roof facets on the short sides of the main wing. The Figures show that all these facets are detected. After selecting segments with slopes less than 150% only the roof and the ground segments remain.

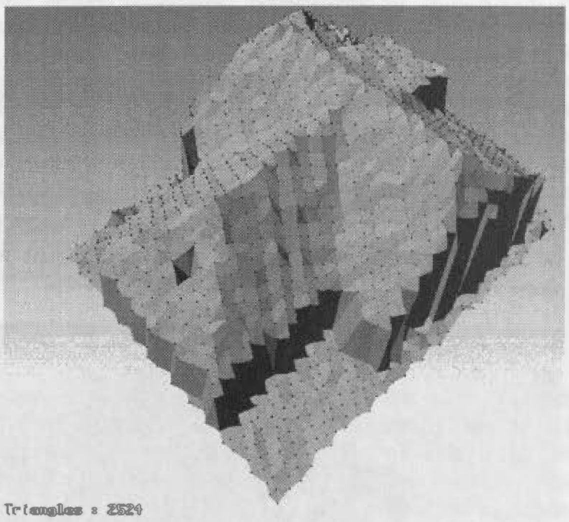


Figure 3: Simple house, triangulated

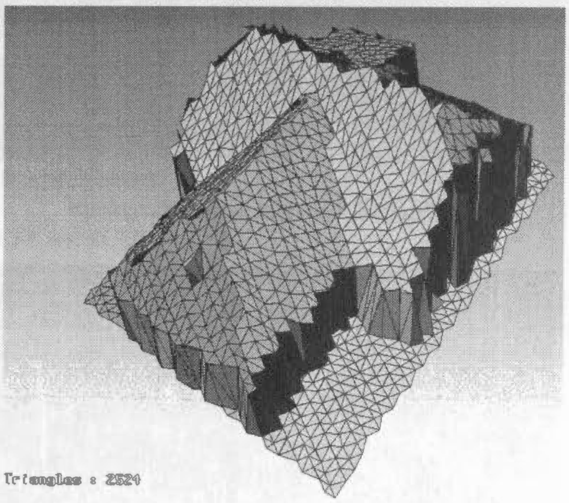


Figure 4: Simple house, segmented

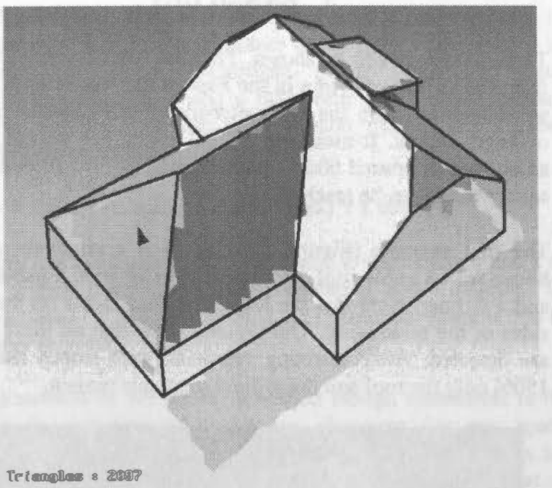


Figure 5: Simple house, selected segments. The lines were added manually for clarification.

The second example concerns a more complex scene with several buildings. Also some vegetation is present, which could be removed after segmentation by selecting segments with a slope less than 150% that contain more that 10 points.

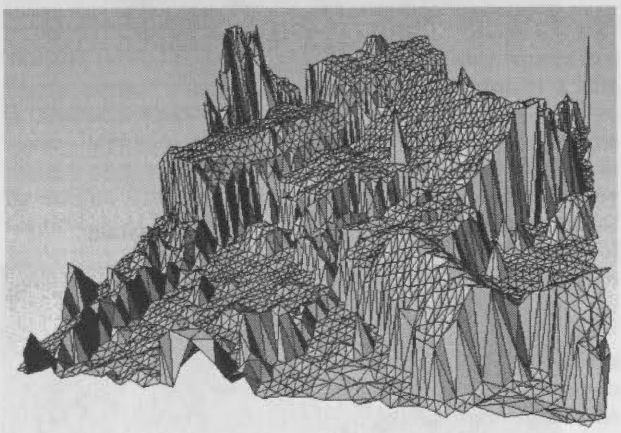


Figure 6: Complex scene, triangulated

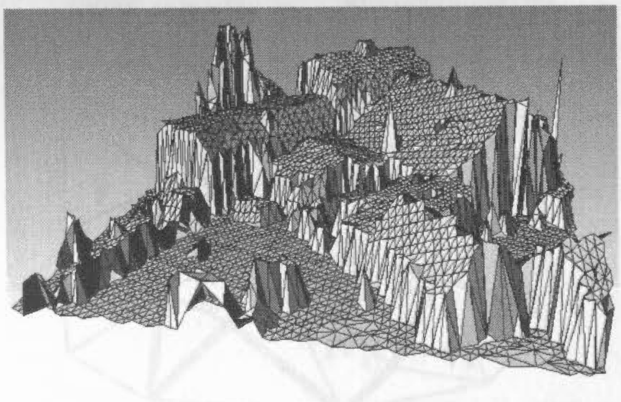


Figure 7: Complex scene, segmented

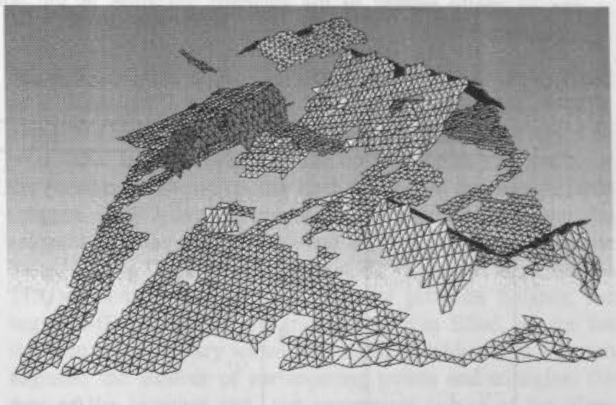


Figure 8: Complex scene, segment selection

The third example is a Toposys laser scan of the roof of the Aula of Delft Technical University. The Toposys scanner is characterized by a very small sampling distance between points within each scan line, and a much larger distance between adjacent scan lines.

The surface of the roof has the form of waves; a profile across the waves would have a sinusoidal shape. These are far from ideal circumstances for TIN segmentation. As a result, many points on segment boundaries are projected to quite different positions into the two segments.

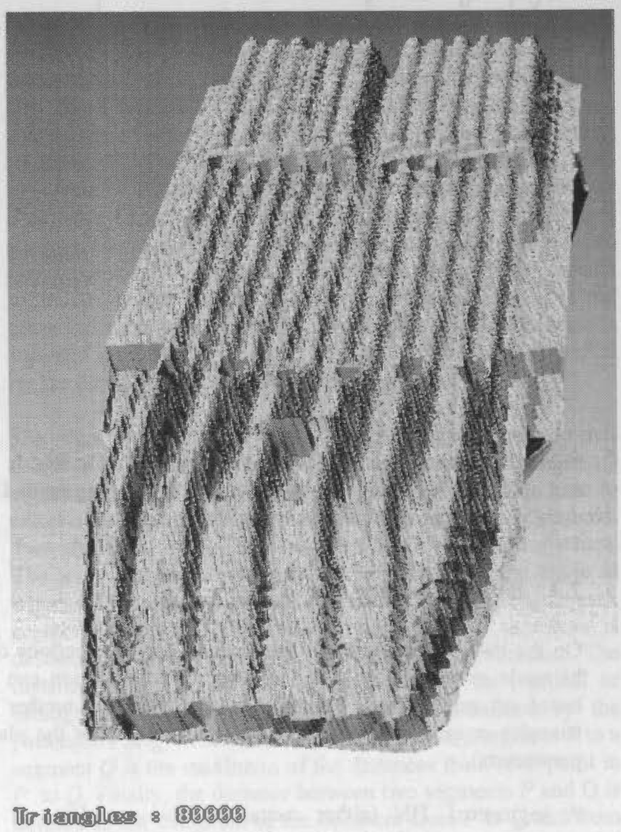


Figure 9: Aula roof

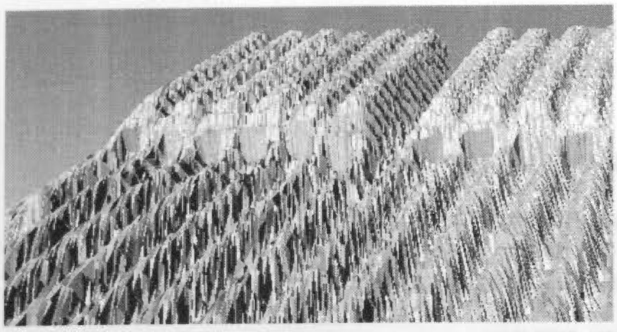


Figure 10: Aula roof detail

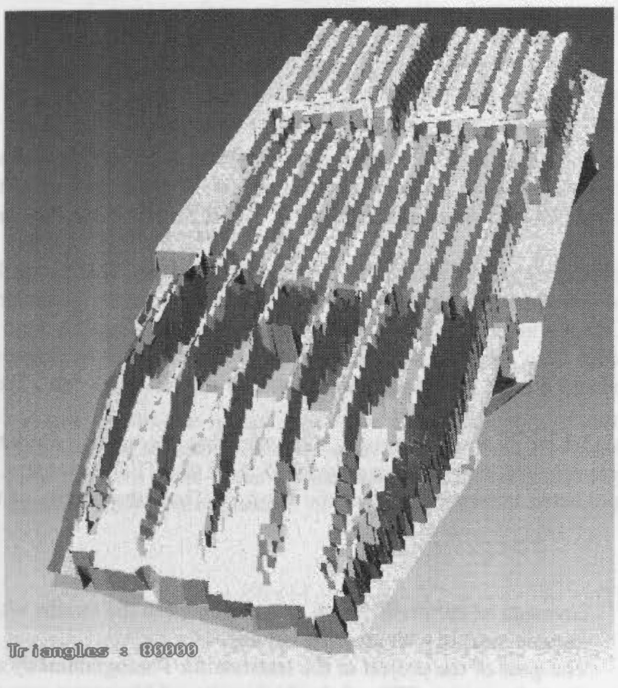


Figure 11: Aula roof segmented

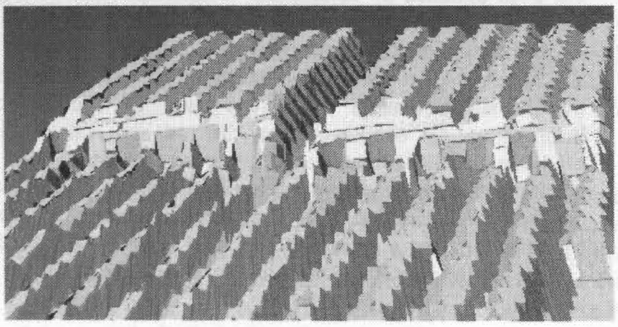


Figure 12: Aula roof segmented detail

4 EVALUATION

TIN segmentation is defined as a recipe and not in terms of the result. It is not guaranteed that the segmentation is optimal in some sense. Perhaps other segmentations are possible satisfying the same maximum-distance requirement and these segmentations may even be 'more optimal', for example have fewer segments. Many segmentation algorithms suffer from order dependency, meaning that the (relative) position of data-items (pixels) in the input data set influences the segmentation results. For example, quadtree-based segmentation (split-and-merge) is generally not invariant to rotations or shifts of the input data. TIN segmentation does not suffer from order dependency, as long as distances between different pairs of adjacent segments are not (exactly) equal.

Performance

Given the execution time of TIN segmentation for a certain data set, a twice as large data set requires twice as many boundary removals. Note that only one boundary is removed during each pass of the main loop of the algorithm. Each pass requires a single traversal of the boundary information to find the boundary with the smallest distance. The subsequent removal of that boundary requires that distances between the newly formed (combined) segment and all its neighbors are recalculated. Although this is a local operation (not depending on the size of the data set), the boundary table has to be traversed again to find these neighbors. The new distances are stored in the boundary table, which avoids that any distance will be calculated more than once. Despite this optimization, the execution time is still Order N-squared. Unfortunately this matters, given the size of laser altimetry data sets. Segmenting the aula roof (55,133 points and exactly 80,000 triangles into 1789 segments) took already 5 hours and 10 minutes using a C implementation on a 350 MHz Pentium PC under Linux.

5 CONCLUSIONS

Compared to 3-dimensional Hough transform, the most striking advantage of TIN segmentation is its simplicity, also towards the user. The whole process is controlled by only one parameter, the maximum allowable segment distance before merging, which is related to the measurement's standard deviation. At first sight, Hough transform also does not seem very complex, but in its original form is performs badly. This is mostly because Hough transform does not ensure that detected planes are actually connected. Experiments show that TIN segmentation performs at least as well as a heavily modified, rather complex 3-D Hough transform [Goos, 2001], which is controlled by five parameters. Further optimization of the TIN segmentation implementation is required to handle data sets with realisitic sizes.

REFERENCES

[Duda and Hart, 1972] Duda R.O., Hart P.E., Use of the Hough transform to detect lines and curves in pictures. *Comm. of ACM*, 15(1), 1972, pp. 11-15

[Geibel and Stilla, 2000] Geibel, R. and U. Stilla, Segmentation of laser altimetry data for building reconstruction: different procedures and comparison, *IAPRS*, vol. 33, part B3, pp. 326-334.

[Goos, 2001] M.I.M. Goos, 2001, Surface identification from irregularly distributed laser altimetry data. M.Sc thesis, Delft University, 69 pp.

[Maas and Vosselman, 1999] Maas, H.-G. and G. Vosselman, Two algorithms for extracting building models from raw laser altimetry data, *ISPRS J. of Photogrammetry and Remote Sensing*, vol. 54, pp. 153-163.

[Vosselman, 2000] G. Vosselman, Slope based filtering of laser altimetry data, *IAPRS*, vol. 33, part B3, pp. 935-941.

ANALYSIS OF SRTM DTM
METHODOLOGY AND PRACTICAL RESULTS

A. Koch, C. Heipke, P. Lohmann

Institute for Photogrammetry and GeoInformation (IPI), University of Hannover
Nienburger Straße 1, 30167 Hannover, Germany
(koch, heipke, lohmann)@ipi.uni-hannover.de

Commission IV, WG IV/6

KEY WORDS: DTM, DSM, Spatial Similarity Transformation, Systematic Errors

ABSTRACT:

In February 2000 the first mission using space-borne single-pass-interferometry was launched – the Shuttle Radar Topography Mission (SRTM). The goal of the mission was to survey the Earth surface and to generate a homogeneous elevation data set of the world with a grid spacing of 1 arcsec. Antennas with two different wavelengths were used: Beside the American SIR-C the German / Italian X-SAR system was on board.
This paper deals with the assessment of the Interferometric Terrain Elevation Data derived from the X-SAR system. These so called ITED-2 data were compared to reference data of higher quality of a well known test site in the South of Hannover (Trigonometric Points and Digital Terrain Model). The approach used is based on a spatial similarity transformation without using any kind of control point information. The algorithm matches the SRTM data onto the reference data in order to derive seven unknown parameters which describe horizontal and vertical shifts, rotations and a scale difference with respect to the reference data. These values describe potentially existing systematic errors.
The standard deviation of the SRTM ITED-2 was found to be ±3,3 m in open landscape, after applying the spatial similarity transformation. Maximum horizontal systematic shifts of 4-6 m were detected, representing only 20-25 % of the ITED-2 grid size. In summary, it can be stated that the results are much better than predicted before the start of the mission. Thus, the quality of the SRTM ITED-2 is indeed remarkable.

1. INTRODUCTION

Interferometric SAR (IfSAR) allows to obtain information about the third dimension of the Earth surface. Thus, the main product of this method is a Digital Terrain Model (DTM) or Digital Surface Model (DSM) in case of using a short wave system.
The SRTM mission (Werner, 2001, Rosen et al., 2001a) has been the first mission using space-borne single-pass interferometric SAR. The main antenna of the system was located inside the cargo bay of the Space Shuttle Endeavour. It transmited and received microwave pulses. The outboard or slave antenna was fixed at the tip of a 60 m long mast and acted as a receiver only. The mast realised the interferometric baseline. The system used two different wavelength. The American C-band system SIR-C operated with a wavelength of λ=6,0 cm, the wavelength of the German / Italian X-band system was λ=3,1 cm. In order to obtain a global coverage between 60 degrees north and 58 degrees south the Shuttle was flown at an altitude of 233 km and an inclination of 57 degrees. Because of the Earth rotation the Shuttle surveyed the ground strip after strip. With C-band it was possible to cover the Earth surface completely. The C-band interferometer operated in the so-called ScanSAR mode (Bamler, 1999).
The X-band antenna operated at a fixed depression angle of 38 degrees and a swath width of about 50 kilometres. The advantage of X-band is the higher relative vertical accuracy resulting from the shorter wavelength. The disadvantage of the X-band system used on board SRTM was the incomplete coverage of the Earth. There are gaps between the swaths which became smaller with growing latitude.
The goal of the project at the Institute for Photogrammetry and GeoInformation (IPI) of the University of Hannover was the assessment of the SRTM X-SAR data. The quality of the data has been obtained by comparison to reference data of higher quality in a well known test site.
Chapter 2 describes the algorithm used for the assessment of the SRTM data. The method is based on a spatial similarity transformation without using any kind of control point information. An overview about the test site and the used reference and SRTM data sets is given next (chapter 3) and the results of the validation process are presented in chapter 4 and 5. The paper concludes with a short summary. A similar investigation was e.g. carried out by Kleusberg & Klaedtke (1999) for airborne IfSAR data and by Rosen et al. (2001b) for SRTM.

2. ALGORITHM FOR MATCHING DIGITAL SURFACE MODELS

The developed algorithm (see also Koch & Heipke, 2001) is based on a spatial similarity transformation. The seven parameters of this transformation describe systematic errors of the SRTM data set detected within the test area. Remaining errors after having applied the similarity transformation can be considered as either local systematic errors or random errors.

2.1 Mathematical Model

Single points P (X,Y,Z) contain height information about a given area.
The points are combined to vectors:

$$\begin{aligned} G_1 &= \{P_{11} \quad P_{12} \quad \cdots \quad P_{1i} \cdots \quad P_{1n}\} \\ G_2 &= \{P_{21} \quad P_{22} \quad \cdots \quad P_{2j} \quad \cdots \quad P_{2m}\} \end{aligned} \tag{1}$$

The reference data set G_1 contains n regularly or irregularly distributed points. G_2 consists of m points, which describe the same physical surface as G_1. G_2 is the data set to be investigated. For the remainder of this paper we consider points P_{1i} and P_{2i} to have the same planimetric coordinates. If for a point P_{1i} no such corresponding point P_{2i} exists a priori (or vice versa) a height Z_{2i} must be interpolated from the other data set at the position X_{2i},Y_{2i} using e.g. a bilinear interpolation.

In the ideal case the following equation is fulfilled under the above mentioned assumptions:

$$Z_{1i}(X_{1i}, Y_{1i}) = Z_{2i}(X_{2i}, Y_{2i}) \tag{2}$$

Because of possible global systematic errors the two elevation data sets can be shifted and rotated against each other and can have different scale factors. Consequently a spatial similarity transformation is introduced:

$$Z_{1i}(X_{1i}, Y_{1i}) = Z0 + (1+m) \cdot \underline{r}_3 \cdot \underline{X}_{2i} \tag{3}$$

where

$$\begin{aligned} \begin{pmatrix} X_{1i} \\ Y_{1i} \end{pmatrix} &= \begin{pmatrix} X0 \\ Y0 \end{pmatrix} + (1+m) \cdot \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \end{pmatrix} \cdot \underline{X}_{2i} \\ \underline{X}_{2i}^T &= (X_{2i} \quad Y_{2i} \quad Z_{2i}) \\ \underline{R}^T &= (\underline{r}_1 \quad \underline{r}_2 \quad \underline{r}_3) \end{aligned} \tag{4}$$

In this way the points P_{2i} are transformed into the coordinate system of the reference data set by means of the seven parameters of the spatial similarity transformation. Z0 is the height translation, (1+m) is the scale. The vector $\underline{r}_3$ contains the rotations ω, φ and κ, it is the third row of the rotation matrix $\underline{R}$ of the spatial similarity transformation. Note that we use the rotation sequence ω, φ and κ. The centre point of rotation is the centre of the test site.
Z_{1i} on the left side of equation (3) is the corresponding height value of the reference data set with the planimetric coordinates X_{1i}, Y_{1i}. X_{1i} and Y_{1i} are computed according to equation (4) by transforming the coordinates X_{2i}, Y_{2i}, Z_{2i} of the investigated data set by means of the seven parameters. The vectors $\underline{r}_1$ and $\underline{r}_2$ are the first two rows of the rotation matrix $\underline{R}$. X0 and Y0 are the planimetric translations of the similarity transformation. In order to determine Z_{1i} in general the mentioned interpolation must be carried out, since we cannot assume that for the computed planimetric position (X_{1i},Y_{1i}) a value Z_{1i} exists in the reference data set.

2.2 Least squares adjustment

Equations (3) and (4) form the base of a least squares adjustment. We introduce the heights Z_{2i} (X_{2i}, Y_{2i}) as observations and consider the parameters of the similarity transformation as unknowns. The observations are assumed to be independent of each other and of equal accuracy resulting in an identity matrix for the covariance matrix of the observations. Equations (3) and (4) can then be formulated as observation equations, one for each height value Z_{2i}:

$$\begin{aligned} &v_i(Z_{2i}) = \\ &Z_{1i}\left(X0 + (1+m)\underline{r}_1\underline{X}_{2i}, Y0 + (1+m)\underline{r}_2\underline{X}_{2i}\right) - \\ &\left(Z0 + (1+m)\underline{r}_3\underline{X}_{2i}\right) \end{aligned} \tag{5}$$

This equation is the fundamental equation for calculating the unknown parameters of the spatial similarity transformation. Because of the non-linearity of equation (5) it has to be expanded into a Taylor series, and the unknowns are computed iteratively starting from approximate values. The design matrix of the least squares adjustment contains the partial derivatives of the observation equations with respect to the unknown transformation parameters. It should be noted that the explained method relies on height variations within the area under consideration, because – with the exception of $\partial v_i / \partial Z0$ - the partial derivatives all depend on $\partial Z / \partial X$ or $\partial Z / \partial Y$ (see equation 6).

$$\begin{aligned} \frac{\partial v_i}{\partial X0} &= \frac{\partial Z_{1i}}{\partial X_{1i}} \\ \frac{\partial v_i}{\partial Y0} &= \frac{\partial Z_{1i}}{\partial Y_{1i}} \\ \frac{\partial v_i}{\partial Z0} &= -1 \\ \frac{\partial v_i}{\partial \omega} &= \frac{\partial Z_{1i}}{\partial Y_{1i}} \cdot \frac{\partial Y_{1i}}{\partial \omega} - \frac{\partial Z'_{2i}}{\partial \omega} \\ \frac{\partial v_i}{\partial \varphi} &= \left(\frac{\partial Z_{1i}}{\partial X_{1i}} \cdot \frac{\partial X_{1i}}{\partial \varphi} + \frac{\partial Z_{1i}}{\partial Y_{1i}} \cdot \frac{\partial Y_{1i}}{\partial \varphi} \right) - \frac{\partial Z'_{2i}}{\partial \varphi} \\ \frac{\partial v_i}{\partial \kappa} &= \left(\frac{\partial Z_{1i}}{\partial X_{1i}} \cdot \frac{\partial X_{1i}}{\partial \kappa} + \frac{\partial Z_{1i}}{\partial Y_{1i}} \cdot \frac{\partial Y_{1i}}{\partial \kappa} \right) - \frac{\partial Z'_{2i}}{\partial \kappa} \\ \frac{\partial v_i}{\partial m} &= \left(\frac{\partial Z_{1i}}{\partial X_{1i}} \cdot \frac{\partial X_{1i}}{\partial m} + \frac{\partial Z_{1i}}{\partial Y_{1i}} \cdot \frac{\partial Y_{1i}}{\partial m} \right) - \frac{\partial Z'_{2i}}{\partial m} \end{aligned} \tag{6}$$

$Z'_{2i} = Z0 + (1+m)\underline{r}_3\underline{X}_{2i}$ is the transformed height value.

The unknown parameters are then computed according to the well-known equations of the least squares adjustment. The standard deviation of unit weight is identical to the standard deviation of the height differences after applying the transformation.

2.3 Special case of unknown shift Z0

The algorithm was implemented such that different unknown parameters can be introduced. If only a vertical shift Z0 is to be obtained the algorithm can be simplified. The obtained transformation parameter Z0 is identical to the mean value of the height differences between the two data sets. The observation equations have the following form:

$$v_i(Z_{2i}) = Z_{1i}(X_{2i},Y_{2i}) - (Z0 + Z_{2i}(X_{2i},Y_{2i})) \qquad (7)$$

The result of the algorithm introducing only a vertical shift Z0 corresponds with calculating a difference DTM.

3. TEST SITE AND USED DATA SETS

As mentioned before the aim of the project was the assessment of the SRTM X-SAR elevation data set. This task can be accomplished by comparing the data with reference data of a well known test site.

The test site of IPI is situated in the south of Hannover. The north-eastern part of the area is characterized by urban regions and flat terrain. The south-western part is more undulated, forest and agrigultural regions cover the area. The size of the test site is 50x50 km². The maximum height difference is about 450 m.

The accuracy of the reference data - provided by the surveying authority of Lower Saxony "Landesvermessung und Geobasisinformation Niedersachsen LGN" in our case - has to be at least one order of magnitude better than the SRTM data. The expected vertical accuracy of the SRTM data is several meters. Thus, highly accurate coordinates of Trigonometric Points (TP) and the Digital Terrain Model of LGN, the ATKIS DGM5, had to be used as reference data sets.

Trigonometric Points are part of the fundamental network of the surveying authorities of Germany. The planimetric coordinates are Gauß-Krüger coordinates, the heights are normal heights. The horizontal and vertical accuracy is 1-3 cm.

The DGM5 is a data set representing the terrain surface. The data consist of regularly distributed points with a grid spacing of 12,5 m. Together with morphological information the data represent a hybrid DTM. The vertical accuracy is about 0,5 m and depends on the terrain undulation. The DGM5 covered parts of the test site, altogether 4,7 million DGM5 points were available.

The SRTM ITED-2 data (Figure 1) represent the surface including vegetation and buildings because of using a short-wave X-band system. Thus, the data set is a Digital Surface Model in contrast to the reference Digital Terrain Model. The data are given in ellipsoidal coordinates referring to the geocentric ellipsoid WGS84. The grid spacing is 1 arcsec in both directions. For comparing the data sets the ITED-2 data were transformed into the coordinate system of the reference data set. A datum transformation between the two ellipsoids WGS84 and Bessel was carried out, additionally the ellipsoidal heights were corrected using geoid undulations. The geoid has an extent of 43-45 m in the test site. After these transformation steps the ITED-2 data are approximately distributed in a rectangular grid. The grid spacing in north-south direction is about 30 m, the grid spacing in east-west direction depends on the ellipsoidal latitude. The test area is situated at a latitude of about 52°. Therefore the grid spacing in east-west direction is about 20 m. The available data set consists of 5,5 million points.

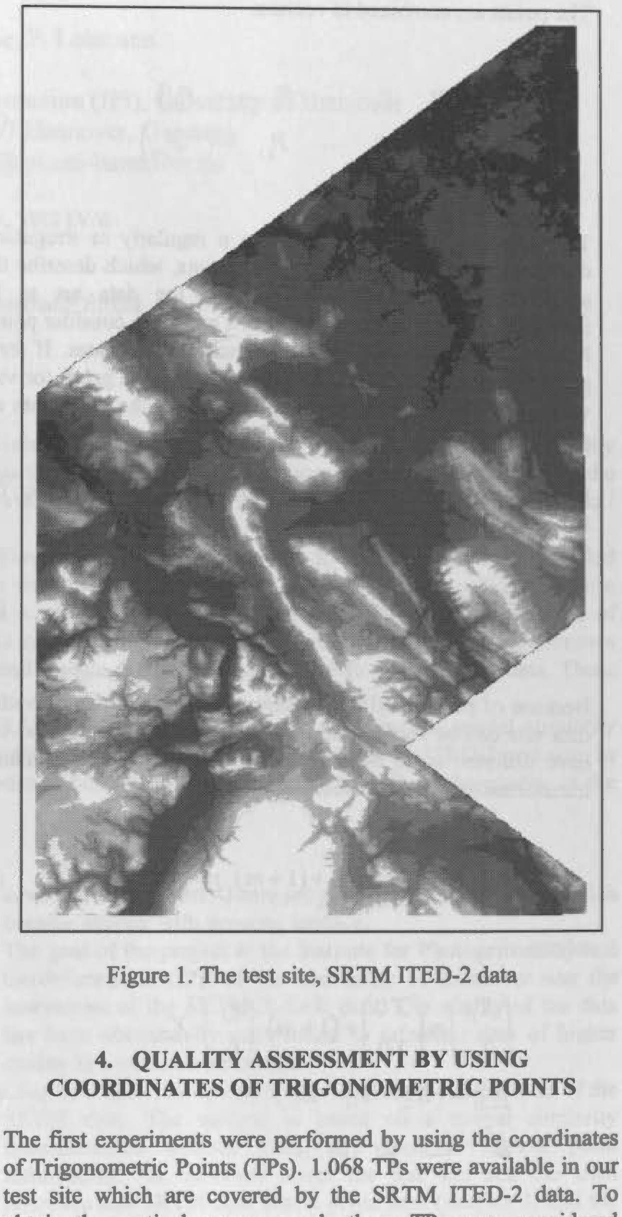


Figure 1. The test site, SRTM ITED-2 data

4. QUALITY ASSESSMENT BY USING COORDINATES OF TRIGONOMETRIC POINTS

The first experiments were performed by using the coordinates of Trigonometric Points (TPs). 1.068 TPs were available in our test site which are covered by the SRTM ITED-2 data. To obtain the vertical accuracy only those TPs were considered which lie clear of vegetation and buildings. To classify the TPs a Digital Landscape Model (DLM) of the surveying authority LGN (ATKIS Basis DLM) was used. The DLM is a two-dimensional representation of the topography. 368 points in urban regions, inside or near forests were excluded.

	Values
Z0 [m]	+3,18
s [m]	±4,27
s_{Z0} [m]	±2,86
n	700

Table 1. Quality measures using TPs

By means of the planimetric positions of the TPs the corresponding height values of the ITED-2 data set were obtained. The height differences were calculated and the quality measures were obtained (see equation 7). Any kind of planimetric systematic error or errors in rotation or scale were neglected. Table 1 shows the results.

The positive value Z0 means that the height level of the SRTM ITED-2 data is lower than the height level of the TPs. The data set contains a systematic vertical shift of +3,2 m. The standard deviation of the height differences s is ±4,3 m. After considering the mean value, the standard deviation s_{Z0} is ±2,9 m. These values confirm the high vertical accuracy of the SRTM ITED-2 data. Unfortunately, a conclusive reason for the shift cannot be given. A possible explanation can be derived from the calibration of the ITED-2 data used for the investigation. An error in the heights selected for the calibration directly influences the results of our study. While in many cases IfSAR orbits are processed and calibrated from coast to coast and thus the ocean serves as absolute reference, the orbit containing the data of the test site was shorter and had to be handled in a different way. If any buildings and vegetation existed in the areas selected for calibration and were not properly accounted for, the absolute ITED-2 heights would indeed come out too low, explaining the obtained results. It should be pointed out, however, that this hypothesis could not be further tested, because no additional information about the calibration sites and procedure was available.

5. QUALITY ASSESSMENT BY USING THE DIGITAL TERRAIN MODEL ATKIS DGM5

In a first step the height differences between corresponding values were calculated neglecting the influence of any kind of "terrain noise" (buildings, trees). In a second step only the height differences outside urban and forest regions were used. The algorithm described in chapter 2 was utilised introducing one and seven unknown transformation parameters.

5.1 Investigations with all DGM5 height values

About 1,2 million reference DTM points were available inside the test site. In contrast to the investigations before, the planimetric positions of the ITED-2 data were used to obtain the corresponding height values of the DGM5 using a bilinear interpolation. Then the height differences were calculated and the quality measures were derived (see equation 7). The following table 2 shows the results:

	Values
Z0 [m]	-2,63
s [m]	±9,08
s_{Z0} [m]	±8,68
n	1.234.815

Table 2. Quality measures using all DGM5 points

The sign of the mean value of the height differences Z0 is negative. It means that the height level of the ITED-2 data is higher than the level of the DGM5. This result is in contradiction to the obtained value of the investigations using the Trigonometric Points (see table 1). A possible reason is the influence of vegetation and buildings. Whereas in chapter 4 only points which are not influenced by terrain noise were used, here the height values are distributed over the complete test area, also across forests and urban regions.

Figure 2 shows the influence of terrain noise on the sign of local vertical systematic errors. The figure represents a positive vertical systematic error in open terrain. The height level of ITED-2 is lower than the height level of the reference data. Additionally, it can be seen that terrain noise (vegetation and buildings) increases the height level of the SRTM ITED-2 data and thus decreases the systematic vertical shift. The value Z0 can thus become negative.

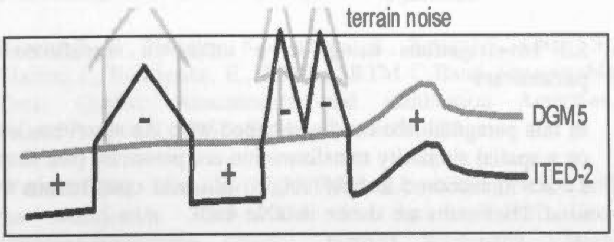


Figure 2. Height differences between DGM5 and ITED-2

The relative frequency distribution in figure 3 confirms the obtained results. The non-symmetric distribution is caused by objects lying above the terrain. Obviously the left negative part of the histogram represents these objects. Additionally in contrast to the calculated mean value Z0 (see table 2) the maximum of the histogram is in the positive part. This means that the vertical systematic error seems to be again positive. This result confirms those of chapter 4.

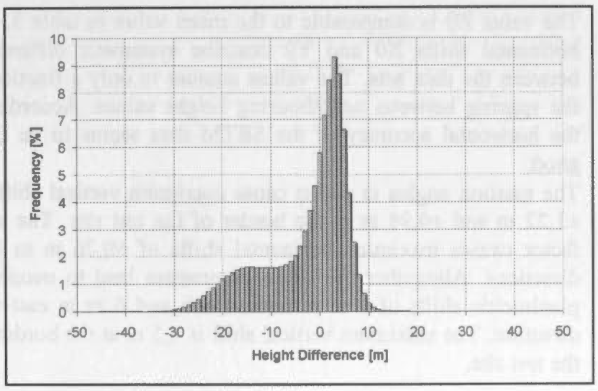


Figure 3. Relative frequency distribution of the height differences between DGM5 and ITED-2

5.2 Investigations with DGM5 height values in "open landscape"

Again the ATKIS-Basis DLM of LGN was used for the classification of the DGM5 height values. Because the mission was flown in February, the SRTM data also represent the terrain in agricultural fields. The vegetation heights are completely negligible. Values in forest and urban regions were excluded amounting to approximately 500.000 points or 40 % of the available DGM5 height values.

The results (see table 3) confirm the investigations before. The sign of the mean value Z0 is positive. The influence of large negative differences caused by terrain noise thus is significant.

As was to be expected the value Z0 does not correspond exactly to the value using the TPs (see chapter 4). There remains a difference caused by borders of forest and urban regions and other secondary effects. The standard deviations are nearly the same. The results show high quality of the SRTM data set.

	Values
Z0 [m]	+ 2,62
s [m]	± 4,32
s_{Z0} [m]	±3,44
n	669.685

Table 3. Quality measures using DGM5 points in "open landscape"

5.3 Investigations using seven unknown transformation parameters

In this paragraph the results obtained with the algorithm based on a spatial similarity transformation are presented (see chapter 2). As in section 5.2. only height values in open terrain were used. The results are shown in table 4.

	Value
X0 [m]	+0,60
Y0 [m]	-2,32
Z0 [m]	+2,28
ω [grad]	-0,003
φ [grad]	+0,002
κ [grad]	-0,007
m [ ]	+0,0000306

Table 4. Transformation parameters

The value Z0 is comparable to the mean value in table 3. The horizontal shifts X0 and Y0 describe systematic differences between the data sets. The values amount to only a fraction of the spacing between neighbouring height values. Accordingly the horizontal accuracy of the SRTM data seems to be quite good.

The rotation angles ω and φ cause maximum vertical shifts of ±1,22 m and ±0,98 m at the border of the test site. The scale factor causes maximum horizontal shifts of ±0,76 m in both directions. Altogether the seven parameters lead to maximum planimetric shifts of 4 m in north-south and 6 m in east-west direction. The maximum vertical shift is 4,5 m at the borders of the test site.

669.466 observations were used in the investigation. The standard deviation of the SRTM heights, comparable to s_{Z0} above, amounts to ±3,3 m, the vertical shift and the standard deviation of the heights reported in this section are rather close to those presented in section 5.2. (see table 3). Thus, the SRTM ITED-2 is free of systematic errors which can be modelled by the spatial similarity transformation.

6. FURTHER INVESTIGATIONS

As mentioned before the quality of the SRTM ITED-2 data, i.e. the order of magnitude of random and systematic errors, is influenced by terrain noise: The larger the percentage of regions containing vegetation and buildings the larger is the systematic vertical shift and the larger is the standard deviation of the remaining residuals.

In order to properly assess the SRTM data not only in open landscape but also in urban and forest areas, two sample sites with a size of 2x2 km² were analysed by obtaining a Digital Surface Model by using analytical photogrammetry. The photogrammetric measurements have a height accuracy of about 0,4 m.

The first sample site is situated in Hildesheim. The area is characterised by single and complex houses with gardens. The second test area is mainly characterised by deciduous forest. Additionally, agricultural fields and a freeway are in the centre of the area. It is a region in the south of Hannover.

		urban region	forest region
DGM5	Z0 [m]	-5,13	-13,19
-	s [m]	±8,46	±15,18
ITED-2	n	6.794	6.814
DSM	Z0 [m]	+1,05	+1,84
-	s [m]	±5,09	±4,57
ITED-2	n	6.612	6.675

Table 5. Quality measures in urban and forest regions, upper part: comparison between DGM5 and ITED-2, lower part: comparison between measured DSM and ITED-2

Table 5 shows the results of comparing the ITED-2 data with the reference DGM5 and the photogrammetrically obtained DSM. The spatial similarity transformation using the seven parameters of table 4 was applied before calculating the height differences. Thus, the expected value for Z0 under ideal conditions (DSM from SRTM and photogrammetry represent the same surface) is zero, and the standard deviation s should be close to the results obtained in chapter 4 and 5. The three upper rows represent the results from the comparison of DGM5 and ITED-2, the lower part shows the results of comparing the DSM of SRTM with the photogrammetrically measured DSM.

Considering the DGM5 both regions are characterised by a negative vertical shift Z0, i.e. the values are influenced by terrain noise. Additionally the standard deviations are very large.

Using the measured DSM the shift becomes positive. That means, that the ITED-2 lies significantly below the DSM (the values of table 4 and 5 must be added to obtain the complete shift). For the urban area this result can be explained with interpolation effects: there are probably some points on the ground influencing the result. In the forest area, an additional explanation may be the fact that the X-band signals somewhat penetrate into the canopy (remember that the mission was flown in February, thus the trees did not cover leaves). Also the standard deviations are larger than in the open landscape. As an overall result, it can be stated that in urban and forested areas, the quality of the ITED-2, while still meeting the predicted values, is somewhat poorer than in open landscape.

7. CONCLUSIONS

This paper contains the results of assessing the quality of the SRTM ITED-2 data. The algorithm used is based on a spatial similarity transformation without using any kind of control point information. The SRTM data was matched to a reference data set of better accuracy, the obtained seven transformation parameters describe potentially existing systematic errors of ITED-2.

First investigations were carried out by introducing just a vertical systematic shift. This procedure corresponds to the calculation of a difference DTM and yields the mean value and standard deviation of the height differences. Investigations using only regions without vegetation and buildings lead to a positive vertical shift of about 2,6 metres. Thus, the height level of the SRTM data seems to be too low. A possible explanation for this result is the fact that the ITED-2 DSM was calibrated over land rather than coastal waters.

The standard deviation of the SRTM ITED-2 was found to be ±3,3 m in open landscape, after applying the spatial similarity transformation. Maximum systematic shifts of 4-6 m were detected, representing only 20-25 % of the ITED-2 grid size. In summary, it can be stated that the results are much better than predicted before the start of the mission. Thus, the quality of the SRTM ITED-2 is indeed remarkable.

REFERENCES

Bamler, R., 1999. The SRTM Mission – A World-Wide 30 m Resolution DEM from SAR Interferometry in 11 Days. In. D. Fritsch, R. Spiller (Eds.), Photogrammetric Week ′99, Wichmann Verlag, Heidelberg, Germany, pp. 145-154.

Kleusberg, A., Klaedtke, H.-G., 1999. Accuracy assessment of a digital height model derived from airborne synthetic aperture radar measurements. In: D. Fritsch, R. Spiller (Eds.), Photogrammetric Week ′99, Wichmann Verlag, Heidelberg, Germany, pp. 139-143.

Koch, A., Heipke, C., 2001. Quality Assessment of Digital Surface Models derived from the Shuttle Radar Topography Mission (SRTM). Proceedings of IGARSS, Sydney, Australia, Supplement CD.

Rosen, P., Eineder, M., Rabus, B., Gurrola, E., Hensley, S., Knöpfle, W., Breit, H., Roth, A., Werner, M., 2001a. SRTM-Mission – Cross Comparison of X and C Band Data Properties. Proceedings of IGARSS, Sydney, Australia, CD.

Rosen, P. A., Hensley, S., Gurrola, E., Rogez, F., Chan, S., Martin, J., Rodriguez, E., 2001b. SRTM C-Band Topographic Data: Quality Assessments and Calibration Activities. Proceedings of IGARSS, Sydney, Australia, CD.

Werner, 2001. Status of the SRTM data processing: when will the world-wide 30m DTM data be available? Geo-Informationssysteme 12/2001. Herbert Wichmann Verlag, Hüthig GmbH & Co. KG. Heidelberg, pp. 6-10.

ACKNOWLEDGEMENT

This research has been partly supported by the German Ministry for Research and Education BMBF through the German Aerospace Research Center DLR under contract no. 50EE9927. We are also thankful to LGN for providing the reference data.

INTERACTIVE ANALYSIS FOR 3D-GIS TOOLS

Désirée Hilbring
Institute for Photogrammetrie and Remote Sensing
University of Karlsruhe, Germany
hilbring@ipf.uni-karlsruhe.de

Commission IV

KEY WORDS: GIS, interactivity, virtual reality

ABSTRACT

Interaction is a very important part of software development. This article explains different interaction techniques for universal GIS applications. Further it investigates these techniques for the dedication in a three dimensional GIS environment. It is possible to define two main categories: universal interaction types and application specific interaction types. The interaction functions belonging to the first category are necessary in almost every 3D GIS application, while the interaction functions belonging to the second category envelope a wide range of specific functions. Thus they are hard to generalise and must be developed for each specific case. This article will show examples of the first category in a 3D software development system and examples for the second category, featuring a specific 3D GIS application.

KURZFASSUNG

Interaktion ist ein wichtiger Teil der Softwareentwicklung. Dieser Artikel erklärt verschiedene Interaktionstechniken für allgemeine GIS-Anwendungen. Diese Techniken werden weiterhin für den Einsatz in einer dreidimensionalen GIS Umgebung untersucht. Es ist möglich, zwei Hauptkategorien zu unterscheiden: allgemeine Interactionsarten und anwendungsspezifische Interaktionsarten. Die Interaktionsfunktionen der ersten Kategorie sind in fast jeder 3D GIS Anwendung notwendig, während die Funktionen der zweiten Kategorie ein weites Feld spezifischer Funktionen umfassen. Die anwendungsspezifischen Funktionen sind schwer zu verallgemeinern und müssen für jeden einzelnen Fall entwickelt werden. Dieser Artikel zeigt Beispiele für die erste Kategorie anhand eines 3D Software Entwicklungssystems und Beispiele für die zweite Kategorie anhand einer spezifischen 3D GIS Anwendung.

1. INTRODUCTION

Besides the visualisation of 2D data the visualisation of 3D data has continued to grow in importance. One reason is, that it is possible to display virtual scenarios in an appearance similar to the real world. Another reason is, that the third dimension helps in displaying complex spatial information for visual analysis tasks. It is possible to present data in a way, which combines realistic world objects (e.g. a terrain shape) with abstract objects (e.g. a 2D topographic map). The familiar look of well known objects helps the viewer to concentrate on the analysis problem.

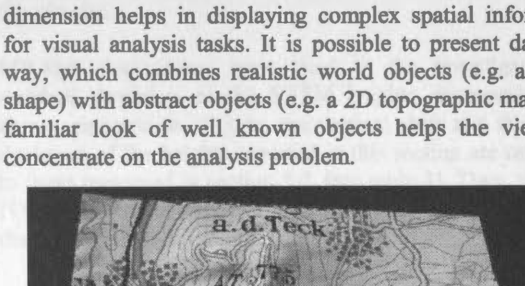


Figure 1. Digital elevation model with topographic map texture

To create such a scenario it is necessary to combine a GIS with a virtual reality for the visualisation of 3D data. But the sole visualisation of the given data is often not enough. The generation of a fundamental analysis from the given data needs interactive techniques to extract the bits of information needed to solve the problem.

For the intervention of the user into the virtual world some kind of devices are needed. Devices like Head-Mounted Displays, data gloves etc. have been popularised in the entertainment media and are recognised by the public as symbols of virtual reality. Here we are at the high end of experiencing a virtual world. On the other hand, an example definition of virtual world is: "Virtual reality is the simulation of a real or imagined environment that can be experienced visually in the three dimensions of width, height, and depth" [Whatis]. That means it is also possible to create a virtual reality with a "low" personal computer. That is the kind of virtual world this article is referencing to. The user views the virtual world at the monitor. To intervene into the content of the displayed virtual world it is possible to restrict the interaction devices to the well known mouse and keyboard.

2. INTERACTION

The intervention of the user in software applications is not restricted to the intervention with 3D content. Commonly the intervention is called user interaction. "Interaction is when the imagery changes in response to user action" [Interaction]. Thus user interaction is a well known task in universal software development. This article concentrates on typical

GIS interaction types applied by the user to intervene into the scene. This envelopes a wide range of activities. We want to name and categorise some GIS interactions into different types, valid for both 2D and 3D GIS scenes.

2.1 Classification of Universal Interaction Types

One classical interaction type is the navigation through the GIS scene. Commonly addressed as navigation is here the change of the view due to user interactions. Typical GIS interactions are:

- Panning
- Zooming
- Zoom to extent

To analyse the scene content the user often needs to select one object for further activities. This is another interaction category mostly referred as "picking". Picking includes the click of the user on an object and the resulting selection. The object has then focus for further action. The following actions are performed directly on the object and can include:

- Moving the object
- Change the shape of the object
- Change the appearance of the object
- Display further information about the object
- ...

Picking is an interaction category, which is performed on special object content.

Another interaction type is the appearance manipulation of the scene. In this category all interaction functions changing the entire scene are included:

- Changing the scale
- Changing the background color
- Changing the appearance of all objects in the scene
- ...

These categories are still missing the really interesting part, in which all kind of interaction techniques can be imaginable for developing analysis techniques for problem solving. These interactions envelope a wide range themselves and can be very specific.

Thus there are some common interactions, useful for every displayed image scene and otherwise there are application specific interactions for problem solving.

2.2 Common 3D-GIS Interactions

Until now, only universal GIS interaction categories were discussed. In the following this article will have a closer look at common 3D GIS interaction types.

Navigation through the virtual world is necessary for exploring 3D content from several viewing positions. The 3D-object content of a virtual world is displayed on a 2D device, the monitor. Thus additionally to panning and zooming the user needs the possibility to rotate the object content. In the cases described by this article the user has only a 2D device to interact with the virtual reality, the mouse. But it is possible to use the normal mouse for 3D navigation. One technique is to allocate each mouse button with one navigation type: one for zoom, one for pan and one for rotation. The rotation is managed by the combination of a click into the 3D scene, which defines a starting point, with the direction of the mouse movement. This technique is the default navigation technique for simple Java3D applications. Other techniques are including virtual input devices for the navigation functions into the program.

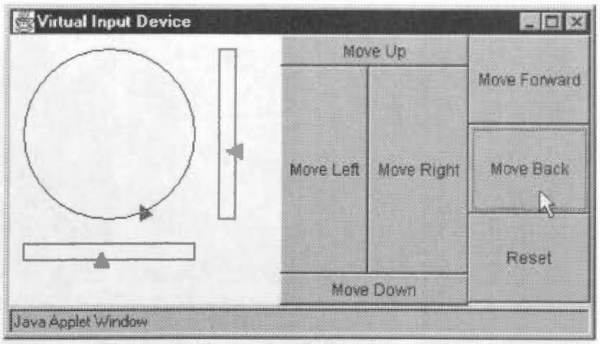


Figure 2. Example of virtual input device

The realisation of the navigation in the virtual world is independent from the transmission way of the user input. It is possible to fulfil that in two different ways. The first approach is changing the actual co-ordinates of the 3D object content, every time the user performs a navigation task, while the second approach actually changes the position of the viewing platform, from which the virtual observer views the 3D scene. The first approach is often used in CAD-Systems and the second approach is often used for large virtual realities including terrain data. Thus the second one is the choice for an interaction type in a 3D GIS.

One big issue during the navigation in a virtual environment is the orientation of the user. It often happens that the viewing platform gets lost in space and the user only sees dark space. For this case the common 2D interaction "Zoom to extent" is helpful, but the orientation in the virtual world can be improved. The following images show an overview window in which the entire 3D scene is displayed. The user can change the view, from bottom to side, so that it is possible to extract the position and height of the viewing platform (represented by the camera) in the virtual universe.

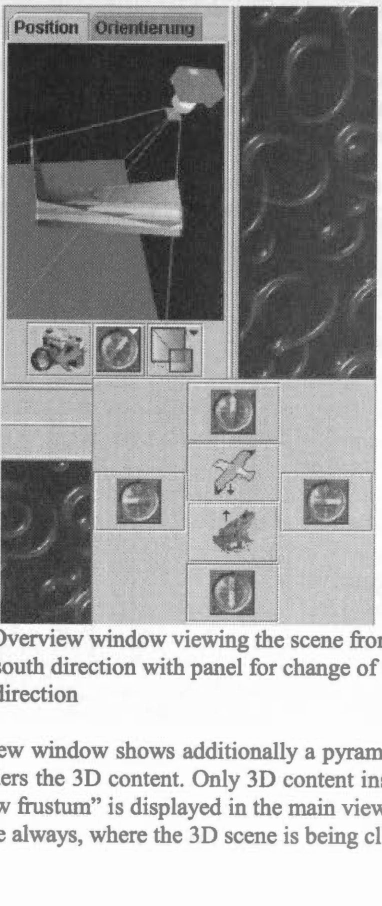


Figure 3. Overview window viewing the scene from the south direction with panel for change of viewing direction

The overview window shows additionally a pyramid frustum which borders the 3D content. Only 3D content inside the so called "view frustum" is displayed in the main view. Thus the user can see always, where the 3D scene is being clipped.

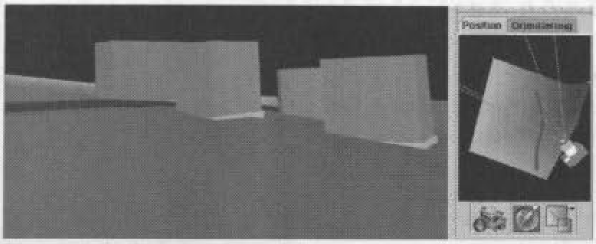


Figure 4. Left: main scene; right: overview window viewing the scene from top with camera south east of buildings in main scene

Picking in a 3D GIS is very similar to picking in a 2D GIS. The mode of operation is the same, while the implementation is different. To find the sighted object in the virtual scene a ray from the users eye is sent into the scene and selects the first, or all objects, intersected by the ray.

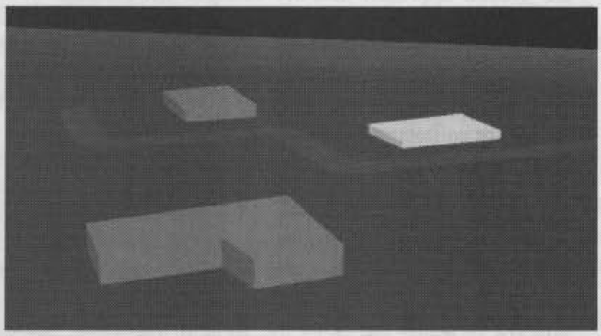


Figure 5. The picked object is highlighted with yellow.

The appearance manipulation for a 3D GIS envelopes activities well known from 2D GIS, and additionally contains classic 3D graphic approaches:

- Objects in a 3D scene can be illuminated with different 3D graphic techniques, like "Gouraud Shading" or "Flat shading". Sometimes the original colour can be useful as well for identifying objects in the scene.
- Important is also, that it is possible to change the position of the sun, ensuring best lighting for any scene.
- Another technique, which changes the appearance of objects, is the scaling of co-ordinate axes. Considering that a 3D GIS displays real world data (for example digital elevation models or geological layers), where the analysis of topographical characteristics is necessary, the scaling of the height axis disband objects and helps to analyse the 3D scene.

Common 3D GIS Interaction techniques are necessary to provide the user with a flexible environment suitable for their specific needs. A number of these techniques are realised in the Geographic Information System GISterm. The project is included in the "Umweltinformationssystem Baden-Württemberg" an environmental GIS project and funded by the "Landesanstalt für Umweltschutz" a local authority concerned with environmental protection [GISterm]. GISterm is developed entirely in Java and is extended by the 3D Service, a service for developing 3D GIS applications. The 3D Service is realised with the means of Java 3D [Java 3D]. The purpose of the 3D Service is to create a virtual reality, include it into GISterm and provide means and methods for the implementation of specific 3D GIS applications. Among other things the 3D Service includes the following interaction techniques:

- 3D GIS navigation featuring rotation with a 2D mouse using the Java 3D approach
- Orientation window for the orientation of the user in the scene
- Manipulating appearance aspects of the entire view with the help of a property dialog
- Picking 3D objects for displaying further attribute information

2.3 Application Specific Interactions

Application specific interactions are hard to categorise. They are including all kinds of activities, like:

- Creating "Fly Throughs"
- Analysing specific parts of a 3D scene
- Postprocessing of created object content
- ...

In the following some examples for specific application interactions are given, featuring GeoPro^{3D}, a special 3D GIS application, realised with the means of the 3D-Service as another extension of GISterm, for visualising and analysing the ground water table and hydrogeological profiles [GeoPro^{3D}].

2.4 Illustration for Application Specific Interactions in an Example 3D-GIS Application GeoPro^{3D}

In contrast to the 3D-Service GeoPro^{3D} implements only specific interactions. It is creating a 3D scene with object content from data stored in a ground water database [GWDB]. The user loads the ground water measurement points into GISterm for the region to be analysed. Then he specifies the relevant points for the calculation of the water table and indicates a time interval for the calculation of the ground water table. With this information GeoPro^{3D} triangulates three ground water layers: minimum ground water layer, medium ground water layer and maximum ground water layer. Each measurement point includes the terrain height, from which GeoPro^{3D} can produce the terrain layer. Some measurement points in the ground water database additionally include hydrogeological stratum information GeoPro^{3D} is using for creating hydrogeological layers. The purpose of GeoPro^{3D} is to detect conflicts between these created layers and proposed construction sites. GeoPro^{3D} can handle two different kind of construction sites: buildings (single or groups of buildings) and routes (roads or railway routes). The user digitises the shapes of the objects in the 2D map and gives further information about the dimension, for creating 3D objects. The following picture shows a virtual reality scene created with GeoPro^{3D}.

For the specific analysis of this example scene the user needs application specific interactions.

The main task for the user is to find out, where the construction sites are cutting the ground water table. Thus the clipping from construction sites from the overall scene is one example for application specific interaction. The function combines the universal interaction "picking" with the clipping of the picked construction site from the scene, so that the observer can analyse the near range of the construction site.

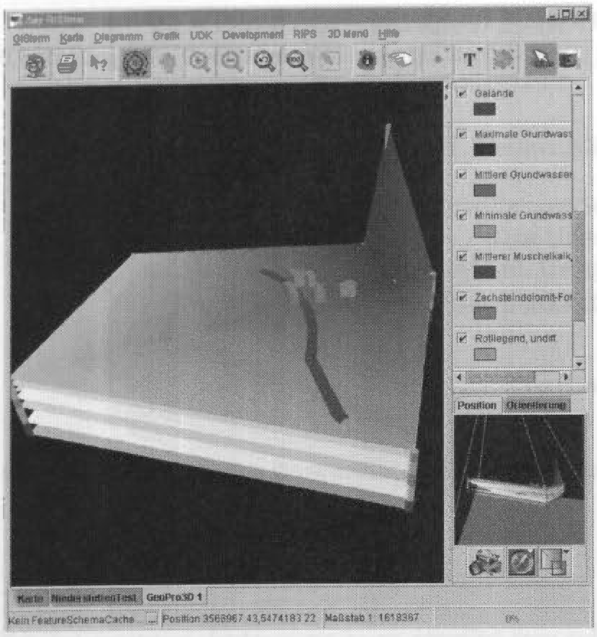

GIStern
Karte
Diagramm
Grafik
UDK
Development
RIPS
3D Menü
Hilfe
☑ Gelände
☑ Maximale Grundwass
☑ Mittlere Grundwasser
☑ Minimale Grundwass
☑ Mittlerer Muschelkalk,
☑ Zechsteindolomit-For
☑ Rotliegend, undiff
Position
Orientierung
Karte
NiedersachsenTest
GeoPro3D 1
Kein FeatureSchemaCache
Position 3566967 43,5474183 22
Maßstab 1: 1618367
0%

Figure 6. Virtual reality created by GeoPro^{3D} and included into GISterm with terrain layer, ground water layers, hydrogeological layers and construction sites

The function needs to distinguish between buildings, which can be clipped the way shown in the following picture and routes needing further analysis functions.

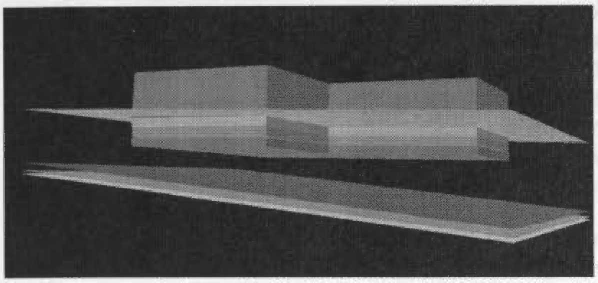


Figure 7. The overground parts of the building (red) are separated by the terrain (green) from the underground parts (grey). There is no conflict between the building and the ground water table (blue).

To analyse routes, the best way is to analyse 2D cross direction profiles along the roads. The user positions a slice through the 3D scene and moves the slice along the run of the street. This application specific interaction will be implemented in future, using the basic specific interaction of creating 2D hydrogeological cross direction slices.

For creating universal 2D cross direction slices the user needs to specify the place of the slice. In activating this functionality the 3D scene is clipped in the middle by a vertical slice. This slice can be positioned using the cursor keys.

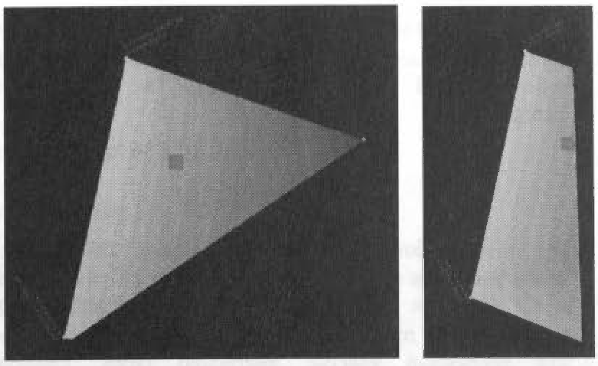


Figure 8. Left: 3D scene viewed from top and cut by vertical slice; right: the slice was moved with the cursor keys to the target place.

If the user has reached the appropriate place of the slice he can activate the cross direction profile function and sees the 2D hydrogeological profile in the display window.

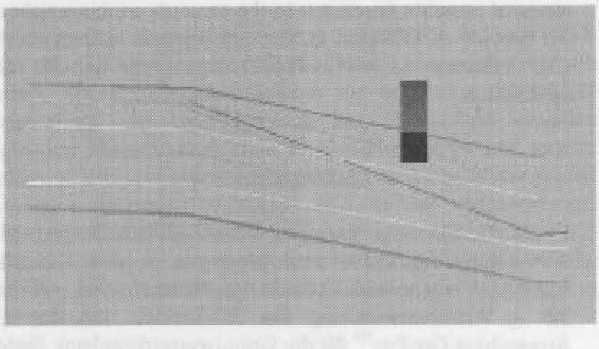


Figure 9. Hydrogeological cross profile from the position in figure 8.

Sometimes the run of the hydrogeological layers GeoPro^{3D} is creating includes some errors (measurement points lying within a layer, not including the hydrogeological strata of the layer).

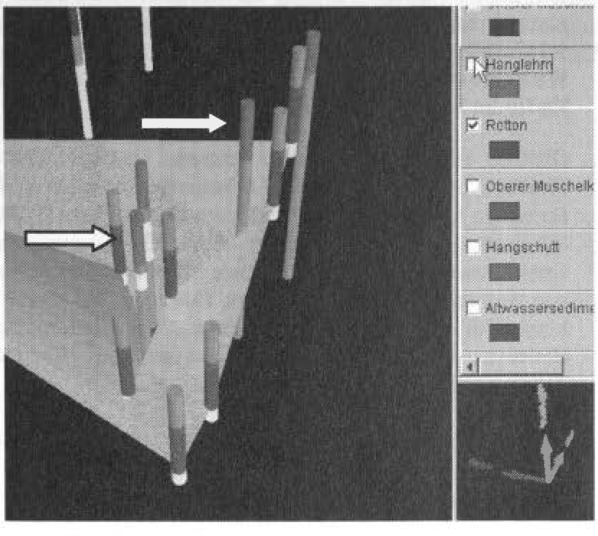

☐ Hanglehm
☑ Rotton
☐ Oberer Muschelk
☐ Hangschutt
☐ Altwassersedime

Figure 10. Not every borehole includes the "Rotton" stratum

This is another task which will be solved by a specific interactive technique, the post-processing of the hydrogeological layers. The user will be able to select points, which should not be included in the triangulation of the hydrogeological layer, and afterwards the layer will be recreated.

3. CONCLUSION

The investigation in this article shows, that different interaction types are necessary for the user oriented analysis in 3D GIS tools. It is possible to distinguish the interaction types into two main categories: universal interaction types and application specific interaction types. Universal interaction types include navigation and orientation, picking, and changing the appearance while application specific interaction types are hard to be categorised and envelope a wide range of possibilities. Thus development systems for 3D GIS software should include methods for all common universal 3D GIS interactions, because they will be needed in almost all special 3D GIS applications. The realisation of application specific interactions must be done in the special 3D GIS application itself, because they envelope such a wide range of possible functions. In the example of this article the 3D-Service of GISterm is the development system for 3D GIS software while GeoPro3D represents the 3D GIS application.

REFERENCES

GeoPro^{3D}: Hilbring, Veszelka, Wiesel (2001): Projekt AJA, Anwendung JAVA-basierter Lösungen in den Bereichen Umwelt, Verkehr und Verwaltung, Phase II 2001, GISterm 3D – Weiterentwicklung des 3D-Service und der 3D-Anwendung GeoPro^{3D} für die Grundwasserdatenbank Baden-Württemberg, S.133-152 in Mayer-Föll, Keitel, Geiger (Hrsg.), Wissenschaftliche Berichte, FZKA 6700, Forschungszentrum Karlsruhe Technik und Umwelt

GISterm: Hofmann, Hilbring, Veszelka, Wiesel (2000): Projekt AJA, Anwendung JAVA-basierter Lösungen in den Bereichen Umwelt, Verkehr und Verwaltung, Phase I 2000, GISterm – Weiterentwicklung des flexiblen Frameworks zur Analyse und Visualisierung von raumbezogenen Daten, S.147-169 in Mayer-Föll, Keitel, Geiger (Hrsg.), Wissenschaftliche Berichte, FZKA 6565, Forschungszentrum Karlsruhe Technik und Umwelt

GWDB: Schmid, Schmieder, Stumpp, Usländer (2001): Projekt AJA, Anwendung JAVA-basierter Lösungen in den Bereichen Umwelt, Verkehr und Verwaltung, Phase II 2001, Informationsaufbereitung in der WAABIS-Fachanwendung Grundwasser in Baden-Württemberg unter Berücksichtigung von Anforderungen des Freistaates Thüringen, S.153-165 in Mayer-Föll, Keitel, Geiger (Hrsg.), Wissenschaftliche Berichte, FZKA 6700, Forschungszentrum Karlsruhe Technik und Umwelt

Interaction: Jack S. Gundrum, Java3D Notes, http://www.psu.edu/dept/cac/ait/viz/java3d/inter.html (visited in March 2002)

Java 3D: Sun Microsystems, Java 3D TM API, http://java.sun.com/products/java-media/3D/ (visited in March 2002)

Whatis: whatis.com , virtual reality, http://whatis.techtarget.com/definition/0,,sid9_gci213303,00.html (visited in March 2002).

TERRAIN VISUALIZATION APPLIED TO VEGETATION STUDIES OF NATIONAL PARKS AND HISTORIC SITES

Roy WELCH, Thomas JORDAN and Marguerite MADDEN

Center for Remote Sensing and Mapping Science
The University of Georgia
Athens, Georgia 30602-2503, USA
{rwelch, tjordan, mmadden}@crms.uga.edu

Commission IV, WG IV/6

KEY WORDS: 3D terrain visualization, DEM, vegetation analysis

ABSTRACT:

Detailed vegetation databases are being compiled from color infrared (CIR) aerial photographs and used in conjunction with 3D terrain visualization techniques to create interactive displays of National Parks and Historic Sites in southeastern United States. The sizes of the parks vary from about 2000 km^2 for the rugged Great Smoky Mountains National Park, located in the Appalachian mountains of Tennessee and North Carolina, to small national battlefields and historic sites of less than 100 ha. The dense forest cover in many of the parks has required the use of Global Positioning System (GPS) surveys in combination with softcopy photogrammetry, DEMs, image processing techniques and geographic information system (GIS) procedures to construct large-scale digital orthophotos and vector-based vegetation databases (Welch et al., 2000). As part of the study, the DEMs generated by stereocorrelation procedures from the color IR photographs were compared to those available from the U.S. Geological Survey. Mapping the floristically diverse parks also required the development of detailed vegetation classification systems suitable for use with 1:12,000 and 1:40,000 scale CIR aerial photographs.

Elevation, slope and aspect of the terrain are all very important environmental factors that determine the type of vegetation associations that will be found in a particular setting. The relationship between vegetation classification and terrain was assessed by draping orthophotos and vegetation maps onto DEMs to create 3-D perspective views (Madden and Jordan, 2001). Flyovers through the vegetation-draped 3D perspectives were constructed to help researchers visualize vegetation changes in relation to elevation, slope and aspect. Statistical analyses also were conducted to quantify the spatial correlation between vegetation and terrain characteristics. Output from these studies include interpretive data sets that will aid park managers and researchers in laying out nature trails, modeling fire behavior and developing displays of interest to park visitors.

REFERENCES

Madden, M. and T. Jordan, 2001. Spatiotemporal analysis and visualization for environmental studies, *International Archives of Photogrammetry and Remote Sensing*, Vol. 34, Part 4/W5: 71-72.

Welch, R., T. Jordan and M. Madden, 2000. GPS surveys, DEMs and scanned aerial photographs for GIS database construction and thematic mapping of Great Smoky Mountains National Park, *International Archives of Photogrammetry and Remote Sensing*, Vol. 33, Part B4/3: 1181-1183.

DENSIFICATION OF DTM GRID DATA USING SINGLE SATELLITE IMAGERY

M.A. Rajabi, J.A.R. Blais

Dept. of Geomatics Eng., University of Calgary, Calgary, AB, T2N 1N4, Canada, – (marajabi, blais)@ucalgary.ca

KEY WORDS: Digital Terrain Model, Shape from Shading, Satellite Imagery, Irradiance Model, Curvature Consistency

ABSTRACT:

Digital Terrain Models (DTMs) are simply regular grids of elevation measurements over the land surface. DTMs are mainly extracted by applying the technique of stereo measurements on images available from photogrammetry and remote sensing. Enormous amounts of local and global DTM data with different specifications are now available. However, there are many geoscience and engineering applications which need denser DTM grid data than available. Fortunately, advanced space technology has provided much single (if not stereo) high-resolution satellite imageries almost worldwide. Nevertheless, in cases where only monocular images are available, reconstruction of the object surfaces becomes more difficult. Shape from Shading (SFS) is one of the methods to derive the geometric information about the objects from the analysis of the monocular images. This paper discusses the use of SFS methods with single high resolution satellite imagery to densify regular grids of heights. Three different methodologies are briefly explained and then implemented with both simulated and real data. Moreover, classification techniques are used to fine tune the albedo coefficient in the irradiance model. Numerical results are briefly discussed.

1. INTRODUCTION

DTMs are used for the analysis of topographical features in GISs and numerous engineering computations as well as scientific applications. Rajabi and Blais [2001] briefly reviewed and referenced a number of sources for DTMs and their applications in both engineering and science.

Stereo measurements from pairs of aerial photographs or satellite images have mainly been used as the primary data in producing DTMs. Information from double or multiple images in overlap areas ensures reliable and stable models for geometric and radiometric processing. Especially recently, with the rapid improvement in remote sensing technology, automated analysis of stereo satellite data has been used to derive DTM data ([Gugan and Dowman, 1988], [Simard et al, 1988], and [Tam, 1990]).

Today, with the need for the better management of the limited natural resources, there are numerous geoscience and engineering applications which require denser DTM data than available. Unfortunately stereo satellite imagery is not available everywhere. Obviously, time and cost are two important factors that often prevent us from field measurements. While interpolation techniques are fast and cheap, they have their own inherent difficulties and problems, especially in terms of accuracy of interpolation in rough terrain.

On the other hand, the availability of single satellite imagery for nearly all of the Earth is taken for granted nowadays. However, reconstruction of objects from monocular images is very difficult, and in some cases, not possible at all. Inverse rendering or the procedure of recovering three-dimensional surfaces of unknown objects from two-dimensional images is an important task in computer vision research. Shape from Shading (SFS) [Horn, 1970] [Horn, 1990], [Zhang et al, 1994] is one of the techniques used for inverse rendering which converts the reflectance characteristics in images to shape information.

This paper discusses the application of SFS techniques to improve the quality of the interpolated DTM grid data with single satellite imagery of better resolution than the DTM data. The idea is highly motivated by the wide availability of satellite remotely sensed imagery such as Landsat TM and SPOT HRV imagery. Section 2 briefly reviews the general SFS problem and the methods implemented in this paper. Section 3 provides numerical tests of the methods explained in Section 2 in more detail. Last but not least, Section 4 provides some concluding remarks.

2. SHAPE FROM SHADING

2.1 Image Formation

SFS is one of the methods which transforms single or stereo 2D images to a 3D scene. Basically, it recovers the surface shape from gradual variations of shading in the image. The recovered shape can be expressed either in terrain height z(x,y) or surface normal N or surface gradient $(p,q)=(\partial z/\partial x, \partial z/\partial y)$.

Studying the image formation process is the key step to solve the SFS problem. A Lambertian model is the simplest one in which it is assumed that the grey level at each pixel depends only on light source direction and surface normal. Assuming that the surface is illuminated by a distant point source, we have the following equation for the image intensity:

$$R(x,y) = \rho\ \mathbf{N} \cdot \mathbf{L} = \rho \frac{pl_1 + ql_2 + l_3}{\sqrt{p^2 + q^2 + 1}} \qquad (1)$$

where ρ is the surface albedo, N is the normal to the surface and $\mathbf{L} = (l_1, l_2, l_3)$ is the light source direction. Even with known ρ and L, the SFS problem will still be a challenging subject, as this is one nonlinear equation with two unknowns for each pixel in the image. Therefore, SFS is intrinsically an underdetermined problem and in order to get a unique solution, if there is any at all, we need to have some constraints.

Based on the conceptual differences in the algorithms, there are three different strategies to solve the SFS problem [Rajabi and Blais, 2001]: 1. Minimization (regularization) approaches 2. Propagation approaches, and 3. Local approaches. A more detailed survey of SFS methods can be found in [Zhang et al, 1999]. The following subsections very briefly review the minimization approach, which is widely used in to solve the SFS problem, and the other variants of the minimization approach which are used here to enhance the solution.

2.2 Minimization Approach

Based on one of the earliest minimization methods, the SFS problem is formulated as a function of surface gradients, while brightness and smoothness constraints are added to ensure that a unique solution exists [Ikeuchi and Horn, 1981]. The brightness constraint ensures that the reconstructed shape produces the same brightness as the input image. The smoothness constraint in terms of second order surface gradients helps in reconstruction of a smooth surface. Brooks and Horn [1985] defined the error functional:

$$I = \iint \left((E(x,y) - N.L)^2 + \lambda(\|N_x\|^2 + \|N_y\|^2) + \mu(\|N\|^2 - 1) \right) dxdy \tag{2}$$

where E(x,y) is the grey level in the image, N_x and N_y are the partial derivatives of the surface normal with respect to x and y directions respectively and the constants λ and μ are Lagrangian multipliers. As it can be seen the functional has three terms: 1) the brightness error which encourages data closeness of the measured images intensity and the reflectance function, 2) the regularizing term which imposes the smoothness on the recovered surface normals, and 3) the normalization constraint on the recovered normals.

The functional is minimized by applying variational calculus and solving the Euler equation: The resulting fixed-point iterative scheme for updating the estimated normal at the location of (i,j) and epoch k+1, using the previously available estimate from epoch k is:

$$N^{k+1}_{i,j} = \frac{1}{1+\mu_{i,j}(\varepsilon^2/4\lambda)} \left(M^k_{i,j} + \frac{\varepsilon^2}{4\lambda}\left(E_{i,j} - N^k_{i,j}.L\right).L \right) \tag{3}$$

where

$$M^k_{i,j} = \frac{1}{4}(N^k_{i+1,j} + N^k_{i-1,j} + N^k_{i,j+1} + N^k_{i,j-1}) \tag{4}$$

There are two comments about this update equation. First, it seems that one has to solve for the Lagrangian multiplier $\mu_{i,j}$ on a pixel-by-pixel basis. However, as it is seen $\mu_{i,j}$ enters the update equation as a multiplying factor which doesn't change the direction of the update normal, therefore we can replace that factor by a normalization step. The second comment is about the geometry of the update equation. As it is seen, the update equation is composed of two components. The first one comes from the smoothness constraint while the second one is a response to the physics of image irradiance equation.

The main disadvantage of the Brook and Horn method or any other similar minimization approach is the tendency of over smoothing the solution resulting in the loss of fine detail. Selecting a conservative value for the Lagrangian multiplier is a very challenging issue in this method. However, in an attempt to overcome this problem, Horn [1990] starts the solution with a large value for the Lagrangian multiplier and reduces the influence of the smoothness constraint in each iterations as the final solution is approached.

2.3 Modified Minimization Approaches

As it was mentioned in the previous section, the update equation is composed of two components, the smoothness part and the data closeness part. As the first attempt to solve the over smoothing problem with the general minimization approach, an adaptive regularization parameter $\lambda(i,j)$ instead of a fixed λ is suggested to be used to adaptively control the smoothness over the image space [Gultekin and Gokmen, 1996]. In each iteration, the space varying regularization parameter at location (i,j) can be determined by the following function:

$$\lambda_{new} = (1 - e^{-\frac{c(i,j)}{V_T}})\lambda_{min} + (e^{-\frac{c(i,j)}{V_T}})\lambda_{old}(i,j) \tag{5}$$

where c(i,j) is the control parameter, V_T is a time constant that regulates the rate of exponential decrease and λ_{min} is a preselected minimum value that $\lambda(i,j)$ may have. The control parameter is defined as c(i,j)=|I(i,j)-R(i,j)|, where |-| denotes the absolute value and the function λ_{new} is an exponentially decreasing function with the following properties:

$$\lim_{c(i,j)\to 0} \lambda_{new} = \lambda_{old}(i,j), \qquad \lim_{c(i,j)\to\infty} \lambda_{new} = \lambda_{min} \tag{6}$$

so that the regularization parameter is only allowed to decrease with the iterations.

Another method to solve the over smoothing problem is to use a robust error kernel in conjunction with curvature consistency instead of a quadratic smoothness. The robust regularizer constraint function can be defined as [Worthington and Hancock, 1999]:

$$\rho_\sigma(\|N_x\|) + \rho_\sigma(\|N_y\|) \tag{7}$$

where $\rho_\sigma(\eta)$ is a robust kernel defined on the residual η and with width parameter σ. Among different robust kernels, it is proved that the sigmodial-derivative M-estimator, a continuous version of Huber's estimator, has the best properties for handling surface discontinuities [Worthington and Hancock, 1999] and is defined by:

$$\rho_\sigma(\eta) = \frac{\sigma}{\pi} \log\cosh\left(\frac{\pi\eta}{\sigma}\right) \tag{8}$$

Applying calculus of variations to the constraint function using the above mentioned kernel results in the corresponding update equation. Here σ, the width parameter of the robust kernel, is computed based on the variance of the shape index. Based on Koenderink and Van Doorn [1992] the shape index is another way of representing curvature information. It is a continuous measure which encodes the same curvature class information as the mean and Gauss curvature, but in an angular representation. In terms of surface normals, the shape index is defined as [Koenderink and Van Doorn, 1992]:

$$\phi = \frac{2}{\pi} \tan^{-1} \frac{\mathrm{N}_{x1} + \mathrm{N}_{y2}}{\sqrt{(\mathrm{N}_{x1} - \mathrm{N}_{y2})^2 + 4\mathrm{N}_{x2}\mathrm{N}_{y1}}} \tag{9}$$

where the second subscripts 1 and 2 correspond to the x and y components respectively. The variance dependence of the kernel in eq. (7) is controlled using the exponential function:

$$\sigma = \sigma_0 \exp\left(-\left(\frac{1}{N}\sum \frac{(\phi_l - \phi_c)^2}{\Delta\phi_d^2}\right)^{1/2}\right) \tag{10}$$

where σ_0 is the reference kernel width which we set to one, ϕ_c is the shape index associated with the central normal of the neighborhood, $\mathrm{N}_{i,j}$, ϕ_l is one of the neighboring shape index values and $\Delta\phi_d$ is the difference in the shape index between the center values of adjacent curvature classes which is equal to 1/8 [Koenderink and Van Doorn, 1992].

The other modification that can be done on the minimization approach is on the data closeness part of the update equation. We know that the set of surface normals at a point which satisfy the image irradiance equation define a cone about the light source direction. In other words, the individual surface normals can only assume directions that fall on this cone. At each iterations the updated normal is free to move away from the cone under the action of the local smoothness. However, we can subsequently map it back onto the closest normal residing on the cone. This has not only numerical stability advantages but also all normal vectors in the intermediate states are solutions of the image of irradiance equation. In other words, the update equation for the surface normals can be written as:

$$\mathrm{N}_{i,j}^{k+1} = \mathrm{R}(\theta)\, \mathrm{M}_{i,j}^{k} \tag{11}$$

where

$$\theta = -\cos^{-1}\left(\frac{\mathrm{M}_{i,j}^{k} \cdot \mathrm{L}}{\left\|\mathrm{M}_{i,j}^{k}\right\| \left\|\mathrm{L}\right\|}\right) + \cos^{-1} \mathrm{E}\,. \tag{12}$$

$\mathrm{M}_{i,j}^{k}$ in eq. 11 is the surface normal that minimizes the smoothness constraint while $\mathrm{R}(\theta)$ is the rotation matrix of angle θ which maps the updated normal to the closest normal lying on the cone of ambiguity (Figure 1). The axis of rotation is found by taking the cross product of the intermediate update with the light source direction:

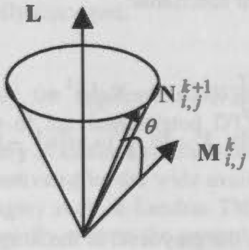


Figure 1. The ambiguity cone. **L** is the light source vector, $\mathbf{M}_{i,j}^{k}$ is the surface normal which minimizes the smoothness constraint and $\mathbf{N}_{i,j}^{k+1}$ is the updated surface normal.

3. NUMERICAL TESTS

3.1 Processing Steps

The main goal of this investigation is to improve the accuracy of the interpolated DTM grid data by applying SFS techniques to the corresponding single satellite imagery, while the original DTM data are used as boundary constraints in the SFS problem.

The basic assumption here is that the satellite imagery has one dyadic order better resolution than the original DTM data. We also assume that 1) the surface is Lambertian (which is questionable in reality), 2) the surface albedo is known (by applying classification techniques to multispectral satellite imageries), 3) the surface is illuminated by a distant point source (Sun), and finally 4) the position of the light source is known.

Our approach deals with a patch at a time (Figure 2) with forty nine points. Sixteen grid points have known heights (dark circles) from DTM grid data and the other thirty three are points with the interpolated heights (unmarked grid points). Our main objective is to improve the accuracy of the interpolation for the five innermost unknown points.

The method essentially consists of three stages: 1) preprocessing, 2) processing, and 3) postprocessing. The preprocessing stage itself has three steps. In the first step, using interpolation (bilinear) techniques, the heights of the unknown points in the patch are estimated. When dealing with the real height data, if there is a gap in height measurements due to any reason (such as existing rivers or lakes), the whole patch is left untouched.

Classification of the image data is the second step of preprocessing stage. The Mahalanobis classifier [Richards, 1986], which is one of the supervised classification methods, is used to classify the image data. The choice of Mahalanobis classifier is because of its simplicity while using the covariance information. After the classification is done, the average grey level data in each class is used to tune the albedo coefficient in irradiance model.

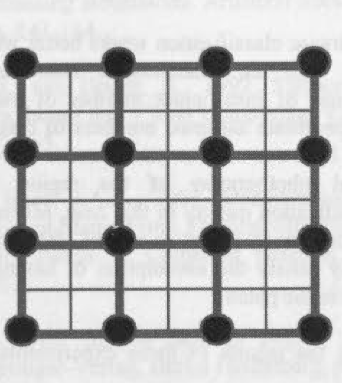


Figure 2. A patch: Circles are the grid points with known heights from DTM grid data and the unmarked ones are the points with the interpolated heights.

In the third step of preprocessing stage, using the known grid points, the relative orientation of the inner most square in the patch with respect to the light source is estimated. If this relative orientation implies that the patch is in the shadow, then there would be no useful shading information to improve the accuracy of the interpolated heights. Therefore, in this case the interpolated heights are considered the final height values.

Otherwise, the processing stage for each patch consists of three steps. In the first step, the smoothed surface normals (based on one of the three methods explained in Section 2 are computed. Then in the second step these surface normals are mapped onto the corresponding ambiguity cones. Finally in the third step, the surface normals are passed to an overdetermined (74 equations and 33 unknowns) linear adjustment process to solve for the heights. In this process, the updated surface normals are considered as the observations while the interpolated height values are the unknowns. This is simply done by approximating p and q with finite differences in terms of heights. The control goes back to the first step of this stage unless the average difference between the calculated and original image grey values of all the pixels in the patch is less than a predetermined threshold.

The last stage, postprocessing, consists of taking arithmetic means of two solutions for the unknown (interpolated) heights located on the boundary of the innermost square in each patches coming from the neighboring patches, except for the outsides of the peripheral patches.

3.2 Experimental Tests

In order to check the validity of the methodologies explained in Sections 2 and 3, a number of numerical examples using both simulated and real world data set are used. The following subsections describe the test procedures in more detail.

3.2.1 Simulated Data Set: First, one synthetic object and its corresponding synthetic imagery are created. The synthetic object is a 1024 by 1024 pixel convex hemisphere with a radius of 250 units, sampled at each 0.5 unit. The corresponding DTM (one dyadic order less than the object, i.e., 512 by 512 pixels) is extracted out of the object. Meanwhile, its corresponding image is created using a Lambertian reflectance model with a much denser version of the object while the albedo coefficient is set to 255. Then the grey level of each pixel is scaled to one of the randomly selected different levels trying to simulate the Earth surface with three different types of coverage .

The differences between the original object, the Interpolated Grid Solution (IGS) and the three SFS solutions, each in turn, in three different cases are analysed. In the first case (Table 3), the assumption is that the albedo coefficient is the same for all the pixels, i.e., same coverage type everywhere, and therefore no classification is done. In the second case, the Mahalanobis classifier with three different classes is used to classify the image pixels. 100 pixels for each class are used as the training data set. Once the pixels are assigned to a class, the average grey value of each class is used as the albedo coefficient for any pixel in that class. In the third case, using the interpolated height values and albedo of 255 for all the pixels, the image of each patch is reconstructed and is compared to the corresponding grey values in the original image file. The mean difference between these two grey values is used to scale the assumed albedo. Table 3 shows the percentages of the improvements in accuracy of DTM densification resulting from these analyses.

	Case 1	Case 2	Case 3
SFS1 vs IGS	18.2%	38.2%	20.5%
SFS2 vs IGS	23.2%	41.8%	26.2%
SFS3 vs IGS	25.6%	42.1%	28.7%

Table 3. Improvement percentages relative to corresponding IGS. Cases 1, 2, 3 show the No Classification, Classification with Mahalanobis classifier and Patchwise Classification cases respectively.

3.2.1 Real Data Set: The second object is a real terrain data set from southern Alberta (Waterton), Canada with 25 metre spacing in UTM coordinate system. A set of 1024 by 1024 grid with over 1300 metre height differences is selected from the four quadrants of NTS 82H04 DTM provincial data file. The originally measured DTM data consists of 100 metre spacing grid in addition to the feature points. These measurements had been used to interpolate the provincial 25-metre and other DTM grids.

The corresponding satellite imagery is a three channel SPOT data file with 20 metre resolution which was originally georeferenced to an extended UTM coordinate system. By extracting the coordinates of the predominant terrain features from the corresponding 1/20000 topographic map sheets, the SPOT imagery is georeferenced to the same coordinate system as the DTM data. For this purpose 23 points with good distribution in the area under investigation are used. A second order polynomial is used for the purpose of georeferencing. The RMS of georeferencing in easting and northing directions are 2.96 m and 1.40 m respectively.

Using PCI software (from PCI Geomatics Enterprises Inc.), a principal component transformation is applied and the first channel with 88.52% energy is selected for these experiments. Then the satellite imagery pixel size is changed from 20 m to 25 m using a bilinear interpolation method.

The efficiency of the SFS methods with the real data set is analysed by trying to reconstruct the 25 m DTM from 50 m DTM and 25 m SPOT imagery. Similar to the simulated data case, the differences between the original object, the Interpolated Grid Solution (IGS) and the three SFS solutions each in turn in three different cases are analysed. In the first case, the assumption is that the albedo coefficient is the same for all the pixels, i.e., same coverage type everywhere, and therefore no classification is done. In the second case, the Mahalanobis classifier with the assumption of three different classes (water, soil and vegetation) is used to classify the image pixels. In doing that all the information in three channels of satellite imagery is used. 100 pixels for each class are used as the training data set. Once the pixels are assigned to their corresponding class, the average grey value of each class in the principal component transformed file is used as the albedo coefficient for any pixel in that class. In the third case, using the interpolated height values and albedo of 255 for all the pixels, the image of each patch is reconstructed and compared to the corresponding grey values in the principal component transformed file. The mean difference between these two grey values is used to scale the assumed albedo. Table 4 shows the results of these analyses.

	Case 1	Case 2	Case 3
SFS1 vs IGS	32.3%	42.2%	46.3%
SFS2 vs IGS	41.1%	48.6%	54.7%
SFS3 vs IGS	47.8%	51.4%	57.2%

Table 4. Improvement percentages relative to corresponding IGS. Cases 1, 2, 3 show the No Classification, Classification with Mahalanobis classifier and Patchwise Classification cases respectively.

4. CONCLUDING REMARKS

The experimental results in Tables 3 and 4 have been analysed and it is concluded that:

1. In all of these cases, applying any of the SFS methods will give us a more accurate densified DTM than the interpolation method. However, the percentage of improvement is different in each case. Generally speaking, it is seen that using a SFS method based on the curvature index will result in a better solution in comparison to the other SFS methods in the same case.

2. It is realized that the improvement percentages in the SFS solutions are higher with the real data set. That seems to imply that our simulated image generation was not as realistic as could be expected.

3. Applying classification methods to calibrate the albedo coefficient in the image irradiance model seems to improve the SFS solutions. However, in case of simulated data set, the Mahalanobis classifier works better than the pacthwise classifier. There seem to be two main reasons for this. First, in the simulation process, the differences between albedo coefficients of different surface types were deliberately considered high. Then it is obvious that Mahalanobis classifier can classify the pixels with better accuracy in comparison to the real data case. Second, the distribution of pixels from different classes in the simulated image is too random to be compatible with the real data set. In other words, in simulating the image with different surface coverages, no correlation between neighboring pixels is considered which is not the case in reality. This seems to imply that patchwise interpolation is not appropriate for these simulated data sets.

4. Patchwise classification works better with the real data set case than the Mahalanobis classifier. This might be because of insufficient number of training data sets or inappropriate assumed numbers of classes. Having access to a land use map or other source of information such as aerial photography of the region can improve the classification quality in this case. Moreover, each patch in our case study is just an area of 150 m by 150 m which can easily satisfy the assumption of having the same ground type in the patch.

Based on the results of these experiments, future work will include:

1. Studying the behaviour of SFS techniques with different terrain types and DTM resolutions as well as satellite imageries of different quality.

2. As the SFS solutions used in these experiments are apparently suffering from over simplified irradiance models, more sophisticated image formation equations should be more deeply investigated especially for satellite imageries.

3. The terrain properties in the statistical analysis results shown here are assumed to be homogeneous and stationary. However, these assumptions are not true in reality. Obviously, domain segmentation is required and has to be taken into consideration. It goes without saying that the effectiveness of the SFS method in densifying DTM should be studied in rough terrain applications.

4. Each scene of satellite imagery usually covers a wide area of the Earth and the resolution can be as high as one metre or less. When dealing with such applications, tremendous amounts of computations have to be expected. Using parallel processing techniques in these types of applications is a must for general implementations.

5. REFERENCES

Brooks, M.J., Horn, B.K.P., 1985. Shape and Source from Shading. International Joint Conference on Artificial Intelligence, pp. 932-936.

Gugan, D.J., Dowman, I.J., 1988. Topographic Mapping from Spot Imagery. Photgrammetric Engineering and Remote Sensing 54(10), pp. 1409-1414.

Gultekin, A., Gokmen, M., 1996. Adaptive Shape From Shading. ISCIS XI The Eleventh International Symposium on Computer and Information Sciences, pp. 83-92.

Horn, B.K.P., 1970. Shape from Shading: A method for Obtaining the Shape of a Smooth Opaque from One View. Ph.D. Thesis, Massachusetts Ins. of Technology.

Horn, B.K.P., 1990. Height and gradient from shading. Int. J. Comput. Vision, 37-5.

Horn, B.K.P., 1990. Height and Gradient from Shading. Int. J. Comput. Vision , Vol. 5, No. 1, pp. 37-75.

Ikeuchi, K., Horn, B. K. P., 1981. Numerical Shape from Shading and Occluding Boundaries. Artificial Intelligence, Vol. 17, Nos. 1-3, pp. 141-184.

Koenderink, J.J., van Doorn, A., 1992. Surface Shape and Curvature Scales. Image and Vision Computing, Vol. 10, No. 8, October, pp. 557-565.

Rajabi, M.A., Blais, J.A.R., 2001. Densification of Digital Terrain Model Using Shape From Shading with Single Satellite Imagery. Lecture Notes in Computer Science, Vol. 2074, Springer, pp. 3-12.

Richards, J.A., 1986. *Remote Sensing Digital Analysis An Introduction.* Springer-Verlag, Berlin Heidelberg, pp.185-186

Simard, R., Rochon, G., Leclerc, A., 1988. Mapping with SPOT Imagery and Integrated Data Sets. Invited paper presented at the 16th congress of the International Society for Photogrammetry and Remote Sensing held July 1988 in Kyoto, Japan.

Tam, A.P., 1990. Terrain Information Extraction from Digital SPOT Satellite Imagery. Dept. of Geomatics Eng., The University of Calgary.

Worthington, P.L., Hancock, E.R., 1999. Needle Map Recovery Using Robust Regularizers. Image and Vision Computing, Vol. 17, pp. 545-557.

Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M., 1994. Analysis of Shape from Shading Techniques. Proc. Computer Vision Pattern Recognition, pp. 377-384.

Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M., 1999. Shape from Shading: A Survey. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 21, No. 8, August, pp. 690-706.

USE OF AIRSAR / JERS-1 SAR DATASETS IN GEOLOGIC / STRUCTURAL MAPPING AT THE NORTHERN NEGROS GEOTHERMAL PROJECT (NNGP), NEGROS OCCIDENTAL, PHILIPPINES

R. A. Camit, L.F. Bayrante, C.C. Panem, O.C. Bien and J.A. Espiridion

PNOC Energy Development Corporation, Merritt Road, Fort Bonifacio, Metro Manila 1201, Philippines
rexcamit@hotmail.com

Commission IV, WG IV/6

KEYWORDS: Airborne SAR, Topographic SAR, Polarimetric SAR, DEM, aerial photography, high resolution, hillshading

ABSTRACT:

A remote sensing study of the Northern Negros Geothermal Project (NNGP) was conducted using Airborne Synthetic Aperture Radar (AIRSAR) AND JERS-1 SAR images. A digital elevation model (DEM) was generated from the topographic SAR (TOPSAR) data sets where various hillshading maps were extracted to emphasize the general geology and the structural configuration of the study area. Due to the high resolution of the imagery, geologic mapping interpretations were done to refine the traditional and existing aerial photography interpretations including the volcanic history and its implication to the project.

The over-all result of this study have dealt with (1) the delineation of the semi-detailed geology at NNGP, thereby, giving way to the reinterpretation of the volcanic history of Canlaon volcano; (2) refinement of the structural map of Panem and Leynes (1996) with the establishment of chronological order of fault events where NW, WNW and ENE predominates the field; (3) delineation of new structural targets at Pataan area (e.g. Mambucal-B and -C, Asia-C, Asia Splay-A, -B, -C and -D, Catugasan-B Splay, Catugasan-C and -D, Kinabkaban-B, Kinabkaban-C and Pataan-C Splay); (4) confirmation of surface and subsurface interception of Talave Formation (rather than Caliling Formation) with findings of PNOC-EDC (2001); and (5) singling out of Hardin Sang Balo (rather than Tubidiao crater / collapse and Saray dome / collapse) as the probable heat source being tapped by the Pataan sector of NNGP.

1.0 INTRODUCTION

The study area is geographically bounded by Panay Island to the northwest, by Cebu Island to the west and structurally bounded by the Negros Trench to the southwest (Fig. 1). The PNOC Energy Development Corporation (PNOC-EDC) drilled 12 exploration and delineation wells (e.g. MC-1, MC-2, HG-1D, CT-1D, PT-1D, PT-2D, PT-3D, PT-4D, PT-5D, PT-6D, PT-7D and PT-1RD) within the study area (Fig. 2). Wells PT-7D and PT-1RD are yet fully tested while drilling of PT-8D is still ongoing as of this writing. The wells are distributed within the Mambucal, Hagdan, Catugasan and Pataan sectors, respectively, of the Northern Negros Geothermal Project (NNGP). PNOC-EDC intends to develop the Northern Negros Geothermal Project with an aggregate capacity of 40MWe.

This study is the result of integrating available remote sensing data (e.g. airborne synthetic aperture radar of NASA-JPL and JERS-1 synthetic aperture radar of ERSDAC) with the available geographic information system (GIS) database of NNGP.

The interpretation of AIRSAR imagery for geologic and structural mapping applications, however, was limited to the project site at the northwestern sector of Mt. Canlaon (Camit et al., 2000). Thus, to compensate for the lack of data in the other sectors, a three-strip mosaic of processed Topographic Synthetic Aperture (TOPSAR) data sets at 20-m contour interval were utilized to map geologic units. In addition, confirmation of said geologic units was likewise observed on JERS-1 SAR imagery from ERSDAC. The relative age of each unit was based on superposition relationships, flow morphology, degree of dissection, vent source, relief and textural differences. The paleontologic dates were derived from four rock samples from wells and surface outcrops (Lubas, 2001).

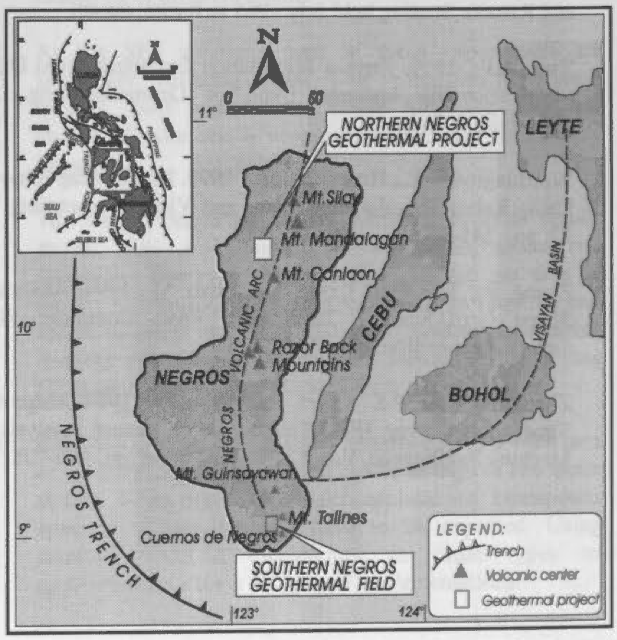


Figure 1. Location Map of the Northern Negros Geothermal Project and the Negros volcanic arc. The Southern Negros Geothermal Field is located at the southern side of the island.

The existing reports on the geology, volcanology and stratigraphy of the area have been freely consulted and used

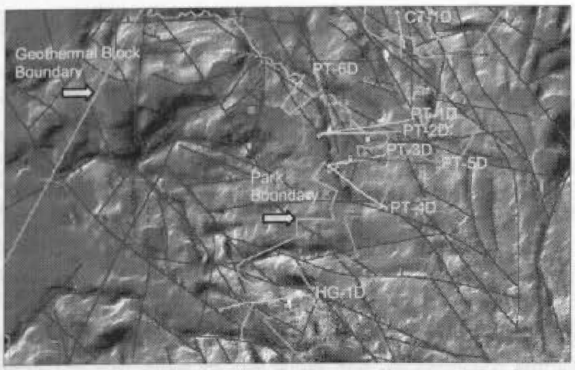

Geothermal Block Boundary
CT-1D
PT-6D
PT-1D
PT-2D
PT-3D
PT-5D
Park Boundary
PT-4D
HG-1D

Figure 2. Location of NNGP wells at Mambucal, Hagdan, Catugasan and Pataan sectors showing geothermal block and Natural park boundaries. Mambucal wells MC-1 and MC-2 are located NW of Pataan (PT) wells.

when applicable to synthesize the geology and stratigraphy within the project (Melendres, 1975; Ferrer, 1979; BMGS, 1982; Umbal and Arboleda, 1987; Pornuevo, 1988; Umbal and Arboleda, 1988; Pamatian et al., 1992; Catane, 1994; von Biedersee and Pichler, 1995; PNOC-EDC, 1998; Martinez-Villegas, 1999).

This study presents: (1) the general geology, structural lineaments and proposed stratigraphy within the Northern Negros Geothermal Project (NNGP) and (2) a reinterpretation of the volcanic history of the Canlaon volcano. This study was based on interpretations made mainly from Topographic Synthetic Aperture Radar (TOPSAR) and JERS-1 SAR imageries and synthesis of earlier geological, petrological and paleontological studies.

2.0 METHODOLOGY

The C-DEM file was mainly used in this study where it was converted to physical parameters, generating the lambertian (shaded relief) surface and extraction of the polarimetric information such as the aspect and slope that was done using the topographic modeling option of ENVI 3.2. Final processing and interpretation of the topographic data are done through ILWIS 2.1. An attempt to delineate the general geology, stratigraphy and volcanic history of the study area was performed through the use of TOPSAR and JERS-1 SAR imageries.

Three imageries were used to extract the structural features at NNGP: (1) slant-to-ground range corrected radar imagery (TOPSAR and POLSAR imageries); (2) georeferenced JERS-1 SAR imagery; and (3) NW-directed / SE-directed hillshading maps (from NAMRIA 1:200,000 m scale map).

3.0 GEOLOGIC INPUTS

3.1 Tectonic Setting

Canlaon Volcano forms the mid-northern segment of the west-facing island arc system that is partly defined by the Negros volcanic arc in the western side of Central Philippines (Fig. 3). The volcanic belt consists of Cuernos de Negros, Razor Back Mountains, Mount Silay, Mount Mandalagan and two small island volcanoes east of Panay Island (30 km north of Negros Island) (Martinez-Villegas, 1999). The arc is the result of subduction of Sulu Sea Basin lithosphere along the Negros Trench (Acharya and Aggarwal, 1981) that has been an active subduction zone since the Miocene (Cardwell *et al.,* 1980). This trench was eventually dammed when the North Palawan - SW Mindoro continental fragment collided with the Philippine Arc during the late Miocene. Geochemical evidence of high potassium content of Negros Volcanics have been similarly associated with source contamination by remnants of the continental crust of said collision (von Biedersee and Pichler, 1995). In addition, this 150 km deep Negros trench has no clearly developed Benioff-Wadati zone (Acharya and Aggarwal, 1980) but has a present day subduction rate of about 2 cm per year.

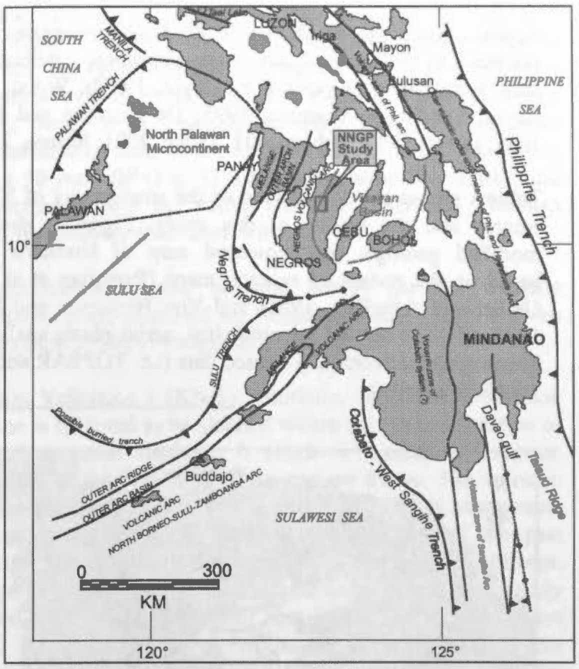
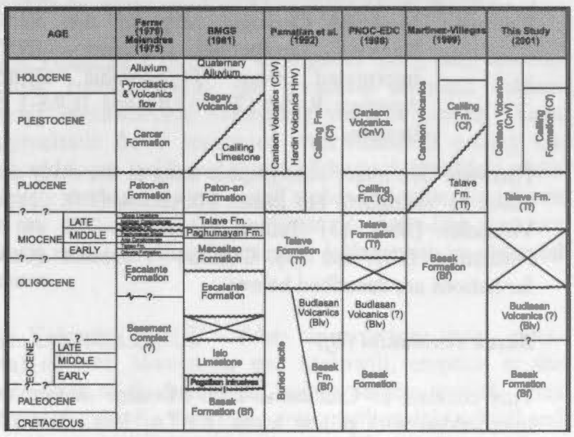

SOUTH CHINA SEA
MANILA TRENCH
PALAWAN TRENCH
North Palawan Microcontinent
LUZON
Iriga
Mayon
Bulusan
PHILIPPINE SEA
Philippine Trench
NNGP Study Area
PANAY
PALAWAN
Visayan Basin
CEBU
BOHOL
NEGROS
NEGROS VOLCANIC ARC
Negros Trench
SULU SEA
SULU TRENCH
MINDANAO
Possible buried trench
OUTER ARC RIDGE
OUTER ARC BASIN
VOLCANIC ARC
NORTH BORNEO-SULU-ZAMBOANGA ARC
Buddajo
SULAWESI SEA
Cotabato - West Sangihe Trench
Davao gulf
Talaud Ridge
0
300
KM
10°
120°
125°

Figure 3. Distribution of Active and Inactive Volcanoes and the Different Tectonic Features of the Philippines(Hamilton, 1979).



AGE		Ferrer (1979) Melendres (1975)	BMGS (1981)	Pamatian et al. (1992)			PNOC-EDC (1998)	Martinez-Villegas (1999)		This Study (2001)	
HOLOCENE		Alluvium Pyroclastics & Volcanics flow	Quaternary Alluvium Sagay Volcanics	Canlaon Volcanics (CnV)	Hardin Volcanics HnV)	Calling Fm. (Cf)	Canlaon Volcanics. (CnV)	Canlaon Volcanics	Calling Fm. (Cf)	Canlaon Volcanics (CnV)	Calling Formation (Cf)
PLEISTOCENE		Carcar Formation	Caliling Limestone								
PLIOCENE		Paton-an Formation	Paton-an Formation				Calling Fm. (Cf)	Talave Fm. (Tf)		Talave Fm. (Tf)	
MIOCENE	LATE	[illegible]	Talave Fm.	Talave Formation (Tf)			Talave Formation (Tf)	Basak Formation (Bf)			
	MIDDLE	[illegible]	Paghumayan Fm.								
	EARLY	[illegible]	Macasilao Formation								
OLIGOCENE		Escalante Formation	Escalante Formation	Buried Dacite	Budlasan Volcanics (Blv)		Budlasan Volcanics (?) (Blv)			Budlasan Volcanics (?) (Blv)	
EOCENE	LATE	Basement Complex (?)	Isio Limestone		Basak Fm. (Bf)		Basak Formation			Basak Formation (Bf)	
	MIDDLE										
	EARLY		Pagatban Intrusives								
PALEOCENE			Basak Formation (Bf)								
CRETACEOUS											

Table 1. Comparative analysis of Northern Negros stratigraphy illustrating the previous stratigraphy of the region and the result of this study.

3.2 Geologic Mapping Results

Initially, the surface geology and inferred stratigraphy were mapped by this study using: (1) Topographic Synthetic

Aperture Radar (TOPSAR) and JERS-1 SAR imageries and (2) synthesis of surface and subsurface petrologic / paleontologic data from existing reports (i.e.; Reyes, 1979b; Wood, 1979a; Reyes, 1984b; De Guzman, 1995; De Guzman and Revilla, 1995; Lubas, 1997; Lubas, 2001; Dulce, 2001; Ramos, 2001).

Table 1 shows the comparison of the stratigraphy of Northern Negros and the results of this study. Figure 4 shows the modified geologic and structural map of Northern Negros based on the review of existing maps (Pamatian et al., 1992; Umbal and Arboleda, 1988; and Von Biedersee and Pichler, 1995), field survey of selected sites, aerial photo analysis and interpretation of remotely-sensed data (i.e. TOPSAR and JERS-1 SAR imageries).

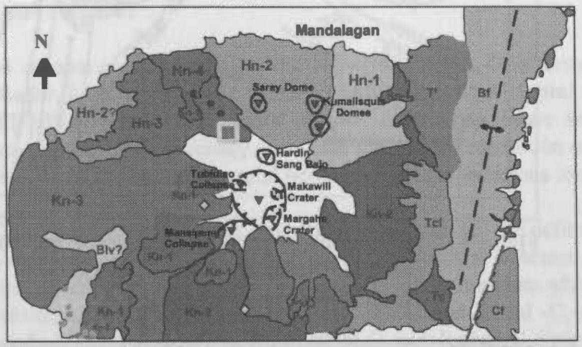


Figure 4. Modified Geologic Map of the Northern Negros Geothermal Project (NNGP) interpreted from Topographic Synthetic Aperture Radar (TOPSAR) and JERS-1 SAR Imagery.

There are five major stratigraphic units in the study area (from oldest to youngest): (1) Basak Formation (Bf); (2) Budlasan Volcanics (Blv); (3) Talave Formation (Tf); (4) Caliling Formation (Cf) and (5) Canlaon Volcanics (CnV). The formations are described below:

Basak Formation (Bf)

This consists of Cretaceous-Late Miocene marine sediments interbedded with pillow lavas (KRTA, 1980; BMGS, 1982). The formation trends N-S and is exposed to the east of the study area. The bathyal marine sediments are composed of mudstone, claystone and gravel conglomerates that dip 45° to the northwest while the lavas are mainly two-pyroxene basalt.

Well MC-1 was previously interpreted to have encountered Basak Formation from 610 mRSL to 718 mRSL where andesitic volcanics and calcareous sediments occur. However, re-evaluation of petrologic data (PNOC-EDC, 1998; Ramos, pers. comm., 2001; Lubas, 2001) indicated that the Basak Formation was never intercepted by Mambucal wells (i.e., MC-1 and MC-2) at depth.

Budlasan Volcanics (Blv)

This rock unit consists of thick columnar-jointed lava flows of hornblende andesite to two- pyroxene andesite. An inferred volcano root ("cinder cone" to Martinez-Villegas *et al.*, 1999) exposed as a flat-topped plateau or dome-like structure is observed roughly 9 km SE of the active Canlaon crater (Pamatian *et al.*, 1992). Well MC-1 was previously interpreted to have intercepted this unit between 406 mRSL to 610 mRSL (Pamatian *et al.*, 1992). Based on the recent analysis of MC-1 rock cuttings by Ramos (pers. comm., 2001), the andesitic volcanic breccias were considered part of the Talave Formation from said depths until TD of well at 718 mRSL.

An outcrop (located approximately 11.7 km SW of the active Canlaon crater) consisting of highly eroded terrane was interpreted from processed radar images as part of the real Budlasan Volcanics. The Budlasan Volcanics delineated by Pamatian *et al.* (1992) to the NW and south of the active Mt. Canlaon has distinct polarized signature similar to the younger Hardin and Canlaon Volcanics compared with the interpreted Budlasan Volcanics southwest of Canlaon volcano identified in this paper.

Talave Formation (Tf)

Based on the analysis of radar imagery, three distinct members of this formation are interpreted and correlated with the lower Talave limestone, middle Tigbao clastics and upper Bairan agglomerate mapped by BMGS (1982) at the eastern portion of the study area. The following lithologic descriptions of the Talave members are primarily obtained from Pamatian *et al.*, (1992) and from observations in this study. The members are designated as Tc, Tcl, and Tl.

An upper member Tc (~Bairan agglomerate?) which appears straddling over deposits of a thicker member Tcl (~Tigbao clastics?) and the Basak Formation was observed from interpreted AIRSAR imagery at the southern section of the north-northeast trending Talave Formation (Fig. 4). In turn, the Tcl member underlies another member Tl (~Talave Limestone) at the northern section of the Talave Formation. The lower Tcl, middle Tl and/or an upper Tc comprise the Talave Formation in the study area. An aggregate thickness of 425 meters was estimated for Tcl and Tl (BMGS, 1982).

The Tcl member mapped by BMGS (1982) consists of laminated to thin-bedded, tuffaceous mudstone interbedded with conglomerate and thin-bedded, gray, sandy limestone. This unit exhibits low to moderate dipping "flat irons" feature at the southeast sector of Mt. Canlaon. This implies bedding structures within the Talave Formation. This formation is low- to moderately- folded beneath the younger Canlaon Volcanic Complex as seen and extrapolated from the processed SAR imagery. A north-northeast trending fold axis trace is inferred along the Basak Formation (Fig. 4). This interpretation is consistent with the inferrence made by Pamatian et al. (1992).

The Tl member is characterized by karst topography with low knobby hills. Within the study area Tl member is found as an isolated feature and onlaps the Tcl member. This unit consists of indurated well-bedded, fossiliferous calcisiltites and calcarenites with beds striking N20°E and dips 15°NW (Pamatian *et al.*, 1992). This member is the equivalent of the Talave Formation of BMGS (1982) and Melendres (1975).

The Tc member at the southeast was also mapped by Pamatian et al. (1992) to have an angular unconformity with the older Basak Formation. This unit consists of sub-rounded to rounded clasts of pyroxene-bearing hornblende andesite to clinopyroxene andesite set in tuffaceous silty to sandy matrix. This deposit is only exposed to the southeast.

This formation was initially described as volcano-sedimentary complex in Pataan, Hagdan and Catugasan wells. However, to be consistent with known paleontologic dating of the sediments, a Late Miocene to Pliocene Talave Formation is proposed for subsurface stratigraphy purposes (Ramos, pers. comm., 2001). This formation has a thickness of 1280 meters at well CT-1D (Ramos, pers. comm., 2001). It consists of fossiliferous limestone interlayered with calcareous to carbonaceous sandstone, siltstone, claystone, mudstone and volcaniclastics. Wells PT-1D and PT-2D did not encounter the Talave Formation. An estimated age range of Late Eocene to Pliocene was previously assigned to this unit (Pamatian *et al.*, 1992).

Caliling Formation (Cf)

This formation onlaps the older Basak and Talave Formations along the east coast of the study area as observed from the interpreted radar imageries of TOPSAR and JERS-1 SAR. This unit is characterized by karstic, friable, fossiliferous, calcisiltites, calcarenites and calcirudites. Its texture suggests massive bedding with dips of 20°-40° NE/SE (Pamatian *et al.*, 1992).

Canlaon Volcanics (CnV)

The term Canlaon Volcanic Complex was initially used by Pamatian et al. (1992) as a sack name for the deposits from vents of the ancestral Canlaon caldera (inferred), Makawili, Tubidiao, Hardin Sang Balo, Margaha, Mansipang and the active Canlaon (Fig. 4). Due to the presence of numerous, distinct vents within Canlaon, Pamatian et al. (1992) adopted the nomenclature Canlaon Volcanic Complex. In addition, Alincastre (1983) and Umbal and Arboleda (1987) also attest that Canlaon Volcano is not a single, vent-type of volcano.

For simplicity, the term Canlaon Volcanics that was used in PNOC (1998) is retained in this study to describe deposits of Canlaon Volcanic Complex (CnVC) at surface and subsurface levels (Table 1). Melendres (1975) and Ferrer (1979) called the same surface deposits as a Pleistocene to Recent "pyroclastic and volcanic flows" while BMGS (1982) called the same formation as Sagay Volcanics. The term Pleistocene to Recent Sagay Volcanics refers to the Quaternary volcanic deposits within Negros Island including the volcanic edifices of Cuernos de Negros, Silay, Mandalagan and Canlaon (BMGS, 1982).

This study confirmed the existence of different volcanic vents that deposited the various pyroclastic units and lava flows around the inferred ancestral Canlaon caldera (Pamatian *et al.*, 1992). Collapse structures (Umbal and Arboleda, 1987; PHIVOLCS, 1987; von Biedersee and Pichler, 1995; Martinez-Villegas, 1999) were mapped at the north, southwest and southeast rim of the inferred caldera of Pamatian et al. (1992).

around the inferred ancestral Canlaon caldera (Pamatian *et al.*, 1992). Collapse structures (Umbal and Arboleda, 1987; PHIVOLCS, 1987; von Biedersee and Pichler, 1995; Martinez-Villegas, 1999) were mapped at the north, southwest and southeast rim of the inferred caldera of Pamatian et al. (1992).

These structures were interpreted in this study as volcanic eruption-related craters and small-scale calderas such as the Mansipang Collapse, Tubidiao Collapse, Hardin Sang Balo crater, Margaha crater ("Lugad crater" of von Biedersee and Pichler, 1995), Makawili crater and the active Canlaon crater.

3.3 Canlaon Volcanic History

Spectral and morphologic evidences; such as polarized signatures, tone, texture and pattern (from the polarized synthetic aperture radar imagery) in combination with superposition sequences (using the digital elevation model of topographic SAR) of volcanic deposits were proven essential in interpreting a multiple vent source for the Canlaon Volcanic Complex. However, the lack of absolute dating of the various deposits is evident in the area. The available dates of surface deposit were from charcoal samples from two different pyroclastic flow units obtained a ^{14}C dating of 1920 ± 50 and 2570 ± 50 years BP (Fig. 4). These pyroclastic flow units could be associated with the active Canlaon crater and Tubidiao crater, respectively (Martinez-Villegas, 1999).

A modification of the volcanic history for the Canlaon Volcanics earlier proposed by Pamatian et al. (1992) is given below (from oldest to youngest) (Figs. 5I):

Canlaon Volcanics 1 (KN-1) - Initially, an ancestral Canlaon volcano is believed to be situated within the present location of inferred ancestral Canlaon. A single-vent series of eruption events led to the breaching of its western flanks. The volcanic events were thought to be confined within a single vent source that deposited Kn-1. At present the remnants of the past activities are still extant at the northeast, south, and southwest. The paroxysmal stage of the volcanic activity was presumably marked by an explosive event that led to the breached west flank. A resulting debris avalanche deposit occupies the southwest sector of the study area. This unit is characterized by its prominent hummocky topography with blocks and matrix facies. The present location of these deposits at La Castellana suggested that the deposits traveled a maximum distance of 33 km and has an estimated volume of 13 km^3 (Catane, 1994).

Canlaon Volcanics 2 (KN-2) - Contiguous (post, intra-caldera) volcanism ensued within the breached ancestral caldera producing Makawili and Mansipang volcanic centers. Lava and pyroclastic flows associated with Makawili occupy the eastern midslopes and lower slopes of the ancient caldera while Mansipang produced mainly airfall tephra and minor lava flows which are deposited to the southwest. A dome-like structure proximate to the resulting Mansipang collapse can be observed in the area.

Hardin Volcanics (HN-1) - After the eruptions (post, extra-caldera) of both Mansipang and Makawili, eruption at the northern flank of the breached Canlaon caldera ensued. This resulted to the Hardin Sang Balo crater with mainly ashfall and pyroclastic flow deposits distributed to the northeast of the study area.

Hardin Volcanics (HN-2) - A renewed eruption at the Hardin Sang Balo crater caused breaching of its northwest crater wall thereby depositing mainly distal pyroclastic deposits and near-vent lava flows to the north. Three prominent dissected dome-like structures are contained within these deposits (e.g. Saray and Kumalisquis domes?). These structures are inferred to be post-Hardin Sang Balo parasitic domes.

Hardin Volcanics (HN-3) - Thick, voluminous lava flows accompanied by minor airfall deposits predominate the volcanic terrain sourced from the same vent of Hardin Sang Balo crater. Two lava flow events are believed to represent these deposits to the northwest and north-northwest. The former probably has a basaltic andesite composition while the latter is of dacitic in composition. This is based on the observed low- to high-relief lava flow morphology of both deposits as seen on radar imagery. The intermediate lavas have traveled over a longer distance and exhibit a widened lobe structure (less viscous) while the silicic lavas formed short to near-vent deposits and show a constricted lobe features (viscous lavas).

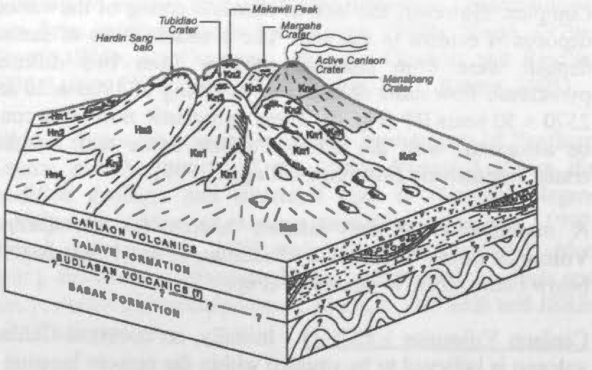

Makawili Peak
Tubidiao Crater
Margaha Crater
Hardin Sang balo
Active Canlaon Crater
Mansipang Crater
Kn2
Kn3
Kn4
Kn1
Hn1
Hn2
Hn3
Hn4
CANLAON VOLCANICS
TALAVE FORMATION
BUDLASAN VOLCANICS (?)
BASAK FORMATION
?

Figure 5. Three-dimensional Canlaon volcanic terrane showing Hn- and Kn- volcanic units. Subsurface geology of NNGP also shown (modified after Pamatian et al., 1992).

Hardin Volcanics (HN-4) - This volcanic sequence includes pyroclastic deposits believed to be transported along narrow valleys within Hn-3 units thereby emplacing its products along distal slopes northwest of the study area

Canlaon Volcanics 3 (KN-3) - Two volcanic sequences (e.g. Tubidiao and Margaha) correspond to this contiguous eruption event that produced KN-3. The distribution of the volcanic deposits in the area suggests that the Tubidiao collapse, an inferred volcanic center, deposited the pyroclastic flows as infills along valleys of previously breached western flanks of the ancient caldera. The Margaha crater (termed Lugad crater by Martinez-Villegas, 1999) influenced lava building (as near-vent deposits) that preceded the deposition of pyroclastic flows along narrow valleys of same Kn-3 lava banks mostly concentrated to the south - southeast slopes.

The Tubidiao pyroclastic deposits correspond to the andesite block-rich flows to the west at Guintubdan area as mapped by Martinez-Villegas et al. (1999) at its west slopes. A ^{14}C dating of charcoal sample obtained an age of 2570 ± 50 years BP. The Margaha volcanic center was the result of eruption at the southern slope of Makawili (Kn-2) thereby creating 1.5 to 2 km diameter crater (Pamatian *et al.*, 1992; von Biedersee and Pichler, 1995).

Canlaon Volcanics 4 (KN-4) - This volcanic event marks the most recent (historical) eruption deposits consisting of mainly pyroclastic flows and airfall tephra from the active Canlaon crater 100 meters to the south of the 1.5-2 km diameter Margaha crater (Fig. 5). However, it is observed that this airfall tephra thinly blankets the underlying lava flows to the south based on the analysis of radar imagery of the Canlaon study area. The pyroclastic flows initially described as valley in-fills correspond to the scoria-rich deposits identified by Martinez-Villegas et al. (1999) at Biak-na-Bato area to the south. A ^{14}C dating of charcoal sample gave a younger age of 1920 ± 50 years BP for this pyroclastic flow unit.

Two-isolated pyroclastic flow units (probably associated with Kn-4) have been similarly identified in this study along the general area of Talongo and Enyawan rivers. These deposits correspond to the pumice-rich pyroclastic flows overlain by andesite block-rich pyroclastic flow deposits mapped by Martinez-Villegas et al. (1999). These pumice-rich pyroclastic deposits could have been slightly older deposits of Kn-4. Unfortunately, no data are available to absolutely constrain the age of the deposit.

4.0 STRUCTURAL CONFIGURATION VERSUS WELL DATA

The structural analysis done here supplements the structural map of Panem and Leynes (1996). Lineaments delineated from the radar imageries (e.g. TOPSAR: digital elevation model and JERS-1 SAR) and the NAMRIA-extracted digital topography map (DEM) was initially compared with the 1996 structural map. Commonly placed or redundant lineaments and spatially unique lineaments were filtered and retained from the resulting composite map. Since the study is based on remote sensing, the observed unique linear features (other than the ground-verified ones) were designated initially as lineaments. These new lineaments were given separate names while the existing names of structures defined by Panem and Leynes (1996) were retained. These new fault traces are Mambucal-B and -C, Asia-C, Asia Splay-A, -B, -C and -D, Catugasan-B Splay, Catugasan-C and -D, Kinabkaban-B, Kinabkaban-C and Pataan-C Splay. Most of these faults (e.g. Asia Splay -A, -B, -C and -D, Kinabkaban-B, Catugasan-C, Kinabkaban-C, Catugasan-B Splay and Mambucal-B) were intersected in the recently drilled wells PT-5D and PT-6D (Fig. 2 and 6).

The latest temperature well profile of well PT-5D further confirmed that the hottest region in the area of northern Negros lies east and beneath wells PT-2D, PT-4D and PT-5D. In the absence of actual tracer tests to approximate fluid flow paths, it is surmised that the channels are predominantly fault- related and to a minor extent interformational- to intraformational-related permeable zones (PNOC-EDC, 2001).

Assuming a N-S stress regime dominates the Negros Island, the E-W and N-S faults comprise the compression and eastern extensional faults, respectively. Due to the movements along the Mindoro–Panay collision zone, a postulated clockwise rotation of the stress regime from N-S to SE (~135°) at its southeastern fringes possibly caused the current structural configuration of the study area. This theory explains opening up of previous N-S fault networks previously correlated with the eastern - subducting Negros Trench where the volcanic

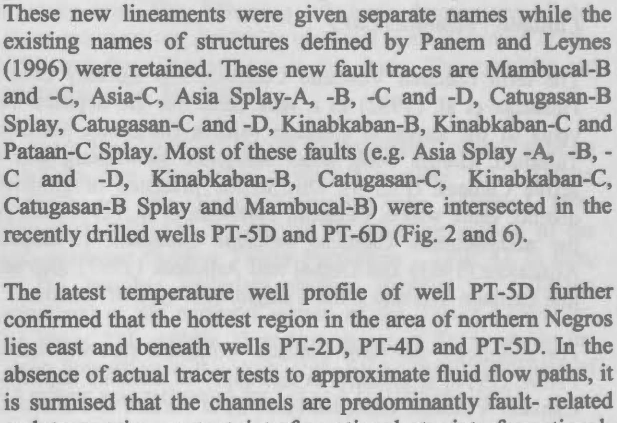

Catugasan C F.
Catugasan D
Catugasan B
Simaguan D
Mambucal C F.
Napatagan A
Napatagan B
Asia Splay B F.
Asia Splay D F.
Napatagan D
5D
Mambucal B F.
Catugasan B Splay
Simaguan B Splay
Simaguan B F.
Mambucal A Fault
Proposed Power Plant
Napatagan C F.
1D
2D
Asia Splay C F.
Asia Splay A F.
Kinabkaban A
PT
4D
6D
Kinabkaban B
Kinabkaban C
Kinabkaban Warm spring
Dinagan
Pataan-A F.
Pataan-B F.
Pataan-C F.
3D

Figure 6. Composite structural map of the Northern Negros Geothermal project (NNGP) showing the structures.

occurrences at Canlaon, Mandalagan and Silay are associated. With the postulated clockwise stress rotation, newly created NW structures might have hosted hydrothermal fluids associated with the volcanic episodes at Mount Canlaon. These hypotheses are supported by scientific findings of previous workers as shown below.

The subduction of the Sulu Sea Basin lithosphere along the Negros Trench was active since Miocene (Hamilton, 1979; Cardwell et al., 1980) and dip at a depth of 150 km beneath Negros (Acharya and Aggarwal, 1980) based on earthquake hypocenter data. The continuity of the Negros Trench (to the south) was consequently disrupted Late Miocene by the Mindoro-Panay and Sulu-Zamboanga collision zones (Barrier et al., 1991) which corresponds to the regional N-S compression phase. This phase is supported by clockwise rotation of paleomagnetic orientation, northeast fold axes and southeast-directed thrust axes observed in Panay, Negros and Cebu noted southeast of the Mindoro-Panay collision zone (McCabe et al., 1985). Consequently, a reported Plio-Pleistocene N-S directed extension might have accommodated volcanism at Negros where Canlaon, Mandalagan and Silay volcanic centers are associated with subduction processes at the Negros Trench (Aurelio et al., 1990) A localized rotation of stress fields are evident at the Northern Negros where NW structures was also observed to cut pre-existing NNW, WNW and NE structures.

Interpreted lineaments from the three imageries (e.g. TOPSAR / JERS-1 SAR, NW- and SE- directed hillshading maps) provide better control and information of the fault kinematics within the study area. It is observed that NW- and WNW-trending fault sets dominate and represent the active fault regime at NNGP (Fig. 6). The observed order of fault events is as follows (from young to old): NW, WNW / ENE, NNW and NE. The E-W and N-S structures are re-activated fault structures.

5.0 DISCUSSION

Previous works (Alincastre, 1983; PHIVOLCS, 1987; Umbal and Arboleda, 1987; Pamatian *et al.*, 1992; Von Biedersee and Pichler, 1995; Catane, 1994; Martinez-Villegas et al., 1999) and this study confirm a multiple vent source for the deposits of Canlaon and Hardin Volcanics. A possible shift in magma composition with time (e.g. from basaltic andesite to andesite to dacite and vice-versa) is also interpreted by Martinez-Villegas et al. (1999). Pamatian et al. (1992) and this study consistently subdivided the volcanic deposits according to their respective vent sources (e.g. from oldest to youngest; KN-1, KN-2, Hn-1, Hn-2, Hn-3, Hn-4, Kn-3 and Kn-4). The source vents are the Ancient Canlaon caldera (?), Hardin Sang Balo crater, Tubidiao crater collapse, Margaha crater and the active Canlaon crater, respectively. Based on the recorded but incomplete eruption history of Canlaon from 1866 to 1989 (COMVOL, 1981; Simkin, 1981), the generally phreatic type of eruption (some strombolian and phreato-magmatic) could only be associated with its latest stages Kn-3 and Kn-4 deposits. Charcoal dating (^{14}C) (Martinez-Villegas et al., 1999) on block-rich and scoria-rich pyroclastic flows of Tubidiao and the active Canlaon crater, indicate ages of 2570 ± 50 years BP and 1920 ± 50 years BP, respectively (Fig. 4).

With knowledge of the stratigraphy, volcanic history and structural networks within the study area, one can surmise a potential geothermal resource evident within the Northern Negros Geothermal Project (NNGP). It is believed that a nearby heat source to the southeast of the Pataan sector of NNGP could be associated with Hardin Sang Balo crater. The possible relationship of Saray dome / collapse to the ESE and Tubidiao crater / collapse to the south has been discounted as possible heat source for NNGP (PNOC-EDC, 2001). Hence, seepage of fluids are postulated to have been channeled to the northwest mainly along well - intercepted permeable major northwest-trending faults and splays with minor contribution of fluid flows along interformational- and intraformational-permeable zones within contacts of Canlaon Volcanics and Talave Formation. The location of the Mambucal thermal area, Hacienda Paz and Hacienda Montilla springs to the northwest are mapped surface manifestations of this brine leakage. Petrologic, geochemistry and geophysical data supports this postulation (PNOC-EDC, 2001).

ACKNOWLEDGEMENTS

We would like to thank NASA – Jet Propulsion Laboratory, Philippine Institute of Volcanology and Seismology (PHIVOLCS), PCASTRD and Earth Remote Sensing and Data Analysis Center (ERSDAC) for allowing us to process and interpret the Airborne Synthetic Aperture Radar (AIRSAR) and JERS-1 SAR data sets, respectively, of Mt. Canlaon area.

REFERENCES

Acharya, H.K. and Aggarwal, Y.P. (1980). Seismicity and Tectonics of the Philippine Islands. Journal of Geophysical Research, Vol.85 no. B6, 3239-3250.

Alincastre, R.S. (1983). Geology of Northern Negros geothermal project: Mambucal, Hagdan and Saray areas. PNOC-EDC Internal Report.

Aurelio, M.A, Barrier, E. and Rangin, C. (1990). Polyphase tectonics in a region traversed by the Philippine Fault: a microtectonic study. J. Geol. Soc. Philippines, Manila, 45, 31-50.

Aurelio, M.A, Huchon, F., Barrier, E. and Gaulon, R. (1990). Displacement rates along the Philippine Fault estimated from slip – vectors and regional kinematics. J. Geol. Soc. Philippines, Manila, 49, 65-78.

Bien, O.C., Esperidion, J.A. and Camit, R.A. (2001). Geological report of well PT-6D, Northern Negros geothermal project. PNOC-EDC Internal Report, 19 p.

Bureau of Mines and Geo-Sciences (BMGS) (1982). Geology and Mineral Resources of the Philippines, Vol. 1, Manila, 406 p.

Camit, R.A., Los Baños, C.F., Maneja, F.C. and Layugan, D.B. (2000). Remote sensing study of the Northern Negros geothermal project, Negros Occidental, Philippines. Proc. 13th Annual Geological Society of the Philippines (GSP) Geological Convention.

Cardwell, R.K., Isacks, B.L. and Karig, D.E. (1980). "The spatial distribution of earthquakes, focal mechanism solutions and subducted lithosphere in the Philippines and northeastern Indonesian Islands" in Hayes, D.E. (Ed). The Tectonics and Geologic Evolution of Southeast Asian Seas and Islands. Geophysical Monograph Series, vol. 23.

Catane, S.G. (1994). Features and nature of emplacement of debris avalanche deposit in the Philippines. Kobe University, Japan, D. Sc. Dissertation, 261 p.

COMVOL (Commission on Volcanology) (1981). Catalogue of Philippine Volcanoes and Solfataric Areas. COMVOL, Quezon City.

Curlis, J.D., Frost, V.S. and Dellwig, L.F. (1986). "Geological Mapping Potential of Computer – enhanced Images from the Shuttle Imaging Radar: Lisbon Valley Anticline, Utah," Photogrammetric Engineering and Remote Sensing, vol. 52, no. 4, 1986, pp. 525-532.

De Guzman, A. (1995). Microanalysis of well MC1 core. DOE report.

De Guzman, A. and Revilla, A. (1995). Micropaleontologic analysis of wells HG1D and CT1D cuttings. DOE report.

Dulce, R.G. (2001). Petrologic summary of well PT-5D. PNOC-EDC internal report, 8 p.

Dulce, R. G. (2001). Feb. 9 Petrologic summary of well PT-5D. PNOC-EDC Int. Rpt., 8 p.

Esperidion, J. A., (2001). Feb. 13 Preliminary well geologic report PT-5D, Northern Negros geothermal project. PNOC-EDC Int. Rpt., 10 p.

Esperidion, J.A., Caranto, J.E.A. and Camit, R.A. (2001). Well PT-5D geological report, Northern Negros geothermal project. PNOC-EDC Internal Report, 33 p.

Ferrer, H.P. (1979). Evaluation of stage-1 drilling, Northern Negros geothermal project. PNOC-EDC Internal Report, 55p.

Ferrer, H.P. (1980). Regional geology of Northern Negros: Its implication to geothermal development. PNOC-EDC Internal Report, 29p.

Hayashi, M. (1996). Report on the zircon analysis of drill cuttings from the Northern Negros geothermal project, Mt. Kanla-on, Negros Occidental. PNOC-EDC Internal Report.

KRTA (1980). Report on Northern Negros geothermal exploration. KRTA Unpublished Report.

Lubas, L.L., De Guzman, A., and Revilla, A. (1997). Micropaleontological analysis of well PT-3D cuttings. DOE report.

Lubas, L.L. (2001). Results of the micropaleontological analysis of samples from wells PT-5D and PT-6D. Department of Energy Report, 3 p.

Martinez-Villegas, M.L., Bornas, M.A.V. Abigania, M.I.T., Listanco, E.L. Quiambao, R.B. and Sikat, B.S.J. (1999). Stratigraphic relationship of eruptive products from Canlaon Volcano, Negros Island, Philippines: Implications for hazard assessment, 9 p.

Melendres, M.M. (1975). Local stratigraphy of Northern Negros. Bureau of Mines and Geo-Sciences.

Ordoñez, E. (1979). Paleoanalysis of core #2. Mambucal-1 Internal Report.

Pamatian, P.I., Salonga, N.D. and Tebar, H.J. (1992). The geology of the Northern Negros geothermal prospect, Mt. Canlaon, Negros Occidental. PNOC-EDC Internal Report, 60 pp.

Panem, C.C. and Leynes, R.D.C. (1996). Faults and lineaments in Northern Negros geothermal area. PNOC-EDC Internal Report.

PNOC-EDC (1995). Draft report on the preliminary resource assessment of the Northern Negros geothermal project, Mt. Kanla-on Negros Occidental, Negros Occidental, Philippines. PNOC-EDC Internal Report. 231 p.

PNOC-EDC (1998). Northern Negros geothermal project - resource assessment update. PNOC-EDC Internal Report, 139 pp.

PNOC-EDC (2001). Northern Negros geothermal project - resource assessment update. PNOC-EDC Internal Report, 126 pp.

Pornuevo, J.B. (1988). Lineaments and structures of Canlaon volcano and vicinity. PNOC-EDC Internal Report.

Ramos, S.G. (2001). Petrologic report: well PT-6D (Original Hole). PNOC-EDC internal report. 9 p.

Ramos, S.G. (2001). Petrologic report: well PT-6D (Sidetrack Hole). PNOC-EDC internal report. 6 p.

Reyes, A.G. (1979a). Geology of Mambucal-1 exploratory well, Mambucal, Murcia, Negros Occidental. PNOC-EDC Internal Report.

Reyes, A.G. (1979b). Petrographic analysis of cores 1-3 from MC-2A exploratory well. PNOC-EDC Internal Report.

Reyes, A.G. (1984b). Notes on the petrography of Northern Negros surface samples collected by R.S. Alincastre. PNOC-EDC Internal Report.

Simkin T., Siebert, L., McClelland, L., Bridge, D., Newhall, C. and Latter, J.H. (1981). Volcanoes of the world. Smithsonian Institution, Hutchinson Ross, Pennsylvania.

Umbal, J.V. and Arboleda, R.A. (1987). Report of investigation for the semi-detailed mapping and hazard risk assessment of Canlaon volcano. PHIVOLCS Unpublished Internal Report, 27 p.

Umbal, J.V. and Arboleda, R.A. (1988). Debris avalanche deposit of Canlaon volcano. Proceedings, Kagoshima International Conference on Volcanoes, 1998, pp. 463-465.

UNITAR/UNDP Centre on Small Energy Resources (1991). Application of Geochemistry in Geothermal Reservoir Development (co-ordinated by Franco D'Amore), U. S. A, 408 pp.

Von Biedersee, H. and Pichler, H. (1995). The Canlaon and its neighbouring volcanoes in the Negros belt. Philippines, Journal of Southeast Asian Earth Sciences, vol. 11, no. 2, pp. 111-123.

Wood, C.P. (1979a). Petrology of cores and cuttings from Mambucal-1. DSIR/KRTA Internal Letter Report.

A VIEWER INTERFACE FOR INTERACTIVE THREE-DIMENSIONAL GEO-DATA VISUALISATION

Admire Kandawasvika[1], Wolfgang Reinhart[1] and Michael Hahn[2]

[1]University of Armed Forces, Munich Neubiberg
[2]University of Applied Sciences, Stuttgart

Commission IV, WG IV/6

KEY WORDS: Geodata, 3-D visualisation, interactive, viewer, interface

ABSTRACT:

Advancements in three-dimensional (3-D) geographic data acquisition techniques has led to continuous growth of geographic information. The challenges imposed are how to handle such data in a more effective way. A research has been carried out which focused on 3-D visualisation of geodata in an interactive mode. Since human needs differ in terms of culture and exposure, which in turn affect how data is perceived by such different people, there is a variety of specific 3-D visualisation application requirements. Henceforth, there will be need for different software interfaces for different geodata applications.

The main objective of this research was to develop a viewer interface or a simple viewing system for the purpose of visualising parts of interest of a Stuttgart city model. Realisation of this task guides the way to subsequent, specific applications, which can be developed according to the user's or customer's needs. There are some complex visualisation systems on the market but most applications can not be implemented fully without some customisation or new interface development.

In order to develop a viewer interface, which provides interactive functionality and dynamic levels-of-detail (LOD); dynamic movements and weather simulations (for inducing the feeling of an environment), an existing software development kit (SDK) from a company called Skyline Software was used. Microsoft Visual C++ and MFC programmming, together with TerraSDK, were used for the development of a first viewer interface. The interface enables the following functions: a user can move dynamically within the scene (virtual city), weather simulations such as snow and rain can be generated, 3-D objects (i.e. billboards or labels, buildings, trees, etc.) can be loaded on terrain or removed from the scene at will; selection and 'jumping' to places of interest is possible, route-following, and observing objects of interest from four cardinal positions (i.e. north, east, west, and south). The whole interface implementation was realised using an ActiveX control.

Furthermore, this research showed the feasibility of developing viewer interfaces for specific applications with the customer or user in mind. Currently, more applications of this nature are developed and realised on the Internet. Factors such as data streaming and data traffic are crucial for the success of this interfaces. Three-dimensional visualisation will integrate more geographic imaging fields such as digital photogrammetry, remote sensing, GIS, etc. There will be continuing demand in realisation of specific viewer interfaces in the fields of forestry, real-estates marketing, virtual tourism, geology, archaeology, and other geodata marketing fields.

WEB-BASED MULTIMEDIA GIS FOR THE ANALYSIS AND VISUALIZATION OF SPATIAL ENVIRONMENTAL DATABASE

Shunfu Hu

Department of Geography, Southern Illinois University Edwardsville, Edwardsville, IL 62026, shu@siue.edu

Commission IV, WG IV/6

KEY WORDS: Multimedia GIS, Visualization, Internet, Everglades

ABSTRACT:

Recently, Web-based GIS applications have become important tools for disseminating geographical information in digital map format on the Internet because of their platform independence, interactivity, and wide accessibility. This trend brings about new insights on the integration of Web-based GIS and multimedia on the Internet since multimedia has already been seen as the strength of Internet applications.

This paper briefly examines the principles of Web-based GIS applications and multimedia technology, and then focuses on the development of a Web-based multimedia GIS that links GIS and multimedia technologies on the Internet. The Web-based multimedia GIS is based upon interactions between three components: 1) a Web-based GIS application developed to manipulate spatial environmental database including georeferenced images and digital maps; 2) a Web-based interactive multimedia application designed to manipulate multimedia information including text, graphics, photographs, digital video and sound; and 3) a mechanism of linking the Web-based GIS application and the Web-based interactive multimedia application. The paper concludes with an example that demonstrates the use of a Web-based multimedia GIS, which allows the user to explore digital maps and associated multimedia information on the Internet.

1. INTRODUCTION

Recently, there has been new trend in developing Web-based GIS applications. Researchers in this field have recognized Web-based GIS applications are becoming an important tool to disseminate geographical information on the Internet because of their platform independence, interactivity, and wide accessibility. Those applications can be generally categorized into two types: (1) to allow various end-users to access spatial database and conduct simple spatial analysis (Peng and Nebert, 1997; Abel et al., 1998; Sadagopan et al., 2000); and (2) to allow end-users to perform sophisticated spatial analysis for decision-making (Peterson 1997; Doyle et al. 1998, Peng 1999, Dragicevic et al. 2000; Meyer et al. 2001; Zhang and Wang 2001).

Web-based GIS applications are based upon the interactions between client and server computer systems through network technology (Figure 1). The client side allows Internet users to access remote computers on the Internet by providing request through standard Web browser software such as Microsoft Internet Explorer, Netscape Navigator, or other custom-generated software such as Environmental System Research Institute (ESRI) ArcExplorer.

The network technology provides Internet software components that communicate with each other on various computers connected by the network. Those components include HTTP (i.e., hypertext transfer protocol), and TCP/IP (transmission control protocol / Internet protocol). Protocols are the language that makes Internet communication possible.

The advancement of computer technology enables the integration of GIS and multimedia technologies, called multimedia GIS, that allows to incorporate not only spatial and/or temporal geographic information in image/vector format but also multimedia geographic information in descriptive text, scanned ground photographs, graphics, digital sound and video (Hu, 1999; Hu, et al., 2000; Hu, 2001; Hu, 2002a; Hu 2002b).

The development of GIS database for environmental management often generates not only digital data sets in vector/image format, but also multimedia information in 1) text (descriptive text, narrative and labels); 2) graphics (drawings, diagrams, charts, snapshots or photographs); 3) digital video (television-style material in digital format); 4) digital audio sound (music and oral narration); and 5) computer animation (changing maps, objects and images). It is challenging to bring data sets in various formats into a coherent computer system that allows end-users such as policy maker, resource manager and the public with different computer skills to analyze and visualize them.

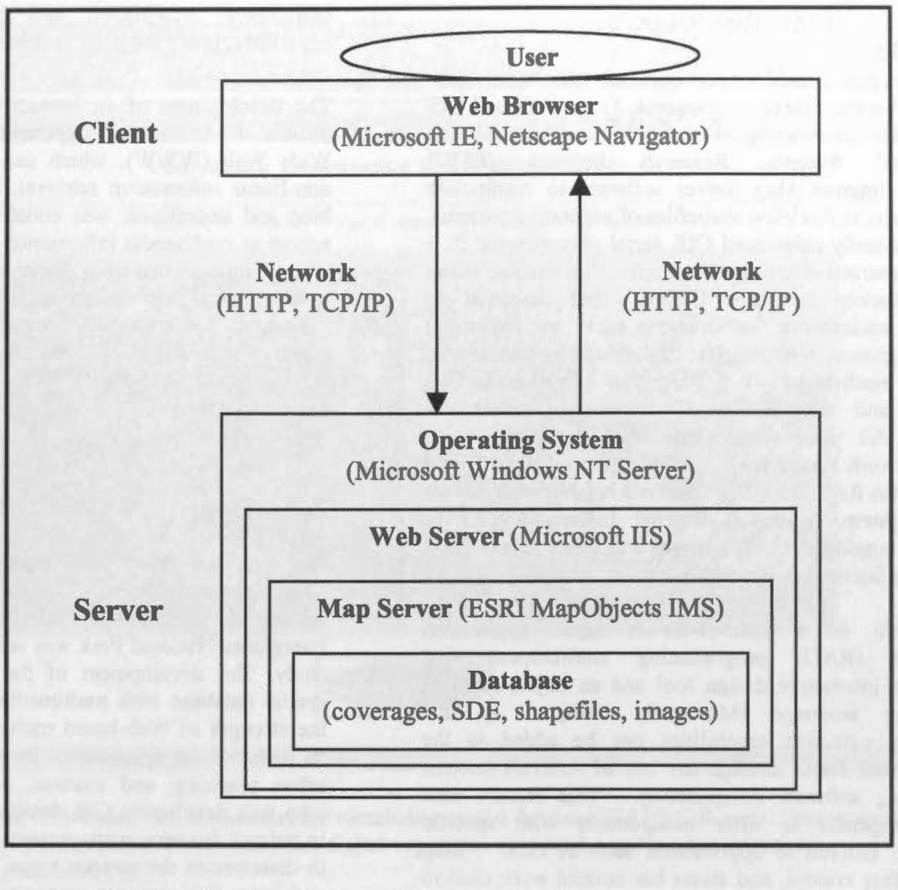

User
Web Browser
(Microsoft IE, Netscape Navigator)
Client
Network
(HTTP, TCP/IP)
Network
(HTTP, TCP/IP)
Operating System
(Microsoft Windows NT Server)
Web Server (Microsoft IIS)
Server
Map Server (ESRI MapObjects IMS)
Database
(coverages, SDE, shapefiles, images)

Figure 1. Web-Based GIS application architecture

2. RESEARCH OBJECTIVE

The objective of this research is to develop a Web-based multimedia GIS approach that: 1) is interfaced with GIS for the analysis and visualization of spatial GIS database in vector/image format; 2) permits the combination of GIS database with multimedia information 3) allow the user to access the databases via standard Internet explorer such as Microsoft Internet Explorer or Netscape Navigator.

The study area is corresponding to the U.S. Geographical Survey (USGS) Long Pine Key and Pa-Hay-Okee Lookout Tower 1:24,000-scale topographic quadrangles in Everglades National Park (Figure 2). A GIS database includes detailed vegetation maps, and scanned National Aerial Photography Program (NAPP) 1:40,000-scale color-infrared (CIR) aerial photographs, which are the data source for the development of digital vegetation maps. The multimedia database contains descriptive text, ground photographs, digital video clips and audio segments highlighting the characteristics of Everglades natural environment (i.e., climate, landform and hydrography) vegetation plant communities, individual species, and invasive exotics, as well as plant-animal interactions, hurricane damage, and post-fire vegetation succession.

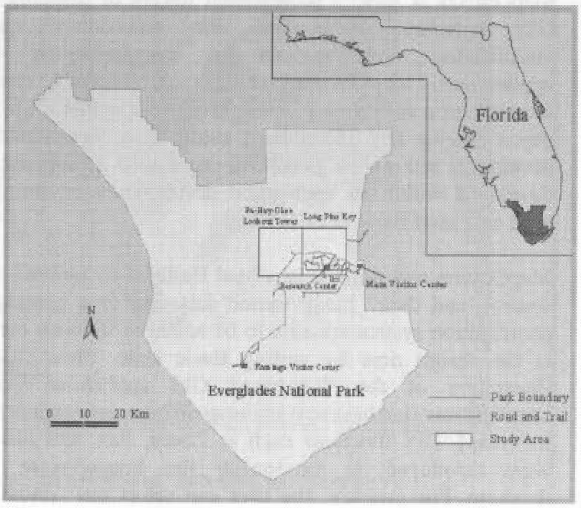

Florida
Pa-Hay-Okee
Lookout Tower
Long Pine Key
Research Center
Main Visitor Center
N
Flamingo Visitor Center
Everglades National Park
0 10 20 Km
Park Boundary
Road and Trail
Study Area

Figure 2. Study area: Long Pine Key (right) and Pa-Hay-Lookout Tower (left) quadrangles in Everglades National Park.

3. METHODOLOGY

The web-based multimedia GIS approach was based upon interactions between three components: 1) a web-based GIS application developed using Microsoft Visual Basic 6.0 and Environmental Systems Research Institute (ESRI) MapObjects Internet Map Server software to manipulate spatial data sets in ArcView shapefiles of vegetation patterns, and geographically referenced CIR aerial photographs; 2) a web-based interactive multimedia application created using hypertext markup language (HTML) and designed to manipulate multimedia information such as hypertext, hyperlinks, ground photographs, digital video, and sound; and 3) a mechanism of linking the Web-based GIS application and the Web-based interactive multimedia application. All three components were developed on a Compaq network server with 1.4 GHz microprocessor, and 256 magebytes RAM, running Windows NT Network Server operating system. Microsoft Internet Information Server (MIIS) was employed as Web server and ESRI MapObjects Internet Map Server 2.0 as map server.

Visual Basic is a state-of-the-art rapid application development (RAD) programming environment that combines an interactive design tool and an object-oriented programming language (Microsoft, 1999). In this environment, extended capabilities can be added to the standard Visual Basic through the use of ActiveX custom controls (i.e., software components). This feature also permits companies to offer components with specific functionality tailored to applications such as GIS. Map control, toolbar control, and status bar control were used to develop the GIS application and graphical user interface, whereas text control, picture control, time control, and scroll bar control were used to develop the interactive multimedia system. The underlying relation among various custom controls was manipulated using Basic language. The following sections discusses the development of the three components mentioned above.

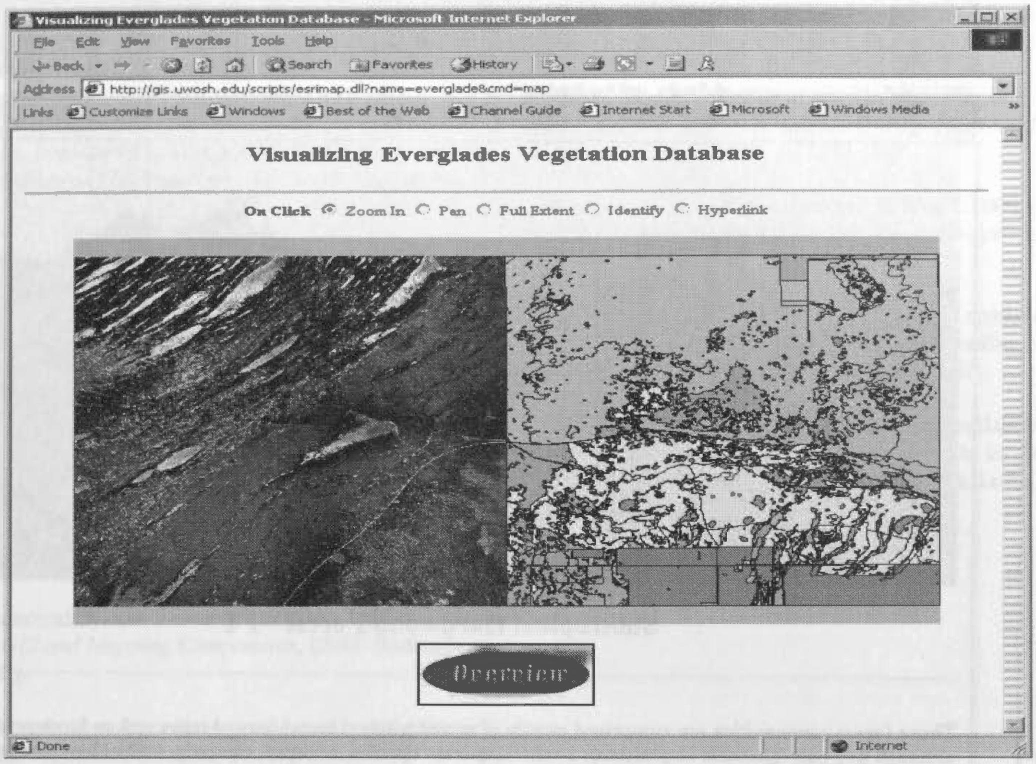
MapObjects is ESRI's professional library of mapping and GIS software components for embedding spatial visualization and analysis in custom-written GIS applications. MapObjects includes an ActiveX custom control and a collection of over 35 programmable OLE (i.e. object linking and embedding) automation objects that let developers add mapping and GIS capabilities to applications developed within an application development environment such as Visual Basic (ESRI, 1996).

MapObjects was utilized in Visual Basic as an ActiveX map control, and then image/vector data tied to a geographic coordination system are able to be added to the map control at the design time by writing Basic code. Plate 1 is an illustration of the Web-based GIS application (URL: http://gis.uwosh.edu/scripts/esrimap.dll?name=everglade&cmd=map). GIS functions such as Zoom, Pan, and Identify were developed to manipulate the image/vector GIS database. For instance, the user can select the "Zoom in" option, and zoom into an area on the digital map, then select the "Identify" option, click on any polygon, and the alphabetic letters representing the dominant vegetation category in that polygon will be displayed on the screen (Plate 2). Furthermore, the user can select "Hyperlink" option and click on one polygon, he/she will be directed to a ground panoramic view of that plant community, followed by a interactive multimedia application containing multimedia information and including hypertext and hyperlinks, about the plant community (Plates 3).

The development of an interactive multimedia application utilized the concepts of hypertext and hyperlinks on World Wide Web (WWW), which can provide interactivity and non-linear information retrieval. A hypertext, displayed in blue and underlined, was constructed to allow interactive access to multimedia information in the form of descriptive text, ground photographs, digital video, and sound. There are numerous ways to design excellent Web-based interactive multimedia applications for use on the Internet. Commercial software includes HTML, Macromedia Director and Dreamweaver, and Adobe Streaming Media Collection, to name just a few.

4. CONCLUSIONS

The main purpose of this study was to provide the user with user-friendly tools to visualize both spatial data and associated multimedia information on the Internet. Everglades National Park was selected as a test area for this study. The development of the new approach to linking spatial database with multimedia information demonstrated the strength of Web-based multimedia GIS technology and its potential for applications in environmental management, urban planning, and tourism. Agencies and organizations who task developing GIS databases, especially applications in natural resource management, can employ this technique to disseminate the various types of collected information to end users (i.e, resource managers, planners, and the public) with different levels of computer skill. Further research will focus on incorporating more GIS analytical functions and modeling capability.

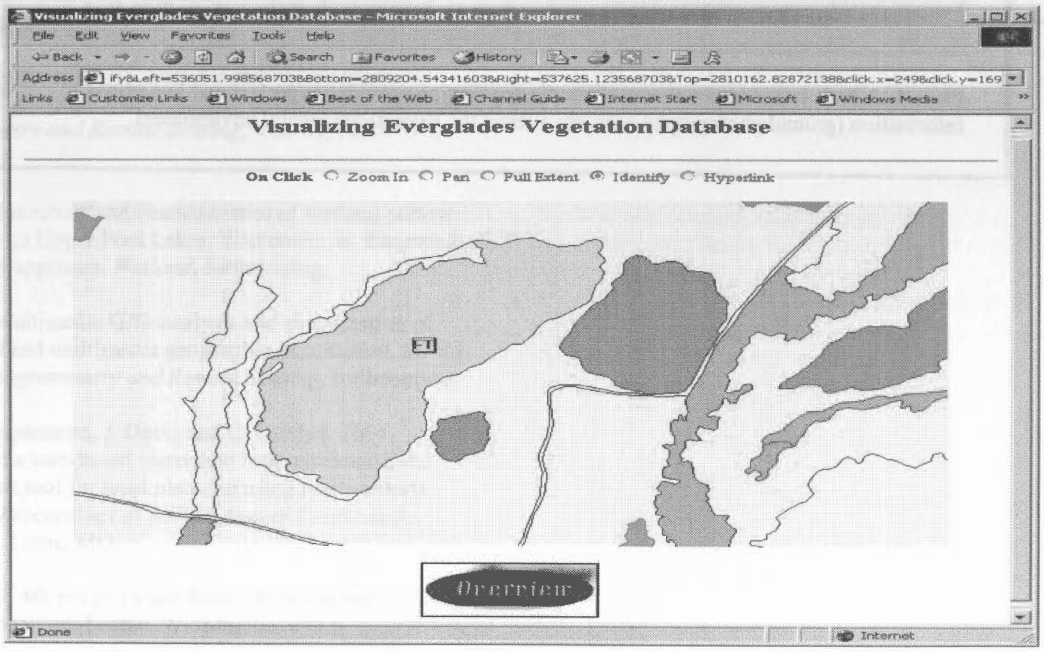

Visualizing Everglades Vegetation Database - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Links Customize Links Windows Best of the Web Channel Guide Internet Start Microsoft Windows Media
Overview
Done
Internet

Address http://gis.uwosh.edu/scripts/esrimap.dll?name=everglade&cmd=map

Visualizing Everglades Vegetation Database

On Click ● Zoom In ○ Pan ○ Full Extent ○ Identify ○ Hyperlink

Plate 1. The Web-based GIS application for visualizing spatial database: NAPP CIR aerial photographs for Pa-Hay-Okee Lookout Tower (left) and digital vegetation map for Long Pine Key (right) in Everglades National Park.

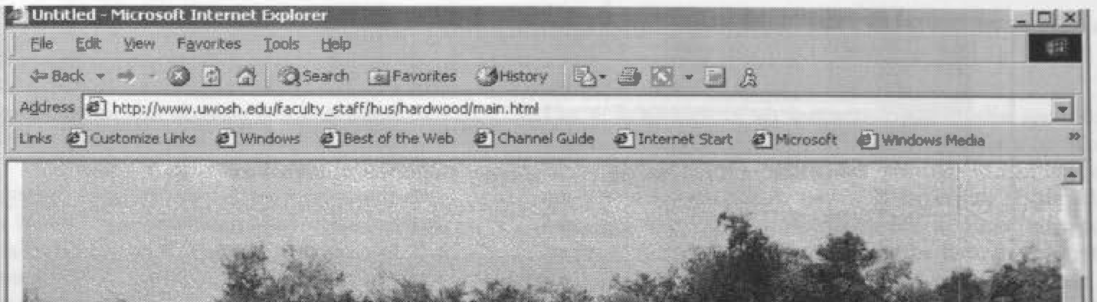

Visualizing Everglades Vegetation Database - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Links Customize Links Windows Best of the Web Channel Guide Internet Start Microsoft Windows Media
E1
Overview
Done
Internet

Address ify&Left=536051.998568703&Bottom=2809204.543416038&Right=537625.123568703&Top=2810162.82872138&click.x=249&click.y=169

Visualizing Everglades Vegetation Database

On Click ○ Zoom In ○ Pan ○ Full Extent ● Identify ○ Hyperlink

Plate 2. The Web-based GIS functions allow the user to manipulate spatial data sets, and to retrieve the dominant vegetation category for a selected polygon from the GIS database.


Untitled - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Address http://www.uwosh.edu/faculty_staff/hus/hardwood/main.html
Links Customize Links Windows Best of the Web Channel Guide Internet Start Microsoft Windows Media
Done
Internet

Subtropical Hardwood Forest - FT

These forest communities are comprised mostly of mixed tropical broad-leaved trees and understory shrubs, although live oak (*Quercus virginiana*), a sub-temperate species, occurs frequently. Called hammocks in the local vernacular, their canopies are typically closed. This leads to more humid conditions and less temperature extremes than in other plant communities. For this reason, orchids and bromeliads, as well as brillantly colored tree snails, are common. This community is also important for migrating warblers and other birds including the resident barred owl. Hammock trees seldom attain heights of more than 15 or 20 meters, perhaps largely because of the peneplanning effects of hurricanes. For example, the tallest pine tree at the

Plate 3. A Web-based interactive multimedia application provides hypertext and hyperlinks to explore multimedia Information (ground photographs, text, etc.) about the Everglades natural environment.

REFERENCES

Abel, D., T. Kerry, R. Ackland and S. Hungerford. 1998. An Exploration of GIS Architectures for Internet Environments, *Computer, Environment and Urban Systems*, 22(1): 7-23.

Bill, R. 1994. Multimedia GIS – Definition, Requirements and Applications, *European GIS Yearbook*, Blackwell, Oxford, pp. 151-154

Doyle, S., M. Dodge and A. Smith. 1998. The potential of web-based mapping and virtual reality technologies for modeling urban environments, *Computer, Environment, and Urban Systems*, 22(2): 37-55.

Dragicevic, S., Balram, S., Lewis, J. 2000. The role of Web GIS tools in the environmental modeling and decision-making process. *4th International Conference on Integrating GIS and Environmental Modeling(GIS/EM4): Problems, Prospects and Research Needs*. Banff, Alberta, Canada, September 2 - 8, 2000, http://www.colorado.edu/research/cires/banff/upload/363/

ESRI (Environmental System Research Institute). 1996. *MapObjects - GIS and Mapping Components*, ESRI: Redlands, California, 202 p.

Hu, S. 1999. Integrated multimedia approach to the utilization of an Everglades vegetation database, *Photogrammetric Engineering and Remote Sensing,* 65(2), 193-198.

Hu, S., A. O. Gabriel, and C. Lancaster. 2000. An integrated multimedia approach for wetland management and planning of Terrell's Island, Winnebago Pool Lakes, Wisconsin, *The Wisconsin Geographer,* 15-16, 34-44.

Hu, S. 2001. Multimedia GIS: analysis and visualization of spatio-temporal geographic information, *International Achieves of Photogrammetry and Remote Sensing*, Vol. 34, Part 4/W5, pp. 42-47

Hu, S., 2002a. Inventory and characteristics of wetland habitat on the Winnebago Upper Pool Lakes, Wisconsin: an integrated multimedia GIS approach, *Wetland*, forthcoming.

Hu, S. 2002b. Multimedia GIS: analysis and visualization of spatio-temporal and multimedia geographic information, *ISPRS Journal of Photogrammetry and Remote Sensing*, forthcoming.

Meyer, J., R. Sugumaran, J. Davis and C. Fulcher. 2001. Development of a web-based watershed level environmental sensitivity screen tool for local planning using multi-criteria evaluation. In: *Proceedings of ASPRS Annual Conference*, April 23-27, St. Louis, MO.

Microsoft. 1999. *Microsoft Visual Basic: Programmer's Guide*. Microsoft Cooperation: Seattle, Washington, 898 p.

Nielson, J. 1990. *Hypertext and Hypermedia*, Academic Press Professional, Boston, Massachusetts, 296 p.

Peng, Z. and D. Nebert. 1997. An Internet-Based GIS Data Access System, *Journal of the Urban and Regional Information Systems Association*, 9(1): 20-30.

Peng, Z. R. 1999. An assessment framework for the development of Internet GIS. *Environment and Planning B: Planning and Design*, 26:117-132.

Peterson, M. 1997. Trends in Internet Map Use. In: *Proceedings of 18th International Cartographic Conference*, ICA Stockholm, Sweden, pp.1635-1642.

Sadagopan, G., J. J. Richardson and R. Singh. 2000. Complying with the freedom of information act, *Geo Info System*, 10(1):28-32.

Zhang, X. and Y. Q. Wang. 2001. Web based spatial decision support for ecosystem management. In: *Proceedings of ASPRS Annual Conference*, April 23-27, St. Louis, MO.

Zhang, Z., J. Tsou and H. Lin. 2001. Virtual collaboration on visual impact assessment: a web GIS approach. In: *Proceedings of ASPRS Annual Conference*, April 23-27, St. Louis, MO.

AUGMENTED MAPS

Joachim Bobrich[*], Steffen Otto

Institute for Cartography and Geoinformatics, University of Hanover,
Appelstraße 9a, 30167 Hanover, Germany – (joachim.bobrich, steffen.otto)@ikg.uni-hannover.de

Commission IV, WG IV/6

KEY WORDS: Augmented Reality, Cartography, Mapping, Visualization, Real-time, Video, Combination

ABSTRACT:

Cartographic visualizations are usually given as printed maps or screen maps. On the one hand the user can directly interact with a flat paper-map in a natural and familiar way like all other physical things in our environment. On the other hand there is a possibility to extend the dimensions of a map to more then two by the means of computer-vision. Three-dimensional landscapes and animation can be created and visualized. The usability depends to a great degree on the experience of the user and the possibilities of the computer program. Augmented Reality (AR) combines these characteristics, in that it places well known ways of handling and all potentialities of computer-science to the users disposal. In our approach the actual position and line of vision referring to the map is determined by the use of pattern-recognition techniques. Unambiguous tags are placed on the border of the map. These tags are captured by a small video camera. The computer determines all parameters for overlaying and combining the video-stream of reality with artificial virtual objects. As this calculation can be performed individually for different users, they can look at the same data and see the combination of reality and virtual objects. So every user can interact with the map in his own way and e.g. activate new layers or point on virtual object to get additional information. This technique enables cartographic applications to visualize and work on digital geographic data based on an analogue map. As an example the visualization of a digital terrain model in combination with the corresponding topographic map is shown.

KURZFASSUNG:

Kartographische Darstellung erfolgen in der Regel auf einem analogen Träger oder in digitaler Form durch Bildschirmdarstellung. Während bei der Papierkarte die Interaktion in einer, dem Benutzer vertrauten Art und Weise erfolgen kann, besteht in der computerunterstützten Darstellung die Möglichkeit dreidimensionale Effekte durch virtuelle Landschaften zu erzielen. Die Bedienbarkeit hängt hierbei stark von der Erfahrung des Benutzers und den Möglichkeiten des Programms ab. Durch Augmented Reality werden die beiden Möglichkeiten vereint und ein System geschaffen, welches dem Anwender eine intuitive, und natürliche Bedienung erlaubt. Durch den Einsatz von Algorithmen zur Mustererkennung ist es möglich die Position und Blickrichtung des Betrachters relativ zu den Mustern zur bestimmen. Diese Muster sind an bekannten Positionen auf der analogen Karte angebracht und mithilfe einer kleinen Digitalkamera erfasst. Der Computer bestimmt die Parameter für eine Überblendung der erfassten Umgebung mit den virtuellen Objekten. Die individuelle Bestimmung der Darstellungsparameter für jeden einzelnen Benutzer erlaubt es, einen Datenbestand gleichzeitig aus unterschiedlichen Richtungen zu betrachten und damit zu interagieren. Als Interaktionsmöglichkeiten werden das Aktivieren zusätzlicher Themen über Schalter sowie die Funktion eines Zeigestabes erläutert. Für die Anwendung in der Kartographie ergeben sich so zusätzliche Möglichkeiten für die Darstellung und Bearbeitung digitaler raumbezogen Informationen auf der Grundlage einer analogen Karte.

1. INTRODUCTION

1.1 Overview

In cartography the definition of map is one of main issues and like in many other sciences these definitions are a subject of adaptation to actual demands. Professor Gao Jun presented in his keynote at the International Cartographic Conference in Beijing (Gao 2001) the different kind of maps. He classified maps as two branches of one tree. The first branch is the real (visible) map, this are maps or drawings displayed on the screen-printed on a media. The second branch is the invisible map, so mental maps or digital maps in a computer like virtual reality images.

Joining paper maps and virtual reality images will generate a new nature of map: the augmented map.

Augmented reality (AR) is defined as the mixture of sensory perception of the real environment and fictitious, virtual objects. Milgram (Milgram et al. 1994) shows augmented reality as a mixture between the real environment and virtual reality (Figure 1).

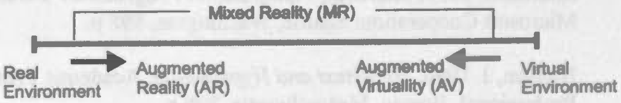


Figure 1. Milgram´s Reality-Virtuality Continuum.

To keep a realistic impression this combination has to be done in real-time. So it is possible to immerse and interact in a natural way with both, the environment and the virtual objects.

---

[*] Corresponding author

Augmented reality is used in many applications. Among those are medicine, car-navigation, energy supply, military (Piekarski et al. 1999) and data-visualisation in general (Vallino, 2002).(Azuma 1997) gives a survey in AR.

In cartography it can be used as a graphical user interface for spatial data.

In this way it is possible to visualize landscapes and other cartographic items in a graphically attractive and flexible way. Conventionally we can present digital landscape models in a perspective way on computer displays, but to get a representation of the terrain in which you can "dive in" and can perceive in a natural way, we have to apply physical models.

1.2 Related work

Through the use of computer-generated pictures, which can be overlaid on reality by a head mounted display (HMD) the viewer will get a stereoscopic impression not only of the real environment, but also of the artificial objects. Reality and virtuality will be melted and immersive interaction is possible.

As a prerequisite for an exact overlay of reality and augmented information, the exact viewing point and viewing direction have to be known. With these parameters we can establish the connection to additional digital geo-data and their thematic attributes. Through the simultaneous visualisation of the real environment and the three-dimensional view of geo-data the human perceptive facility is broadened. The determination of the users position and viewing direction can be accomplished in several ways:

- magnetic field sensors,
- positioning and orientation sensors (GPS, INS),
- or using image processing techniques.

(Broll et al., 2000) presented a virtual round table, where several participants can look at the same scenario on a desk. Tracking is made by an inertial tracking device and pattern recognition. Real Objects (cups) are placeholders as tangible interfaces. (Reitmayr et al., 2001) describe a mobile collaborative AR-System that supports true stereographic graphics, a pen, a pad interface and direct interaction with virtual objects. (Schmalstieg et al. 1996/2000) show the collaborative environment for shared space work in AR.

2. AUGMENTATION OF MAPS

To enrich the functionality and performance of printed maps, we have to examine the following points:

- Information we can handle by traditional maps
- Possible operations with analogous maps
- Additional (invisible) potential of digital information
- New optimised handling techniques
- Integration of these points in an AR-Map

Traditional maps are designed to visualize spatio-temporal phenomena in a flat, two-dimensional way. Cartographic methods have been developed to reach some pseudo-three-dimensional or pseudo-temporal effects. But they are all static and that's where the problem lies. The map cannot react on the users behaviour. There is no real interaction between user and map like in a virtual-reality-environment, where an action of the user causes reactions of the system.

The possible operations with analogous maps are limited to panning, rotation and zooming (approaching / getting off). Thematic selection of overloaded maps depends of the experience of the user (MacEachren 1995).

The enrichment of printed maps is the combination with dynamic, variable digital data. So processes can be visualized in real-time and, because of the variety in dimension, in more than two dimensions. Normally we use traditional maps as a close range application, outside or inside of buildings. We have to find a method to augment maps fitting all these needs. These demands are fulfilled be methods of pattern-recognition.

ARToolKit (ARToolKit 2002, Billinghurst et al., 2001) uses these techniques to determine the relative position of a video camera to a marker (square with unique patterns) in the video-stream. To speed up the process the video is turned into inverse black and white by a threshold operation. In this picture is then searched for square regions. All squares found are inspected for pre-trained patterns inside. If there is a match, a marker is found. Using the known pattern-size and detected orientation the position-matrix can be calculated. This matrix describes the real world coordinates of the real camera relative to the tag. This matrix is then used to set the position of the virtual display coordinates. Since the virtual display and real camera coordinates are the same, the graphic overlay is drawn precisely over the real pattern. An OpenGL API can be used to draw virtual objects in the virtual display. As the patterns are simple and easy to recognize, this operation can be executed in real-time. The algorithm runs with at least one pattern. Figure 2 describes the working method of ARToolKit (© Kato et al. 2000)

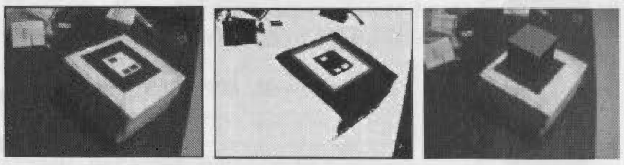


Figure 2. Working method of ARToolKit.

To realize an ARToolKit-application at least a video camera and a computer connected to it is necessary. By using a head mounted display (HMD) a more immersive experience will be created (Figure 3). So video capture and video and VR-object output are done in the same place. HMDs can be used in opaque or transparent/see-through-mode. In opaque mode the video and virtual data will be displayed. In see-through mode only the virtual objects are shown and the user can see the virtual objects overlaid the real environment.

The patterns are used as identical points to georeference the physical map and the artificial objects. So we can merge real and virtual information. The layout of an augmented map is shown in Figure 4. The actual map is bordered with four patterns. So the viewer can look at the map, always grasping one reference point to virtual reality at least. In order to recognize the pattern we have to pay attention to keep distance to other graphical elements.

As a concrete example the archaeological map of a castle is taken. Every summer our students survey historical locations within their courses. One result is the digital terrain-model of the surrounding of the castle. The graphic outcome is a new

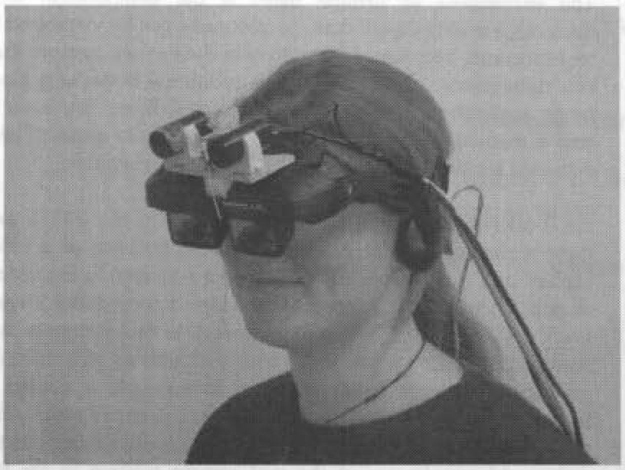


Figure 3. Cameras and HMD in see-through-mode.

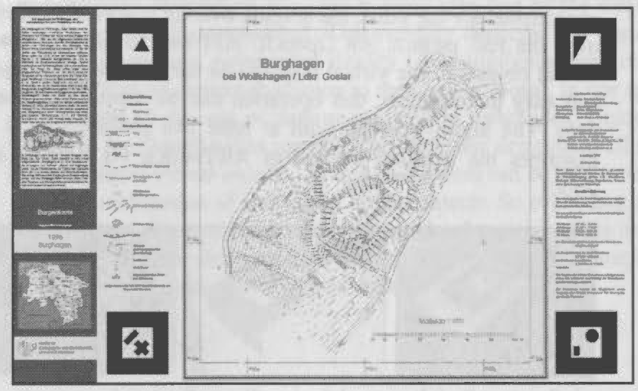


Figure 4. Layout of augmented map.

archaeological map. In the new map-design this map is in the middle of the sheet. If folded, you can see the blue title page. It is on the most left hand side and illustrates the greater location and gives a brief historical outline of the object. The legend between the left patterns describes the map-style in detail. The left text-block contains - apart from geodetic information - the local coordinates of the pattern and the local and global coordinates of one point of reference in the map. Assuming that the local coordinates are parallel to the global coordinates, digital terrain-models and other digital geodetic data can be matched to the analogous map (Figure 5 and Figure 6).

Since four patterns refers to only one virtual object and real-time processing is necessary the position of this object is determined by the first corresponding pattern.

At this stage of research the paper map is connected to digital data. Terrain-models can be shown on top of analogous representations. A further extension is in functionality. Map-users can interact by switching themes or layers, using an augmented signal disc. In our examples it would be useful to show all sites of archaeological discovery.

Repeated recognition of the pattern on the disc will on the one hand change the appearance of sphere on it and on the other hand alter the virtual theme on the AR-map (Figure 7). Themes can be switched on or off; symbolization of virtual map-object can be changed and so on.

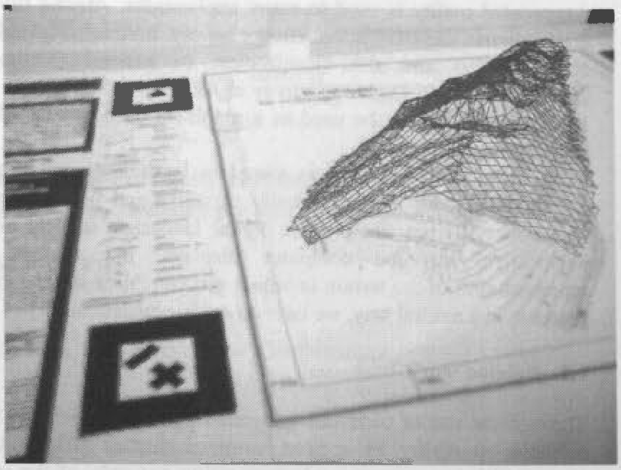


Figure 5. Overlay with 3D-TIN.

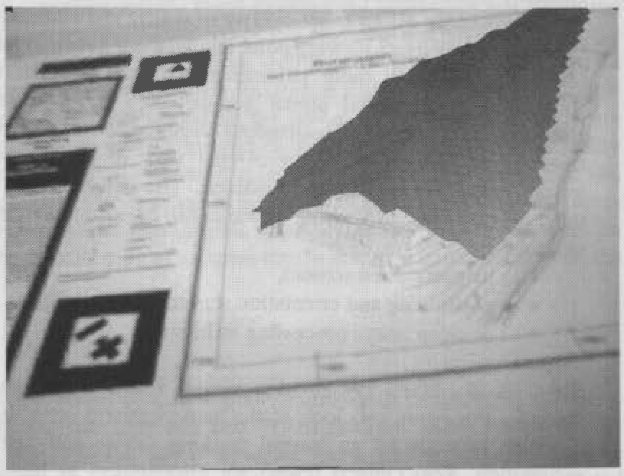


Figure 6. Overlay with DEM.

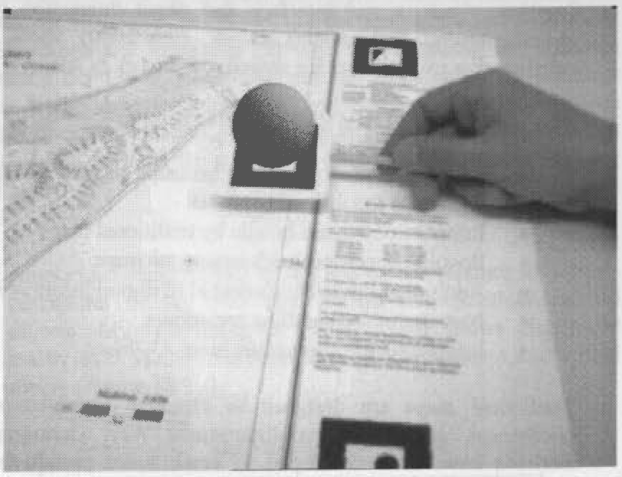


Figure 7. Augmented switch (red=off, green=on).

Another option is also a pattern as a proportional input device, which will modify the appearance of a theme in accordance with the discs angle of inclination. Linking the time index to the pattern can control the course of animations.

Additional information can be obtained by using an augmented pointer (Figure 8). Here virtual objects (visible or not) intersect with the virtual end of the pointer. Depending on the object and the position of the pointer further actions can be triggered.

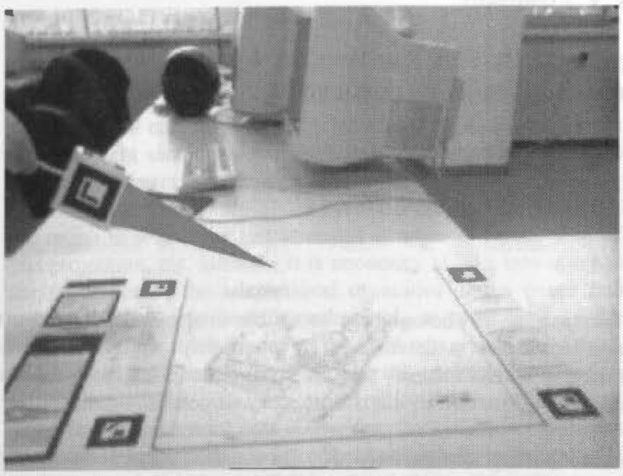


Figure 8. Augmented pointer.

In contrast to other immersive visualisation-methods the information can be projected into the HMD's of several users individually, thus this system greatly enhances collaborative work.

3. CONCLUSION

Pattern recognition techniques are applied for the visualization and inspection of topographic maps. The patterns are placed in the unused parts of the maps, so that the map can be utilized conventional as well. But the connection of the analogue map and a 3D digital landscape model allows for a three-dimensional presentation on top of the map. The user can interact with this model like with an analogue model: move and rotate it and change the point of view and distance. The only restriction is the recognition of at least one pattern in his field of view.

The advantage of this kind of data-visualisation is the possibility of natural interaction between user and system. Different operations are possible. Whereas rotation and movement of all degrees of freedom can be done in both, reality and virtuality, more interactions like zooming, scaling, selection of attributes, change of style, selection of themes and animation are possible.

These functions can be activated by different ways. The view of the real and virtual objects is changeable by the position and viewing direction of the user. Apart from keyboard-input virtual objects can be manipulated by virtual sensors to change their representation. The extension of this concept is that virtual objects can intersect with other virtual objects and cause a reaction.

A further improvement would be markers made with ink in the invisible spectrum. Special cameras could detect the pattern without disturbing the user of the map.

An augmented map represents a further step in interaction with digital cartographic data. Analogue maps are mixed with digital models. An immersive working method is possible for every single user of the system simultaneously. Furthermore by the Internet it does not matter where this user is located; several users can manipulate their own themes and place them into the common virtual environment.

REFERENCES

ARToolKit: A software library for Augmented Reality Applications. Copyright (C) 2002. Hirokazu Kato, Mark Billinghurst, Ivan Poupyrev. WWW: http://www.hitl.washington.edu/resarch/shared_space/download

Azuma R., Baillot Y., Behringer R., Feiner St., Julier S., MacIntyre B., 2001. *Recent Advances in Augmented Reality.* IEEE, Nov. 2001, pp 34-45

Azuma R, 1997. *A Survey in Augmented Reality.* Presence – Teleoperations and Virtual Environment. Vol. 6, no. 4, Aug. 1997, pp. 255-385.

Billinghurst M., Kato H., Poupyrev I., 2001. *The MagicBook: Moving Seamlessly between Reality and Virtuality* In IEEE Computer Graphics and Applications, May/June, pp. 2-4.

Broll W., Meier E., Schardt Th., 2000. *The Virtual Round Table*, Proceedings of ACM Collaborative Virtual Environment (CVE 2000), pp 39-46, San Francisco

Gao Jun, 2001. *Virtual Terrain Environment - a New Annotation for Maps and the Cartography.* In Proceedings of ICC 2001,Aug 6th. -10th, Beijing.

Kato H., Billinghurst M., Poupyrev I., Imamoto K., Tachibana K., 2000. *Virtual Object Manipulation on a Table-Top AR Environment.* In Proceedings of ISAR 2000, Oct 5th-6th.

MacEachren, A.M. 1995. *How Maps Work.* The Guilford Press, New York

Milgram P., Kishino F., 1994. *A Taxonomy of Mixed reality Visual Displays*, IEICE Trans. Information Systems, vol. E77-D, no. 12, 1994, pp. 1321-1329

Piekarki W., Bernhard, G., Bruce T. 1999. *Integrating Virtual and Augmented Realities in an Outdoor Application,* Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, 20 - 21 October, 1999 San Francisco, California

Reitmayr G., Schmalstieg D., 2001. *Mobile Collaborative Augmented Reality,* In Proceedings of International Symposium on Augmented Reality (ISAR), pp 114-123, New York, Oct. 29-30 2001

Schmalstieg D., Fuhrmann A., Hesina G., 2000. *Bridging multiple users interface dimension with augmented reality.* In Proceedings of International Symposium on Augmented Reality (ISAR), pp 20-29, Munich, Germany, Oct. 5 -6 2000. IEEE And ACM

Schmalstieg D., Fuhrmann A., Szalavari Z., Gervautz M., 1996. Studierstube – an environment for collaboration in augmented reality. In Proceedings of Collaborative Virtual Environments Workshop '96, Nottingham, UK, Sept 19-20 1996

Vallino J., 2002. Augmented Reality Homepage – Introduction http://www.cs.rit.edu/~jrv/research/ar/introduction.html (accessed 19.02.2002)

MODELLING OF VISUALIZATION

L. Prisniakova, V. Prisniakov

Institute Geotechnical Mecanics , Dnipropetrovsk, Ukraine, Kazakova No. 4, kv. 30,
Tel/Fax: 380+562+463430, Email:< kprisn@a-teleport.com>

Commission IV, WG IV/6

KEY WORDS: Information, Visible image, Interpreter, Stimulus brightness, Adaptation, Weber's fraction, Choice response.

ABSTRACT:

The pattern recognition, determination of its shape, dimension, location by its photographic or visible image is depended on the activity of image interpreter.. Improvement of activity of image interpreter is determined by possibility of its mathematical description. The known mathematical model of information processing in human memory was drawn on description some actions of the image interpreter. The starting-point is a equation of processing of information by human memory. It connects a magnitude of information in memory I with the time τ and the rate of information inputting into memory $\dot{R}$. On the theoretical solution of this equation the dependencies on stimulus brightness and the received sensation, the description of darkness and light adaptation; the relationship between the threshold sensitivity and stimulus magnitude, the dependence of the choice response time and probability appearance of signal were obtained. Other effects such as: the dynamics of the magnitude of preadaptation on the shape of the adaptation curves; the effect of luminance of different colors on the adaptation curve; the effect of adaptation to light on subsequent darkness adaptation were analyzed. The formula for the Weber's fraction were obtained which opens new possibilities of the analytical assessment of the processes, connected with human sensor systems sensitivity The resulting formula describes its influence of degree of brightness, the size of the test-object, the degree of contrast between the irritant and background. In the special case that the operator has alternative stimulus the choice response time is added by latent period of the simplest response and the choice response time respective.

1. INTRODUCTION

The pattern recognition, determination of their shapes, dimensions, location by their photographic or visible images are depended on the activity of image interpreter. He has obtained the information, has processed hers, has remembered hers and has level with image in his memory. Visibility (or probability of detection) is depended on the brightness of background and object in him, on level of visual adaptation, on stay in field of vision. Improvement of activity of image interpreter is determined by possibility of its mathematical description.

2. MATHEMATICAL MODEL

2.1. Original equation

Process of pattern recognition is defined as the processing of information by human memory in relation to ***equation:***

$$dI/d\tau = \dot{R} - (I-\varphi)/T \quad (1)$$

We obtained, on the theoretical solution of this equation,

$$I=\varphi+ \dot{R}\,T +(I_0 -\varphi - \dot{R}\,T) \exp(-\tau/\varphi) \quad (2)$$

the dependencies on stimulus brightness and the received sensation, the description of darkness and light adaptation; the relationship between the threshold sensitivity and stimulus magnitude, the dependence of the choice response time and probability appearance of signal. The aim of our report is to show the possibility of mathematical modeling of pattern recognition as information processing in human memory.

2.2. Adaptation

If we interpret the changing sensitivity of receptors during ***adaptation*** like the fading of information in memory, we can describe the process of adaptation by using a formula for information retention in memory. We derived this formula from a differential equation (1) (or equation (2) under $\dot{R} = 0$) describing the conservation of the flow of information

$$\bar{I} = \bar{\varphi} + (1- \bar{\varphi}) \exp(-\tau/T) \quad (3)$$

where $\bar{I}=I/I_0$ is relative sensitivity of the receptor; I is current value of information about stimulus intensity; I_0 is the initial value of information about the intensity of the stimulus; $\bar{\varphi} = \varphi/I_0$; φ is the minimum value of the amount of information retained in human memory after a sufficiently long period of time; (it is acquired by memory through life experience and by genetic endowment; a consequence of both nature and nurture). Let me show you how this theory of loss information predicts all adaptation data for light and darkness adaptation. Different experimental conditions alter the values of the parameters T (a memory information_processing time constant, determined experimentally) and $\bar{\varphi}$. Let us examine the application of equation (3) in more detail and investigate the effect of various parameter values for T and $\bar{\varphi}$. Comparison of analysis of theoretical and experimental data shows that as luminance, B_d, increases, both the time constant of the process of light adaptation, T, and the parameter φ decrease. Computations of the experimentally derived values of T and φ made it possible to plot relations between $T(B_d)$ and $\bar{\varphi}(B_d)$ which can be approximated in the range of values of B_d by the following functions:

$$T = 3 - (1/3)\ ln(B_d),\quad \bar{\varphi} = b_1 + b_2\ ln(B_d),$$

where $b_1 = 0.25$; $b_2 = -0.04$.

The curvature of the lines characterizing the relation between light sensitivity and darkness adaptation depends on a number of factors. For example, the degree of luminance during the preliminary exposure of the eyes to light, on the section of the retina to which the test stimulus is applied, on the area of this projection, etc. Besides, it is necessary to take into account the difference of the adaptational capacities of the cones and rods. It has already been hypothesized, as mentioned as the result of the initial luminance of the light perceived by the eyes. One can use formula (3) for predicting the process of adaptation to darkness. The comparison of the predictions against numerous experimental data shows that the process of darkness adaptation can be computed precisely enough from formula (3) under an average value of *T*; the values of *T* and $\bar{\varphi}$ for cones and rods being different. For example, the comparison of Piper's experimental data against the predicted results based on formula (3) shows a satisfactory fit. It is worthwhile nothing that when τ=10 min the experimental data show a discontinuity which corresponds to the time lag between adaptation periods of cones and rods. Both sections are approximated very well but for different values for the rods and cones.

The analysis of the effect of duration, τ_0, of preadaptation exposure to luminance, on the shape of the adaptation curve that is on the values of *T* and $\bar{\varphi}$ yield the formulas

$$T = a_1 \sqrt[4]{\tau_0}\ ;\quad \bar{\varphi} = a_2\ ln(2+\tau_0/\tau_0)+a_3$$

If dark adaptation follows a period of long adaptation to a certain color, the numerical values of *T* and $\bar{\varphi}$ change. In this case the calculated predictions fit the experimental data better under an assumption of linear changes in the variables

$$T = T_0 + \theta\tau.$$

In this case the solution of the equation for the retention of information in sensory memory yields:

$$\bar{I} = \bar{\varphi} + (1 - \bar{\varphi})(1+\theta\tau/T_0)\ exp(-1/\theta) \quad (4)$$

We can also analyze other effects such as:

-**the dynamics of the magnitude of preadaptation on the shape** of the adaptation curves; computations from (3) yielded the average values of variables *T* and $\bar{\varphi}$ which satisfactorily fit the experimental data (under the same value of pre-exposition brightness $I_0 = 30\ mlb$: nominator is retinal cones, denominator is rod cells):

I_0	*4700*	*447*	*44*	*4*
T, min	*2.2/2.8*	*1.3/2*	*1.8*	*1.5*
$\bar{\varphi}$	$0.11/4.710^{-4}$	0.01	0.01	0.1

Table 1 Magnitude T and $\bar{\varphi}$ as function of pre-exposition brightness of white light

- **the effect of luminance of different colors** on the adaptation curve; results of computations of experimental data (from formula (3) are

Colour	***T***, min	$\bar{\varphi}$
Violet	1.35	0.022
Yellow	1.3	0.063
Red	0.91	0.094
White	*1.35*	*0.034*

Table 2.

- **the effect of adaptation to light** on subsequent darkness adaptation is mainly reflected in the values of $\bar{\varphi}$, while the values of T_0 and θ change insignificantly.

3. THE GENERAL PSYCHOPHYSICAL LAW

The problem of the relation between magnitudes of sensation and stimulus is solved in the general form for the first time. The solution the differential equation (1) gave the general Psychophysical Law:

$$I = (1+ p)/(1 - exp(-pR))/p \quad (5)$$

This exponential curve describes the experimental data with brightness as stimulus well.

4. THE HUMAN SENSOR SYSTEM SENSITIVITY

Difference in brightness implies the existence of not only absolute but discrimination thresholds as well. Brightness range of light varies from the absolute lightness threshold (10^{-5} -- 10^{-6} cd/m^2) to the very high values (up to 10^6 cd/m^2). Visual sensations of the observer may be described in terms of the absolute brightness of a stimulus and the relative brightness of light stimuli which can vary.

Solving equation for threshold values we have the formula for the Weber's fraction

$$\Delta R_0/I = ln(1+p)/(1+p)(1- exp(-pR/\varphi)), \quad (6)$$

where ΔR_0 is the sensory discrimination threshold.

This formula opens new possibilities of the analytical assessment of the processes, connected with human sensor systems sensitivity.

The analysis of our computations showed that the influence of difference factors can be estimated both qualitatively and quantitatively with the help of equation (6). This influence is reflected, first of all, in the values of φ, p, $\dot{R}\,T$. The analysis of the experimental data obtained showed that φ, p, $\dot{R}\,T$ values depend upon the light wave length λ. The corresponding relations between φ and $\dot{R}\,T$ variables and λ are

$$\varphi = \varphi_1\ exp(\theta\tau),\quad \dot{R}\,T = a + b\lambda,$$

where $\varphi_1 = 6.6\ 10^{-8}$ unit, $\theta = 0.025\ 1/s$; $a = 32un$; $b = -0.03\ 1/nm$.

This formula is obtained by approximation of the computations based on experimental data.

5. THE TIME OF PATTERN REGOGNITION

In the special case that the operator has alternative stimulus the choice response time is added by latent period of the simplest response and the choice response time respective. Solving Eq. (1) we have the formula

$$\bar{\tau}_p = 1 + \ln[1+p(1/W-1)]/\ln(1+p) \qquad (7)$$

The theoretical predictions have been checked. This theory works well for mathematical modeling real situation and extends possibility study on function of image interpreter.

CONCLUSION

The work shows the prospects of mathematical description of process of pattern recognition by man as information processing in human memory. Theoretical formulas predicts the adaptation processes for light and darkness adaptation; they can also analyze of luminance of different colors on the adaptation curve; the effect of adaptation to light on subsequent darkness adaptation; theoretical relations describe the influence of degree of brightness, the size of the test-object, the degree of contrast between the irritant and background.

NOMENCLATURE

$\bar{I} = I/\varphi$;

I – the amount of information;

φ -the amount of information retained by memory after sufficiently long period of time;

W – probability appearance of signal;

$\dot{R}$ – the rate of information imputing into memory;

τ - time;

$\bar{\tau}_p = \tau_p/\tau^*$;

τ^* -latent period of the simplest response; $p = \varphi/\dot{R}\,T$;

T is memory information processing time constant, determined experimentally (it depends first of all on the individual operator's characteristics, condition of his work, information forthcoming rate etc);

ΔR_0 is the sensory discrimination threshold; I_0 is the initial value of information about the intensity of the stimulus.

References

1. Prisniakov V. F., Prisniakova L. M. Mathematical modelling of information processing by the operators of the man – mashine systems. – Moscow. Machinebuilding Publishers. 1990.
2. Prisniakov V. F., Prisniakova L. M. Mathematical modelling of emotions// Kibernetic and System analysis. – Institut Kibernetic National Academy of Science of Ukraine. - #1, 1994.
3. Link S. W. The wave theory of difference and similarity.- Hilsdale. –New Jersey. 1992.

APPLICATION OF DIFFERENTIAL SPLINES FOR RELIEF SIMULATION

Khrystyna Burshtynska, Alexandr Zayac

State University "Lviv Polytechnika", Ukraine
bursht@polynet.lviv.ua

Working Group IV/6

KEY WORDS: DEM, differential spline, relief, collocation, Kriging method, variogram, contour.

ABSTRACT

For automatic map-making comprising as partial processes the isoline map compiling, three-dimensional surface image construction as well the additional characteristics that are different from them, it is necessary to create a mathematical model of the relief.

Such modeling depends on the means of presenting the output information as well as the approximation methods and the surface interpolating.

The output information about the relief is usually presented in a discrete form with regular, semi-regular and irregular arrangement of points.

The most frequently used construction methods of DEM are known to be as follows: polynomial method, multiquadrics method, spline method, distance-weighting method on the basis of triangulation, Kriging method.

The result of our research is as follows: for effective approximation of relief a mathematical model is constructed on the basis of collocations when the locality surface is approximated by low degree polynomial with additional interpolating the differential spline.

The experimental research for various types of relief proves that a highly accurate reproduction of the surface, excluding places with a certain difference of inclinations with the approximation error is about 1/2 - 2/3 of relief cross-section.

1. INTRODUCTION

Digital Elevation Models (DEM) have an sugnificant importance for solving a series of the engineering type applied problems or as the basis for creating joint information model of terrain. DEM is applied in computer map-making, which consists of contour map compiling, three-dimensional surface model construction, profile lines, geting slope and aspect maps, erosion processes.

DEMs for huge teritories, even for whole states are created with data digitized from contours. In a large scale, applied problems, which accuracy depends on input map accuracy, solves with DEM.

2. THEORY OF DEM CONSTRUCTION

DEM building accuracy depends on such main factors as: input information accuracy and presenting methods and mathematical relief presentation methods. The actual question is accuracy of DEMs based on contours scanned from maps.

Maximal automatization is acheeved using map scanning.

Relief is presented in discrete form with points arrangement at contours and specific places.

Input information presenting methods cause less opposite thoughts than mathematical relief modelling methods. The most frequently used methods for DEM construction are: polynomial method, spline method, multiquadrics method, distance-weighting method, triangulation method, Kriging method.

It was stated from our research that model based on mean square collocation [Moritz H., 1983] is efficient for relief modelling. The main equation of this model for relief approximation is:

$$Z = AX + s + n, \tag{1}$$

where Z - elevation value vector; s - relief component after trend removal; n - random value or noise dependent from measurement error. The expression

$$T = AX \tag{2}$$

presents systematic component or trend and approximated by polynomial of low power.

Since measurements are carried out not on terrain but on map, so when simplifying mathematical model we'll consider $n=0$.

In mean square collocations method the signal is presented by covariance function dependent from distance between points. The expression for signal evaluation will be:

$$\hat{s}_k = C_{ki} \cdot C_{ii}^{-1} \cdot \Delta Z_i \tag{3}$$

where i= 1,2,... ,n - number of input points. Or, in expanded record

$$\hat{s}_k = \begin{bmatrix} C_{k1} \\ C_{k2} \\ \vdots \\ C_{kn} \end{bmatrix}^T \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} \Delta Z_1 \\ \Delta Z_2 \\ \vdots \\ \Delta Z_n \end{bmatrix}, \tag{4}$$

where $\hat{s}_k$ - estimate of signal at k point, which elevation is definad; C_{ki} - vector of covariances between defined and input points; C_{ii} - matrix of covariances between input points; ΔZ_i – vector of differences at input points between measured elevations and the approximated expression Eq. 2.

$$\Delta Z_i = Z_i - T_i. \tag{5}$$

This model was realised in Kriging method with such the difference: covariances were changed by semi-variances and covariance function was changed by variogram [Burshtynska Kh., Tumska O., Lelukh D., 2000]. The expression Eq. 4 by Kriging will be:

$$\hat{s}_k = \omega_1 \Delta Z_1 + \omega_2 \Delta Z_2 + \ldots + \omega_n \Delta Z_n, \tag{6}$$

where ω_i – weights, found from solution of such system of equations:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{k1} \\ \gamma_{k2} \\ \vdots \\ \gamma_{kn} \\ 1 \end{bmatrix}. \tag{7}$$

In those equation values γ_{ij} – semi-variance, find from variogram [Davis J.C., 1977]

$$\gamma_{ij} = C_{ii} - C(R_{ij}), \tag{8}$$

where R_{ij} – distance between points.

Weights are found under condition $\sum\limits_{i=1}^{n} \omega_i = 1$.

Covariance function or variogram depends on relief type and defines its statistic characteristics. The most Kriging problem is variogram definition. It is defined a priori but during DEM construction variogram parameters changes with approximation field type.

As it is shown in [Moritz H., 1983] covariance function or variogram may be presented by different analitic basis functions to simplify the solution. Differential splines shows good approximation properties. For DEM construction we used such differential spline:

$$F = \sum_{i=1}^{n} \lambda_i R_i + a + bx_i + cy_i, \tag{9}$$

where

$$\begin{gathered} R_i(x,y) = [(x-x_i)^2 + (y-y_i)^2] \times \\ \times \ln[(x-x_i)^2 + (y-y_i)^2] \\ R_i(x_i, y_i) = 0 \end{gathered} \tag{10}$$

There are such conditions for coefficients definition λ_i, a, b, c:

$$\begin{gathered} F_i = Z_i, \qquad \sum_{i=1}^{n} \lambda_i = 0, \\ \sum_{i=1}^{n} \lambda_i x_i = 0, \quad \sum_{i=1}^{n} \lambda_i y_i = 0. \end{gathered} \tag{11}$$

Coefficients λ_i, a, b, c are defined from solution of equation system:

$$\left[\begin{array}{c|c} G & M \\ \hline M^T & \theta \end{array}\right] \left[\begin{array}{c} \lambda \\ \hline \tau \end{array}\right] = \left[\begin{array}{c} l \\ \hline O \end{array}\right], \tag{12}$$

where G – symmetric matrix of $n \times n$ size with elements $R_i(x, y)$;

M - matrix of $n \times 3$ size, which rows are filled with numbers 1, x_i, y_i;

θ - zero matrix of 3×3 size;

λ/τ - vector of defined coefficients λ_i, a, b, c;

I - vector of input data, $F_i = Z_i$;

O - zero vector with three rows.

To compare approximation properties excepts input function Eq. 9 we had issued modified function:

$$F' = \sum_{i=1}^{n} \lambda_i R_i^{'} + a + bx_i + cy_i, \tag{13}$$

where

$$\begin{gathered} R_i^{'}(x,y) = \sqrt{(x-x_i)^2 + (y-y_i)^2} \times \\ \times \ln\sqrt{(x-x_i)^2 + (y-y_i)^2} \end{gathered}. \tag{14}$$

3. EXPERIMENTS AND RESULTS

We choose the part of topography plan with scale 1:5000 cross-section 1 m, inclination angles from 0,1° to 10°, with different relief types (plains, soft slopes, slopes with sharp inclination jumps) for comparative analisys of presented relief approximation methods. We didn't took any information from the lower section because of the sharp bluff there.

Input information was received by map scanning and data digitized from contours becides specific points elevations record. Map scanning was done with 300 *dpi* resolution. Contours digitizing was done manual with mean step of 1 mm (step changes with contour curvature).

The programs for Krining method of relief modelling realization were build on *Object Pascal* language in visual programming environment *Borland Delphi 5.0*. We used wandering surface method to build DEM with the presented methods. We realized point search programm by radius and by direction (quadrant, octant).

During approximation methods programm realization for DEM building we determine that unknown find problem is unstable. This make impossible finding of equations inverse matrix by classic method. For solution stabilization we used regularization method by Tikhonow [Zhurkin I., Neyman J., 1988.]. This method main point is that we obtain inverse matrix of equations coefficients from relationship:

$$A^{-1} = \left(A + \alpha I\right)^{-1}, \qquad (15)$$

where α – regularization parameter, calculated as local field variance;

I – identity matrix.

Digital Elevation Model build with 25 m step, input points for approximation methods were defined uniformly: eight by octant.

DEM creation accuracy depends on variation curve form. See Fig. 1 for specified relief types variograms at sections (see Fig. 2): a) plain relief (section D); b) soft slope (section E), c) slope with sharp inclination jump (section A).

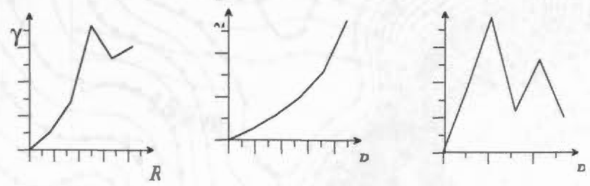


Figure 1. Variogram forms

Form of variograms *a* and *b* shows trend presence, at whole variograms have the soft character. Surface modelling by linear law is good after trend removal. Variogram *c* with sharp curve jumps indicates complicated relief character. We get the best approximation results from DEM building by Kriging method using such an exponential variogram:

$$\gamma(R) = kR^{0.5} \qquad (16)$$

where k – scale coefficient, defined through local field semivariance.

DEM construction accuracy evaluation was done: by divergence of contour input points elevations; by control points; visually, by divergence input contours and reproducted by DEM contours.

For accuracy estimation we used about 3500 input points, which elevations were defined by 4 nearest DEM points so far as section boundaries. Control points (points, which elevations are subscribed at topographical plan) are evenly distributed across the section. Their total number is 55. See Table 1 for results of accuracy estimation.

Table 1

DEM construction accuracy estimation

Method	By contour points, m			By control points, m		
	1	2	3	1	2	3
Differential spline	0.56	-0.85	0.09	0.28	-0.28	0.12
Modified differential spline	0.60	-0.89	0.09	0.27	-0.24	0.11
Kriging* (linear variogram)	0.68	-0.54	0.08	0.26	-0.23	0.12
Kriging** (linear variogram)	0.70	-0.68	0.10	0.26	-0.24	0.12
Kriging*** (linear variogram)	0.68	-0.54	0.08	0.26	-0.24	0.12
Kriging* (exponential variogram)	0.70	-0.68	0.10	0.32	-0.23	0.12
Kriging** (exponential variogram)	0.68	-0.54	0.08	0.32	-0.23	0.11
Kriging*** (exponential variogram)	0.70	-0.68	0.10	0.32	-0.22	0.12

* local trend is approximated by mean value;

** local trend is approximated by plane;

*** local trend is approximated by bilinear surface.

There are maximum and minimum elevations of points divergences in the columns numbered 1, 2 correspondingly, and mean square errors in the columns numbered 3.

Mean square errors pictures general view at approximation accuracy without picturing accuracy at the local sections. So, all methods look like uniformly by mean square errors value, but visual comparing of input contours and contours produced by DEM shows sugnificant divergences at the local sections.

Fig. 2 pictures the map of digitaled contours (pictured with points) and contours produced by DEM, constructed with Kriging method (solid lines); a) using exponential variogram (exponent power 0.5); b) using linear variogram. Divergence analisys shows inexact surface approximation when using linear variogram at sections with sharp jumps of inclination angles (sections A and B), where surface reproduction is more accurate when using exponential variogram.

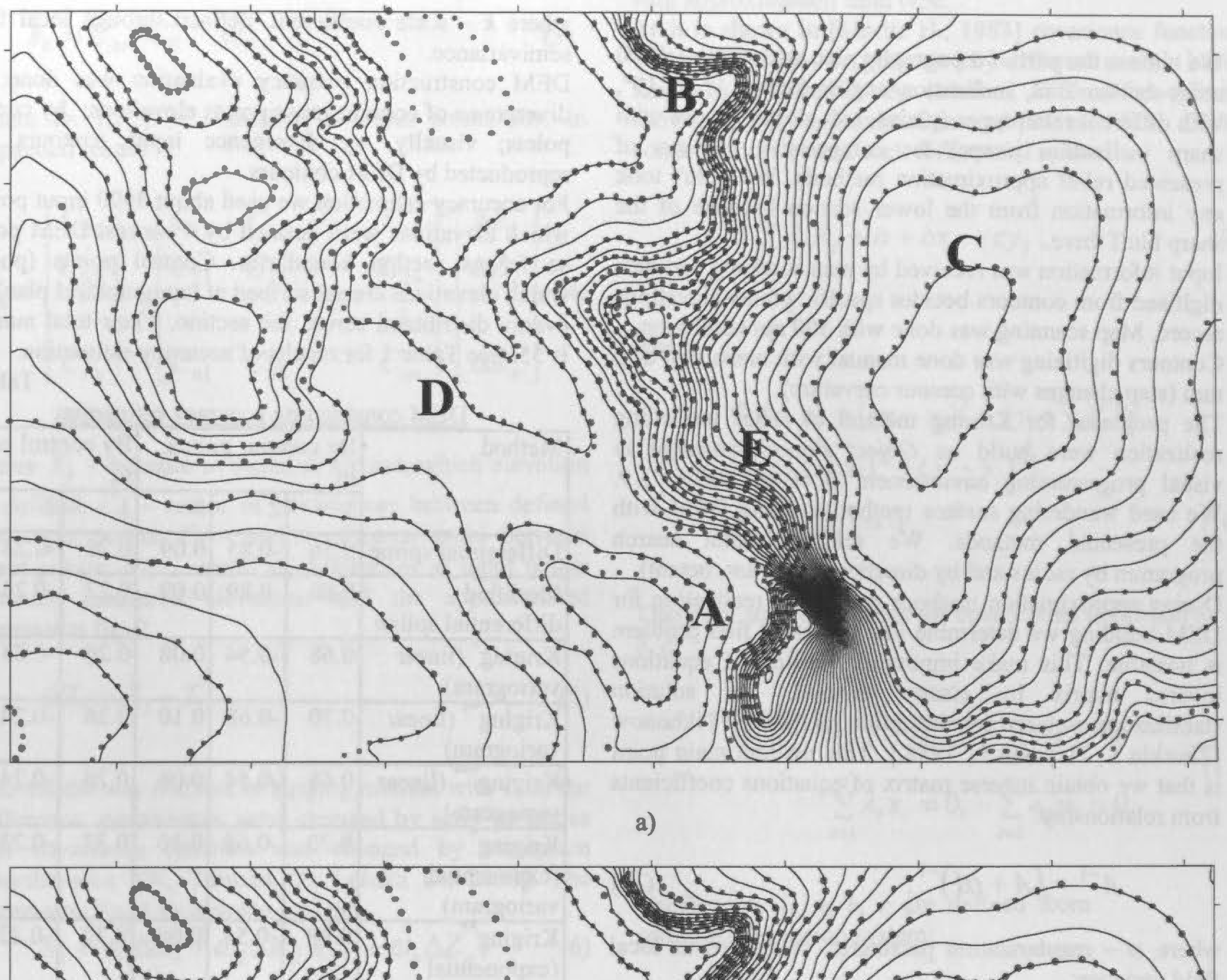
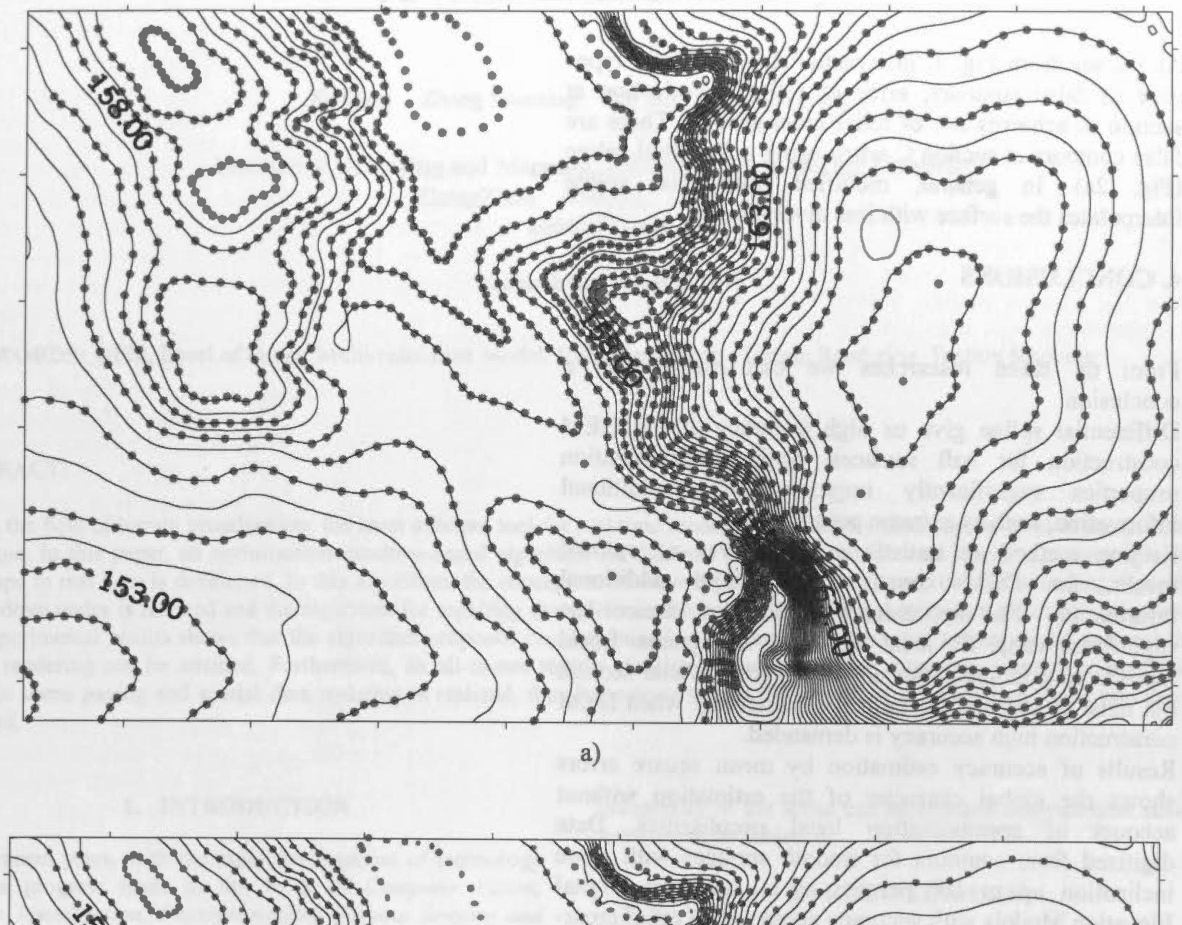

B
C
D
E
A
a)
164.00
154.00
154.00
164.00
159.00
b)

Figure 2. The map of digitaled contours and contours created by DEM constructed with Kriging method.
a) from exponential variogram; b) from linear variogram


158.00
163.00
158.00
153.00
163.00
158.00
163.00
158.00
153.00
163.00

a)

b)

Figure 3. The map of digitaled contours and contours created by DEM constructed with differential splines. a) from Eq. 9; b) from Eq. 13

As we see from Fig. 3, differential splines of two types give us false contours; error of relief reproduction at section A achieves 3/4 of relief cross-section. There are false contours at section C when using differential spline (Fig. 2a). In general, modified differential spline interpolates the surface with less divergences.

4. CONCLUSIONS

From up taken researches we can make such a conclusion:

Differential spline give us high accuracy during DEM construction for soft surfaces. Spline approximation properties sugnificantly improves with additional information, such as extreme points.

Kriging method as statistic method gives us optimal results for DEM construction without additional information. The approximation function choice has sugnificant value for local empirical variograms. Local reproducting function can not be set for a whole section but must be defined for each local fragment when DEM construction high accuracy is demanded.

Results of accuracy estimation by mean square errors shows the global character of the estimation without account of approximation local peculiarities. Data digitized from contours for smooth surfaces with slope inclination up to 10° permits us to construct Digital Elevation Models with accuracy about 1/8 of relief cross-section and accuracy in the range 2/3 – 3/4 of relief cross-section for sections with sharp jump of slopes inclination.

References

1. Moritz H., 1983. *Advanced Physical Geodesy*. Nedra., Moscow, 1983. P.-392.
2. Burshtynska Kh., Tumska O., Lelukh D., 2000. Computer Technology of Determination of Area of Digital Elevation Model. In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Vol. XXXIII, Part B4, pp. 149-154.
3. Davis J.C., 1977. *Statistics and Data Analysis in Geology*. Mir., Moskow, P.-571.
4. Zhurkin I., Neyman J., 1988. *Metody wychyslenij w geodezii*. Nedra, Moskow, P.-303.

A DYNAMIC MULTI-RESOLUTION MODEL AND IT'S APPLICATION TO TERRAIN RENDERING

Xu Qing Zhang Baoming Tan Bing Ma Dongyang

Institute of Surveying and Mapping, Information Engineering University
ZhengZhou 450052, P.R.China
xq64@371.net

Commission IV, WG IV/6

KEY WORDS: DEM, Level of Detail, Multi-resolution Model, View Dependent, Terrain Rendering, Texture Mapping

ABSTRACT:

In the field of terrain visualization, the most efficient tool for real time rendering of complex landscape is level of detail (LOD) technique. In this paper, an optimization quadtree-based algorithm for dynamic generation continuous levels of a given complex landscape in real time is developed. In this algorithm, the conception of view-dependent is introduced, the fast searching algorithm for quadtree nodes is realized and the algorithm for repairing the cracks intrinsically caused by quadtree data structure is improved. The experimental results shows that the algorithm proposed can dynamically generate multi-resolution terrain model and real time terrain rendering can be attained. Furthermore, an all-in-one terrain visualization system which integrates adaptive triangulation, dynamic scene paging and spatial data updating is realized, thus large-scale terrain visualization at interactive frame rate can be achieved.

1. INTRODUCTION

In recent years, with the rapid development of technology and the progress made in the fields of *Computer Vision, Science Visualization, Photogrammetry, Remote Sensing and Computer Graphics*, it's common to generate high detailed terrain model with the data gained by photogrammetry. These complex terrain models consist of millions of triangles or more satisfied people's demand for high sense of realism, however, the rendering performance of these data is seldom considered, which challenges the ability of rendering efficiency and computation capability of computers greatly. The most common approach is to exploit the traditional 3D graphics pipeline, which is optimized to transform and texture-map triangles and can provide high performance of geometry acceleration. But it's still far to meet the practical demand even with the most advanced graphics pipeline available. With Digital Earth project presented and the demand for large-scale terrain simulation, the visualization of large-scale Digital Terrain Model (DTM) proves to be the bottleneck.

General speaking, a complete system to display views of large datasets at high frame rates consists of components to manage disk paging of geometry and texture, LOD selection for texture blocks, LOD for triangle geometry, culling to the view frustum and triangle stripping. Thus, terrain simplification, multi-resolution representation and real-time dynamic rendering have become the focuses of large-scale terrain visualization. As to these topics, three kinds of effective methods were developed, i.e. the visibility preprocessing[1,2], the image-based rendering[3,4] and LOD(level of detail) algorithm[5-17]. Considering the efficiency and performance of terrain rendering, the most common used is LOD model, also called multi-resolution model, which is used to represent terrain models at multiple levels of detail. In this way, the complexity of the scene can be reduced and real-time smooth browse can be achieved.

2. RELATED WORK

Although the model representations achieved are not always the same, they can always fall into two types: the triangle-based multi-resolution terrain model[5-9] and tree-based multi-resolution terrain model [10-17].

Several methods use **Delaunay triangulation** to develop multi-resolution hierarchies [5, 6]. In particular, Cohen-Or and Lev-anoni [6] support on-line view-dependent LOD with temporal coherence, but must resort to "two-stage" geomorphs. Klein *et.al* [7] proposed one algorithm with precise error control mechanism, TINS are reconstructed by Delaunay triangulation when the viewpoint changed, so that any local simplification may leads to global triangulation and restrict the rendering efficiency. And now, the main idea is to save the information of *vertex split* and *vertex collapse* at the pre-triangulation stage, in the process of real-time rendering, only use the information saved to avoid the time-consuming work of triangulation. Such as the *Progressive Meshes* algorithm proposed by Hoppe[8]. This algorithm defines a continuous sequence of meshes for increasing accuracy and store an arbitrary mesh as a much coarser base mesh together with a sequence of n detail records, from which approximations of any desired complexity can be efficiently retrieved. Jie Li[9] introduced an algorithm called Sorted Decremental Triangular Mesh (SDTM) ,which is based on three transformations: Edge Collapsing, Surface Flattening and Line Straightening, through each transform can remove a vertex without holes and thus re-triangulation process can be avoid .

Miller[10] uses a quadtree to preprocess a height field defined on a uniform grid. For each frame at run time, a priority queue drives quadtree refinement top-down from the root, thus allowing specified triangle counts to be achieved directly. Taylor and Barrett [11] extract mesh approximations from rectangular **quadtree** hierarchies. Liu[12] presented the VDDMTM (view-dependent dynamic multi-resolution terrain model), the dynamic terrain simplification and the evaluation function are discussed and the bucket sorting method is adopted to repair cracks, but lacking of continuity constraints. Both Lindstrom et al. [13] and Duchaineau et al. [14] define **bintree** hierarchies, based on binary subdivision of right isosceles triangles, and demonstrate real-time view-dependent LOD. Because these representations are based on regular subdivision, they offer concise storage. Duchaineau et al. are able to create optimal approximating meshes through incremental changes at each frame. However, the meshes are only optimal within a restricted space of meshes, since the regular subdivision structure constrains both vertex locations and face connectivities. As a result, the approximations may be far from optimal when one considers the space of all possible triangulations of the domain. Chen Gang[15] concerns both quadtree-based and bintree-based algorithm .[17] depicted a visualization system integrated some key problems of terrain visulization.

Focusing on the LOD control and culling to the view frustum and dealing with in-memory geometry and texture management, Our algorithm, a quadtree-based view-dependent algorithm applying to continuous multi-resolution terrain rendering is proposed , which can maintain the original topological structure without adding additional vertexes and the properties of original vertexes can all be used directly.

3. MESH REPRESENTATION

3.1 Terrain Quadtree

When using quadtree to represent the terrain model, each node in the quadtree covers a rectangular region of the terrain. The root node which covers the whole terrain is the coarsest representation of the terrain, the lower nodes only cover a quarter of their father nodes. For two neighboring layers in the terrain quadtree, the top nodes have lower resolution than the bottom nods and hence the former induce more sampling errors than the later, thus less rendering quality, however, with fewer nodes to be rendered, the efficiency of the former is much higher than the later, which is key point in interactively browsing a terrain. Thus to gain a good balance between the efficiency and quality of terrain rendering, the key problem of quadtree-based terrain rendering is to decide a node set as a representation for the terrain quadtree considering both the static error (view-independent) and dynamic error (view-dependent).

In general, the original DEM data for terrain visualization can be expressed as a 2D matrix *fpDEM[nMaxRow][nmaxCol]*.where *mMaxRow* and *nMaxCol* are the sampling points in the longitude and in latitude separately. When the DEM data are expressed as a terrain quadtree, each layer of the quadtree consists of the nodes gained by equidistantly resampling with a certain distance. For two neighboring layers, the top layer has half the resolution of the bottom layer. And any none-leaf node in DEM quadtree has four son nodes and each son node cover a quarter region of its father nodes.

3.2 Static Errors Evaluation

As the errors of terrain nodes can be defined accordingly to the terrain fluctuation, we give a static error evaluation function as following:

Suppose that $h(x,y)$ is the elevation function and *A(Star(P))* the average error plane. Node static error can be expressed as formula (1):

$$\varepsilon = \int_{d(c_i)} \left| h(x,y) - A(Star(P)) \right| dxdy \quad (1)$$

Where average error plane is a plane with least-squares fit from all the sampling points .Node static errors are often be computed in the stage of preprocessing and be saved in the *fError* field.

3.3 Dynamic Error Evaluation

In order to establish the direct relation between the viewpoint and the chosen resolution level, thus the farer the distance between the model and the viewpoint, the coarser the representation is used, the nearer, the more complex, we try to represent the importance by distance between the model and the viewpoint. Just as shown in **Fig.1** , suppose the distance between the edge and the viewpoint is ***d***, the length of the project plane is ***L***, α is the field-of-view angle and is constant during the real time rendering, τ (pixels value) is the pixel number of the projected segment of the edge in the screen, and is the length of the line, which is parallel to the screen .

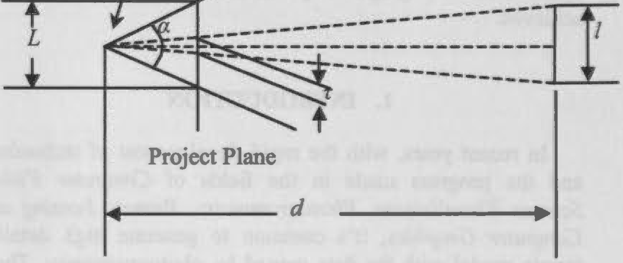


Fig.1 The principle of projective projection

We can easily get the following formula:

$$\tau = \frac{l \times L \times \lambda}{2 \times tg\frac{\alpha}{2} \times d} \quad (2)$$

where **λ** is the number of pixels per world coordinate unit in the screen coordinate system.

According to formula (2) , τ computed is the maximal length of projected segment of this line for a given ***d*** .To a certain edge, the larger the distance ***d*** is the, the less the τ is. At a given error tolerance, when the distance ***d*** is far enough, the length of its projected segment can be less than τ so that the line can be ignored.

General speaking, there are two key factors that closely related to the dynamic error: the static error and the distance between the node and the viewpoint. Considering all these factors together, the following dynamic error evaluation function is given:

$$e(Ci) \approx \frac{k' \bullet L}{\alpha \bullet D} \quad (3)$$

where *L* is length of the terrain that the node covered, *D* is the distance between the center of the node and the viewpoint, α is field-of-view, and k' can be used as a parameter to dynamically adjust system performance so that a comparatively steady rate is sustained.

4. REPARING CRACKS

4.1 The Cause of Cracks

Using terrain quadtree to serve the purpose of multi-resolution representation, there may exist cracks intrinsically. The cause of cracks can be showed in **Fig.2**.Nodes C_1 and C_3 has lower resolution than its neighboring node C_2, which cause the cracks (terrain region uncovered) exsist. In **Fig.2**, There is crack lies between nodes C_1 and C_2, so to C_1 and C_3.

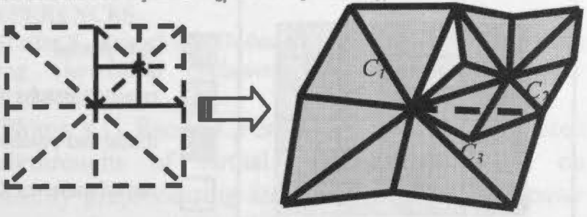


Fig.2 the cause of cracks

4.2 Repairing Cracks

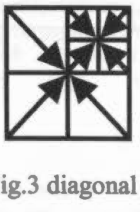
The most common used method to repair cracks is restricted quadtree method **[10]**, which demands that the resolution between one node and its neighboring nodes be restricted to at most one layer. Zhanping Liu[3] proposed one bucket sorting method to repair cracks, but fails to keep the topological structure of cracks regions. Cheng Gang[4] adopt one method that keeps the right topological structure with the dependency of vertexes. There are some still other methods that adding additional triangles to cover the cracks, which may cause discontinuity. We proposed a new method to repair cracks in this section which can be expressed as following:

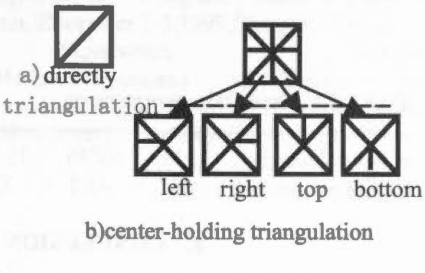
For the sake of convenience, some basic concepts are presented here. And any node mentioned here belongs to the set of the sampling vertexes in the original terrain model.

Definition 1: the son nodes of each none-leaf node in DEM quadtree are named as **Fig.6**, and the upper, the lower, the left, the right are defined as the regions covered by nodes [3, 2]、[0, 1]、[3, 0]、[1, 2] separately.

Definition 2: As to each node select into the node set, its diagonal triangulation refers to line from one corner to the center of its parent node, as shown in **Fig.3**.

Definition 3: The triangulation of quadtree consists of two patterns: directly diagonal triangulation (in **Fig.4 (a)**) and center-holding triangulation (in **Fig.4 (b)**).





a) directly triangulation

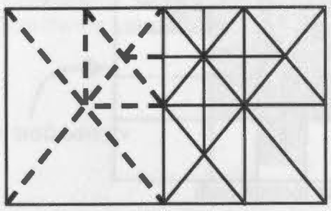
b)center-holding triangulation

Fig.3 diagonal triangulation Fig.4 quadtree triangulation

The following is the method to repair cracks :

1) If the nodes selected have higher resolution than its neighboring nodes or with the same , the directly diagonal triangulation is executed.

2)If the nodes selected have lower resolution than its neighboring nodes ,at first, we executes the diagonal triangulation, and then judge whether the son nodes satisfy 1), this process only involves those son nodes near their higher resolution neighboring nodes. If not, the iterate process will continue until 1) is satisfied.



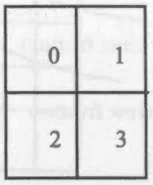


Fig.5 repairing cracks between the left and the right

Fig.6 the index of son nodes

For example, as **Fig.5**, left node (pTreeNode1) has lower resolution than the right neighboring node (pTreeNode2), the dashed is the lines should be added.

The follwing is the iterative process:

```
BEGIN:
SeamErase (pTreeNode1, pTreeNode2) {
pTreeNode1—>bRight EQ TRUE
IF pTreeNode1 lies in the 1 region of parent node
    pTreeNode1—>pParent->bTop=TRUE;
ELSE IF pTreeNode1lies in the 2 region of parent node
    pTreeNode1—>pParent->bBottom=TRUE;
SeamErase ( pTreeNode1->pSon[1], pTreeNode2->pSon[0])
SeamErase ( pTreeNode1->pSon[3], pTreeNode2->pSon[2])
}
END
```

The other relationship between one node and his neighborings can be delt with just as the same.

5. REAL-TIME TERRAIN RENDERING

Dynamic nodes selection is the process to choose an appropriate node set *S* as a representation from the terrain quadtree so that all the nodes in set S should not only cover the whole terrain but satisfy the given error threshold, furthermore, any two nodes in set *S* should not intersect.

In real-time rendering, the key problem is to maintain a relatively steady browsing rate. Thus it's necessary to use the culling to view frustum, which is a dynamic process of choosing only those tree nodes visible to be rendered with the changes of viewpoint and moving directions. This may reduce the triangles to be rendered further.

Fig.7 illustrates the changing regions applied to the active mesh as a user moves forward and to the left through a model. Regions of the model entering the view frustum (on the left an upper) are instantaneously choosed to be rendered, while regions leaving the view frustum (on the right and near the viewer) are instantaneously out of rendering. **Fig.8** shows the culling to view frustum, the regions involved in the two blue lines are visible in the frustum, and the details of regions far from the viewpoint are instantaneous coarsening.

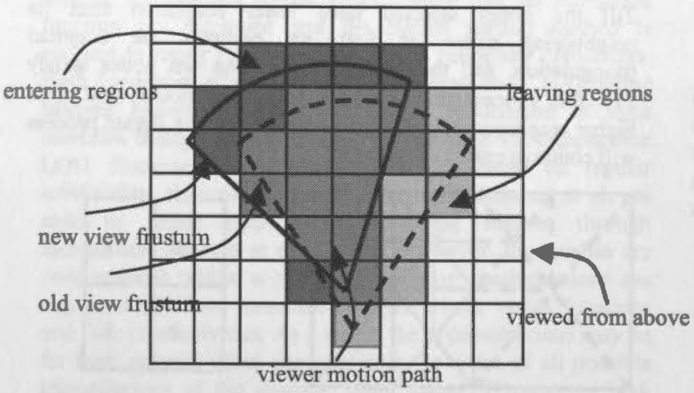


Fig.7 The changes of viewpoint

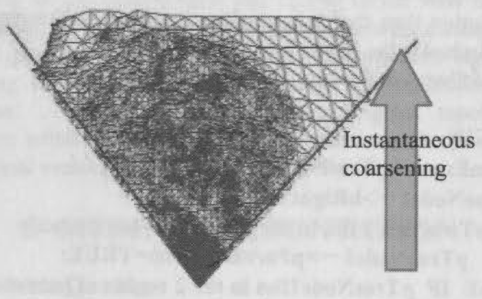


Fig.8 the example of culling to view frustum

In real-time browsing, this is an iterative process and the nodes are adapted dynamically with the changes of error threshold, viewpoint and view direction.

6. SCENE MANAGEMENT AND UPDATING

6.1 Storage and Retrieval

However, the triangulation model described above can't satisfy the demand for large-scale terrain visualization. As to very large terrain database, The input terrain data and texture data are partitioned into blocks of $(2^k+1) \times (2^k+1)$ grid points respectively. This partitioning provides efficient spatial selectivity and physical clustering on external storage. To accelerate the process of data initialization, the errors of each terrain nodes and the normals of each grid points are pre-processed. When stored, all blocks are stored in its corresponding files and data in each file are stored as a quadtree node, furthermore, its child nodes may be set to null for some less fluctuant area. In this way, the terrain data in node can be easily retrieved and the coordinates, normal vector and texture coordinates of grid points can all be used directly, which deserve the demand for some more space for storage.

6.2 Dynamic Scene Updating

To solve the problem of visualizing very large terrain database, a window paging concept as shown in Figure 11 is introduced because we can't load the whole terrain data in main memory and only some blocks of the whole terrain data is necessary to a certain frame. As shown is Figure 9, our visible scenes always represents a window onto the real world in accordance with the current viewpoint and view parameters. Assume that the whole blocks we considered is $n \times n$ and the page of each frame $m \times m(m<<n)$, and the viewpoint is always at the center of the window page. As depicted in Figure 11, with the viewpoint moving, only a few patches need to be updated. Furthermore, the window page is not updated for every small variation of view position and view frustum parameters, an update only occurs when the variation exceeds a specific threshold. This help to reduce the data management costs significantly severe loss of display quality. The process of updating can be achieved with two work threads, one thread for scene updating and the other for displaying, especially for multi-processor computer.

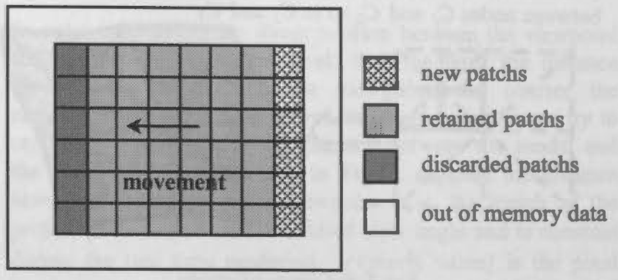


Fig.9 dynamic scene update

7. EXPERIMENTS

With the above algorithm proposed, the browsing of large-scale terrain with high rate is achieved. **Fig 10** illustrates the effect of texture terrain visualization, (a) is the textured full-resolution terrain model, (b) is the simplified multi-resolution model and (c) the simplified model with texture mapping. Compared with (a), even with error threshold of 5 meters, the sense of realism and significant visual details are retained. Thus texturing may improve the visual impression of the terrain and allow higher error tolerance. **Fig.11** illustrates the multi-resolution representations with different error threshold (10,5,2 and 1 meter), with the error threshold decreasing, the terrain are rendered with more and more details. **Fig.12** illustrates the effects of view-dependent LOD control, where the center of the picture is the current view position, the view direction is directly to the north and the viewer motion path is from left to right, the rendering regions with the highest resolution are always at the view position.

We also performed some rendering and scene updating tests based the platform of Windows 2000, the hardware is PII400 with Oxygen GVX1 card. Our whole test scene was a mountainous terrain consising of 48 × 87 patchs, each 6.5km×6.5km large and the always in-memory blocks is 9*9. **Fig 13** shows the scene map. **Table 1** shows the performance of flying through the terrain. **Fig 14** is one frame at interactive frame rate.

TABLE 1 Frame rates comparision

nuber of triangles	5274	10275	15783	20487	43784
frames per second	86	45.1	35.7	22.1	14.1

8. CONCLUSION

The real-time rendering of large-scale surface is one of the most challenging problem in terrain visualization. We have proposed an optimization quadtree-based algorithm for dynamic generation continuous levels of a given complex landscape in real time. The experimental results shows that if the rendered triangulation can be constricted to 1000~2000 and the using of texture mapping, the algorithm can provided a high frame rates maintaining high image quality. Even under the ordinary circumstance, real-time dynamic browsing can be attained.

Our future work will focus on the management of geometry and texture data and the dynamic data-loading from disk interactively. Adapting our algorithm to very large terrain simulation, such as global terrain data is still a challenging problem.

REFERENCES

[1]Teller,S.,Manocha,D.,Hudson.T. et al., Visibility culling using hierarchical occlusion maps *Computer Graphics* 1997,31(3):77

[2]Wang,Y.G.,Bao,H.J.,Peng,Q.S., Accelerated walkthroughs of virtual environments based on visibility preprocessing and simplification. *Computer Graphics Forum*,1998,17(3):187

[3] Sun Hongmei, Ren Jicheng. et.al. Implementing interaction in image-based rendering. *The sixth International Conference on Computer Aided Design and Computer Graphics*, December 1-3,1999,Shanghai, China.

[4]Chen,S..E,QuickTime VR: an imaged-based approach to virtual environment navigation.*Computer Graphics*,1995,29(4):29

[5]Cignoni, P..Puppo,E.. et.al. Representation and visualization of terrain surfaces at variable resolution. *The Visual Computer* 13 (1997),pp.199-217.

[6]Cohen-Or,D.. and Levanoni,Y. Temporal continuity of levels of detail in Delaunay triangulated terrain. In *Visualization'96 Proceedings*(1996),IEEE,pp.37-42.

[7] Reinhard Klein, Gunther Liebich, and W.Straser. Mesh reduction with error control. In *Proceedings of Visualization'96*(1996),pp.311-318.

[8] Hoppe H. Progressive mesh. Computer Graphics .*SIGGRAPH'96 Processing.*

[9] Li Jie, Tang Zesheng. Real-time,continuous level of detail rendering for 3D comples models. In *Proceedings of CAD & Graphics'97*. Fifth international conference on CAD &CG December 2-5,1997,Shenzhen,China.

[10]Mark C. Miller. Multiscale compression of digital terrain data to meet real time rendering rate constraints. PhD thesis, University of California, Davis,1995.

[11] Taylor,D.C and Barrett,W.A. An algorithm for continuous resolution polygonalizations of a discrete surface. In *Proceedings of Graphics Interface'94*(1994),pp.33-42.

[12] Liu Zhanping, Wang Hongwu, et.al. VDDMTM: a view-dependent dynamic multi-resolution terrain model. *The sixth international conference on Computer Aided Design and Computer Graphics*, December 1-3,1999,Shanghai,China.

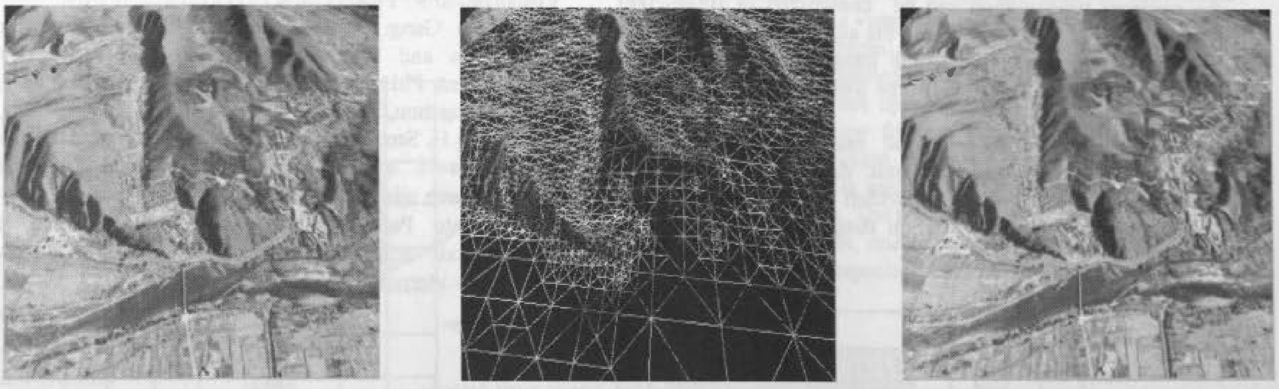
[13] P.Lindstrom, D.Koller, W.Ribarsky,L.F.Hodges,N.Faust,and G.A.Turner. Real-time,continuous level of detail rendering of height fields.In *Proceeding SIGGRAPH 96*,P:108-118.ACM SIGGRAPH.

[14] Mark Duchaineau, Murray Wolinsky *et al*, ROAM Terrain:Real-Time Optimally Adapting Meshes. In Visualization'97 Proceedings(1997),IEEE,pp.81-88.

[15] Chen Gang. A Research on Multi-resolution Surface Description and Real-time Rendering for Virtual Terrain Environment. PhD thesis, Institute of Surveying and Mapping, 2000, Zhengzhou,China.

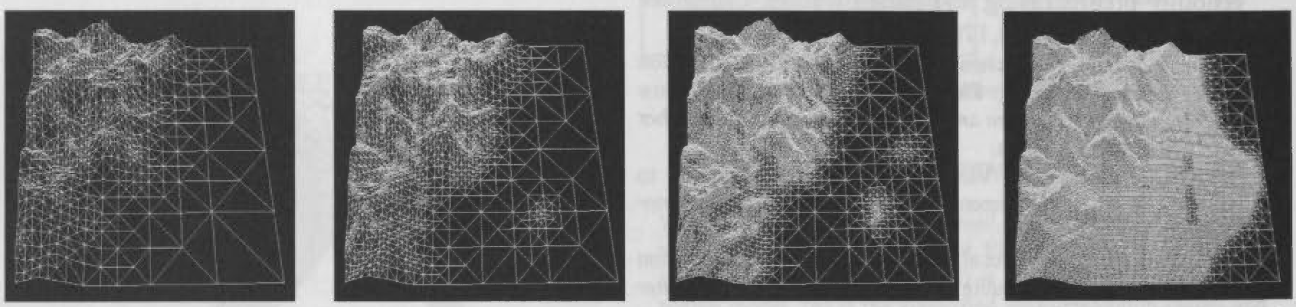
[16]Hoppe H. Smooth view_dependent level_of_detail control and its application to terrain rendering. http://research.microsoft.com/~hoppe.

[17] Renato Pajarola, ETH Zurich. Large scale terrain visualization using the restricted quadtree triangulation. http://www.vterrain.org (accessed 18 Mar. 2001)



a) full-resolution model (textured) b) multi -resolution model (grid) c) multi-resolution model (textured)

Fig 10 textured terrain visualization



a) 10 meter max. error 13% b) 5 meter max. error 27% c) 2 meter max. error 47% d) 1 meter max. error 86%

Fig 11 triangulated surface with different error threshold

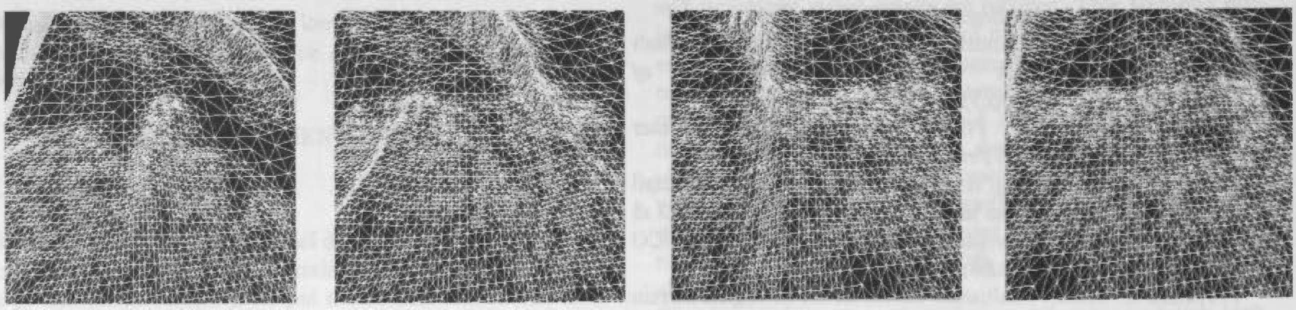


Fig 12 moving of view position (from left to right)

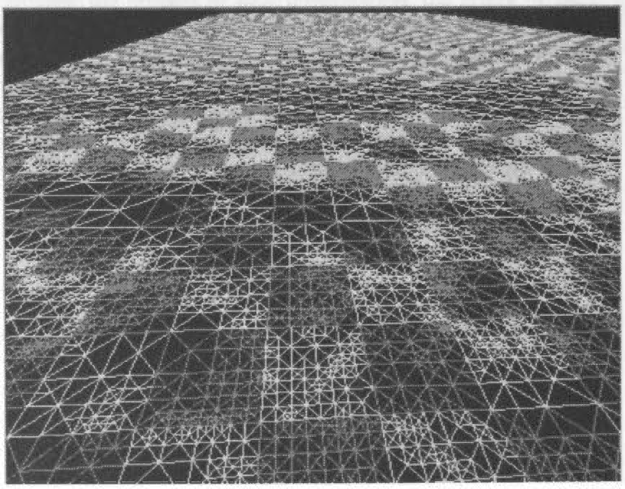


Fig 13 scene map

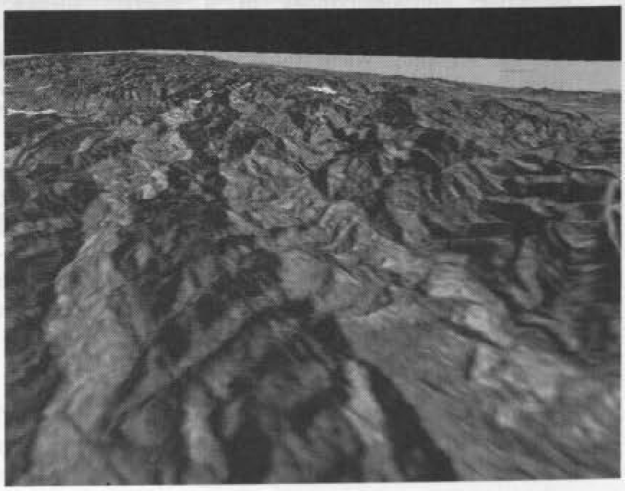


Fig 14 one frame of flying through (num:15783 FPS:35.7)

BUILDING MODELING AND VISUALIZATION FOR URBAN ENVIRONMENT

Efstratios Stamboulogloui Jie Shan*

Geomatics Engineering, School of Civil Engineering, Purdue University, West Lafayette, IN 47907-1284, USA
jshan@ecn.purdue.edu

Commission IV, Working Group IV/6

KEY WORDS: Visualization, Virtual Reality, Photorealistic, Building Modeling, 3-D City Model

ABSTRACT:

One of the primary objectives for geospatial data collection and processing is to generate realistic visualization products of sufficient resolution. Because of the ever growing global urbanization and the complexity of geographic features, realistic urban visualization becomes a major challenge for geospatial research, development and practice. This paper discusses various building modeling approaches and presents its realistic visualization results. Three modelling approaches, flat roof, triangulated irregular network and constructive solid geometry, are described. To create a photorealistic views, textures of for building roofs and walls are respectively acquired from aerial and ground photographs. Addressed in this paper are modelling generation, texture mapping and photorealistic visualization. Various examples are presented for the discussion and demonstration. Photorealistic views over the Purdue campus are shown. Our results and analyses justify the need for a GIS package to further integrate advanced 3-D modeling and animation functions to more effectively model and visualize complex urban environment.

1. INTRODUCTION

One of the primary objectives for geospatial data collection and processing is to generate realistic visualization products of sufficient resolution. Because of the ever growing global urbanization and the complexity of geographic features, realistic urban visualization becomes a major challenge for geospatial research, development and practice. To achieve this objective, urban geospatial features need to be effectively modeled at the required levels of resolution and detail (Brenner and Haala, 1999). In the past studies, most geospatial features, including terrain, roads, water bodies, vegetation and residence blocks, have been successfully modeled via various two-dimensional (2-D) data structures and tools. However, 3-D geospatial features, such as buildings, may not be well represented by and suitable for existing data models (Braun, et, al, 1995). This demands for new approaches to modeling and thereafter visualizing complex 3-D geographic features (Förstner, 1999). Moreover, because of the popularity of GIS and its capability of the integrating and handling different types of geographic data, it is desired that an advanced GIS package be able to model and visualize complex urban environment (Zlatanova and Gruber, 1998).

This paper will discuss various building modeling approaches and presents its realistic visualization results. We conceptualise the building modeling as geometric modeling, texture modeling and their association. This concept is related to the requirement on the levels of resolution and detail of visualization. Different modeling approaches are then studied in terms of their efficiency, fidelity and their integration with geographic information system (GIS). As the simplest modeling approach, buildings are characterized by their footprint polygons with each associated with a height as the polygon attribute. The three dimensional shape of a building is obtained by extruding the footprint polygon with its height attribute. This modeling approach results in buildings with flat roofs or stepwise flat roofs and can be handled by standard GIS functions. The second modeling approach uses triangulated irregular network (TIN). This approach can accommodate complex building shapes and incorporate common terrain data for visualization purpose. As for the third modeling approach, the constructive solid geometry (CSG) principle is utilized. With this approach, a complex building is assembled by or built with simple geometric primitives, such as plane, sphere, cube and cylinder. Boolean operations applied to these geometric primitives are used to create the complex shape of a building. Issues to integrate this complex model into a GIS are discussed. In order to achieve a photo-realistic building modeling and appearance, image textures collected from aerial and/or ground photographs need to be associated to the roofs and walls of buildings. Also discussed is the approach to mapping texture images to geometric models, which will affect the efficiency of view navigation and its fidelity.

The rest of the paper is organized as follows. Section 2 describes and compares three building modelling approaches with simple and complicated buildings. It also shows the integration of building data with other geospatial data, such as terrain and roads. Texture collection and mapping are discussed in Section 4, where aerial and ground photographs are used for building roofs and walls respectively. Section 5 addresses the issues related to the generation of a photorealistic view, including appearance definition, texture mapping, use of light and sky design. The paper is summarized in Section 6. Sample results of photorealistic visualization over the Purdue campus are presented.

2. BULDING MODELS

2.1 Flat roof model

The flat roof model of a building is an extrusion in the vertical direction of a 2-D building footprint. In this modelling, the building footprint can take any complex 2-D shape, however, building roofs are flat or piecewise flat. All the side-walls are vertical. The extrusion can be based on the number of floors or the real height of the building. A 2-D building footprint usually consists of a number of polygons. The topology of the building only involves the topology of the 2-D polygon. The common border of polygons and intersection of edges need to be treated as one single entity respectively so that the extrusion in the vertical direction will not cause topologic conflict or 3-D visual artefacts on the roof. Figure 1 shows a 3-D view of the extruded buildings along with other geographic features over a part of Purdue campus.

---

* to whom contacts may be addressed.

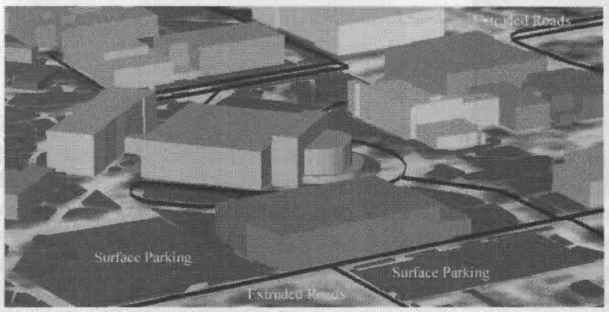


Figure 1. Extruded buildings over Purdue

The other distinction of this modelling approach is its simplicity and supportability by most GIS packages. In addition, the thematic attributes, such as building name, usage and number of floors or building height can be directly associated to the building object.

2.2 TIN Model

In the previous modelling approach, different geographic features are often organized separately as different layers. In the 3-D view, these layers are stacked together according to their natural sequence, such as ground (DEM, digital elevation model) at the bottom, roads and vegetation in the middle, and buildings on the top. Although this multiple layer structure is convenient for spatial and thematic query, it may bring non-realistic artefacts in the 3-D view. At large scale, high resolution view, it can be noticed that features are 'floating' slightly over the ground, so are the vegetation and roads. To overcome this problem, all layers need to be combined into one single file. The flexibility of TIN (triangulated irregular network) structure allows us to accomplish this task. In addition, this will let image textures be draped over the combined virtual model. Since functional surfaces don't support vertical walls, i.e., cannot have two different Z values at the same (X,Y) location, the TIN approach in current GIS packages, e.g., ArcGIS, creates the walls as near-vertical through the use of breaklines and buffering. The building footprint polygons are added to the terrain TIN as breaklines. They then get buffered inward by a specified small distance to create polygons that represent the rooftops. Figure 2 shows a part of Purdue campus draped with aerial images over a combined terrain and buildings TIN model.

2.3 CSG model

To model complex buildings, more advanced tools are needed. Constructive solid geometry (CSG) is introduced for this purpose and used widely in CAD (computer aided design) and computer graphics. This model is featured with fixed topology and geometry. In this approach, a complex building model is built from a composed set of parameterised primitives (Ware, 1994). A 3D object represented using a hierarchy of vertices, edges and faces. The object's surface is a collection of faces that intersect each other only at the common edge. Each face is made of edges and vertices that bound the surface patch it self. The edge of the face does not have to be a straight line, nor does the face have to be a planar surface. Instead of lines and places it is possible to use parametric curves and surfaces to describe the shapes of the elements. The positions and shapes of these geometric elements are usually referred to as the geometry of the boundary model. On the other hand the connections and relationships between the elements are referred to as the topology of the boundary model. (Lammi, 1996).

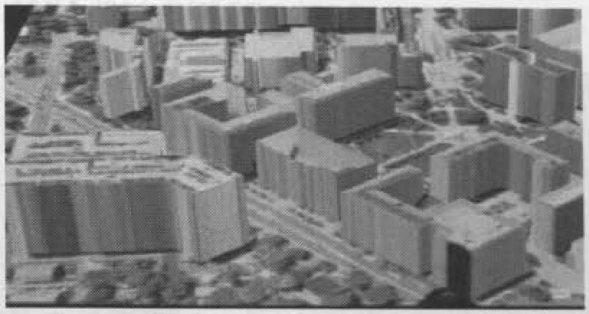
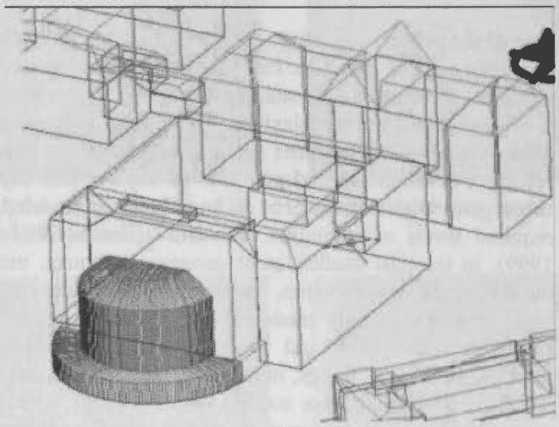
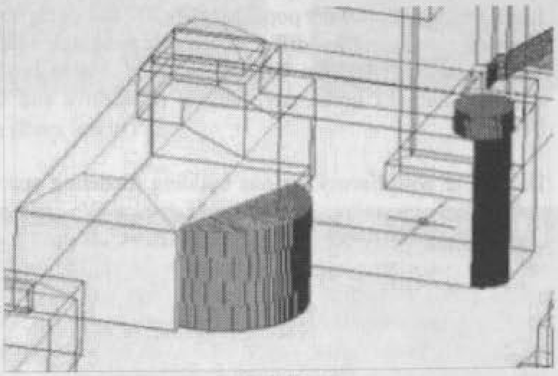


Figure 2. Combined TIN of buildings and DEM



a. Building 1



b. Building 2

Figure 3. Complicated building models from CAD

Different data structures that can be used to represent boundary models include polygon based, vertex based, edge based, and face based models. An example of a polygon-based model would be a polyhedral model. This contains only a set of planar face elements but no topological information about them. It only describes the geometric shape of the surfaces. A polyhedral model can be categorized within the generic models. On the other hand, the vertex, edge, and face-based models, inherently enable the representation of topological information. For a boundary model to be valid, the objects that comprise it must fulfill the following criteria:

- The set of faces is closed
- Faces intersect each other only at common edges and vertices.
- Edges of the faces do not intersect themselves.

This principle has been addressed in variety of studies, e.g., (Gruen and Wang, 1998), (Braun et al, 1995), (Ioannidid et al, 2000). From the previous one can rule out self-intersecting objects, objects that are open. Also from the previous one can understand that each edge only belongs to exactly 2 faces. In general more than one technique is used for the representation of models, in what is called a hybrid approach. In our approach when we created the model in the CAD environment we used a 3D face as a primitive, and made sure the topology is valid by creating the model using some rules. Figure 3 shows two complicated building models over the Purdue campus.

3. TEXTURE MAPPING

A photorealistic view requires buildings have realistic textures other than simple shading (Nakos and Tzelepis, 1998). After having finished the model in AutoCAD we decided to work with 3D Studio Max (Petersom, 1998; Matossian, 2001) to create the photorealistic views of the campus. To apply texture to individual objects in 3D studio, like different parts of walls and roofs, when importing from AutoCAD into 3D Studio, it is important to distinguish between objects, since they are handled separately only when they are either of different color, have different attributes (in this case 3D face and mesh are handled as different objects) or are in different layers. That made it necessary to create layers in AutoCAD for the roof objects, and different layers for the fountain and walls. Putting texture on the faces is the next step. In our study, building roof textures are obtained from aerial images while terrestrial images are used to acquire building wall textures.

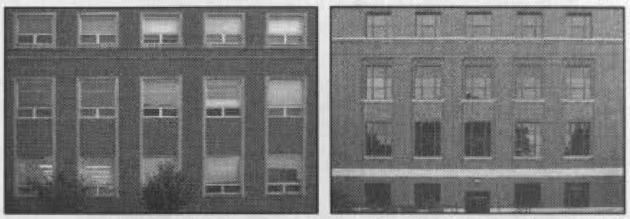
3.1 Roof Texture

To create the roof texture, we used the aerial photograph, which was draped on top of the roofs and base. All the roofs were handled as one object together with the face placed at zero elevation that had the same extents as the image. The result is very realistic. If the walls were the same object as the roof, then the texture of the roof looks like it's continuing on the walls making the result very unrealistic. To keep good resolution, we did not use *.jpg image format as it sometimes requires more rendering time due to decompressing for display.

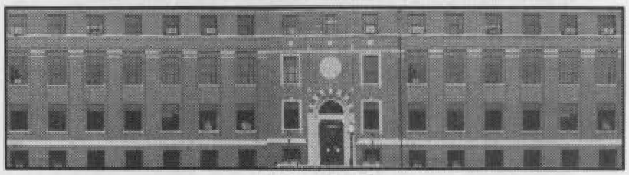
3.2 Wall Textures

To create the wall textures, we took photographs of some building walls on the ground. The photographs were taken using a Kodak digital camera, and a Canon 35mm roll film camera. The pictures from the roll film were developed and scanned using a Microtek scanner. Precision for these images was not as important as with the roof texture so the resolution was kept relatively low. Taking the pictures was not easy due to occlusions, trees (Figure 4a), long facades (Figure 4c) or not enough room between buildings. Capturing the whole façade in one frame was most of the times impossible. All pictures were taken from ground level, with very few being perpendicular to the facades due to lack of space. The roll camera had a very wide-angle 24mm lens making things easier, but this created distortion on the images (Figure 4b). Once the texture images are collected, they need to be associated to each building wall. This is done by using a 2-D perspective transformation.

It should be noted that having all walls photographed is practically impossible and unnecessary. For the building walls that we didn't use picture derived photorealistic textures, we used bitmaps of tiled and concrete bricks. Some buildings used textures from ground photograph of a similar building. We tried to use different ones for each building to create individuality. It is found the bitmap library is useful for this purpose. Figure 5 presents two buildings mapped with image and bitmap textures.

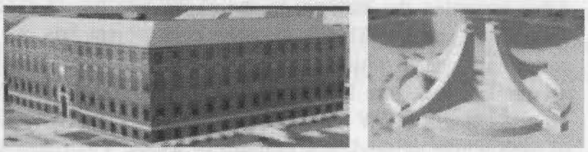


a. Tree occlusion b. Wide-angle distortion



c. Long façade

Figure 4. Textures for walls



a. Building b. Fountain

Figure 5. Building with mapped textures

4. PHOTOREALISTIC VISUALIZATION

The objective in this step is to create a realistic view of the built models. For this purpose, we used one of the most commonly used software in the market, 3DS Max (Petersom, 1998; Matossian, 2001). As addressed earlier, in order to assign textures to 3D objects within 3DS Max, they have to be 3D faces or meshes. Also for different faces to be imported as different objects, they have to be in different layers. Our final AutoCAD model was changed to reflect these needs. Whatever object was going to be used with the aerial image as a texture, was put in a layer called "roof" so that it could be handled as one object. Every object that was not vertical was considered to be part of the roof object. For some buildings we made different textures for each wall which also meant that each wall was in it's own layer, but for most we used one texture that was mapped to all walls which means that all walls were put in one layer. For really small walls we decided to use a brick maps. To create the texture, first one has to create a material using the material editor (Figure 6). For a simple texture map, the bitmap is used as diffuse color map. We also tried to use the same bitmaps with bump maps. The result for the aerial image was not realistic because of the relatively low resolution of the bitmap, and because there are many trees and other structures that create "noise". For the wall maps, due to the big distance from the walls when viewing, the difference is not significant. A bump map creates the illusion of surface by perturbing normals using the intensity values of a map.

When we imported the drawing file in 3DS Max we noticed that in some cases some of the 3D faces didn't show. After some research we found out that 3DS Max assigns a normal to the faces depending on the order the nodes were created. Back faces are not displayed nor rendered. So for some of the walls the normal is facing the inside of the building. To solve this when we created the materials, we made them 2-sided to make sure that texture will be mapped to both sides of the faces.

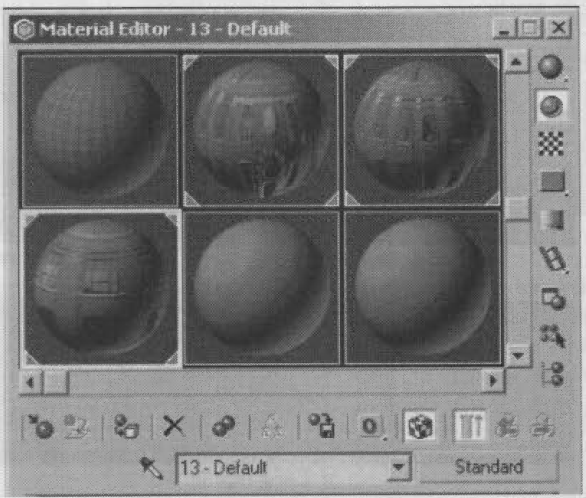


Figure 6. Material editor for the appearance of an object

After assigning the material to the object one has to align the map to the object. This is because each object has it's own coordinate system which is different than that of the bitmap. To accomplish this, one has to apply mapping coordinates to the objects using a UVW modifier. UVW is the coordinate system of each object where U corresponds to width, V corresponds to the height, and W is the axis perpendicular to the UV plane. We used a box mapping fitted to the object. For the faces we had to manually rotate the coordinate system of the map to coincide with that of the object (Figure 7). As we mentioned before, this is essentially a two-dimensional projective transformation. To map the textures rigorously on the walls, one should have had coordinates of points in object space and their corresponding coordinates in image space. We actually do have these coordinates, since we have both the images, and the 3D faces, but 3D studio cannot use this extra information. The mapping is thus done the way described previously.

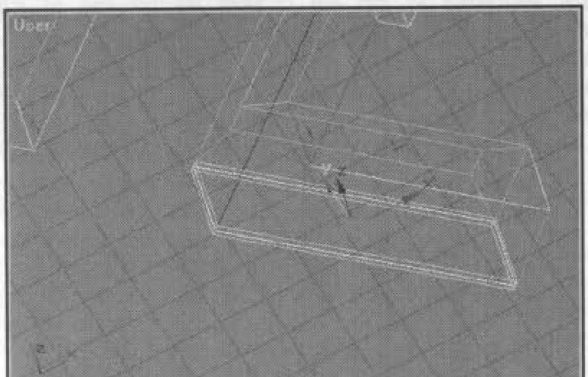


Figure 7. Aligning the map gizmo to the object.

The final touches included creating lights and an environment. We created four omni lights, which radiate light in all directions from a single source point. They can be moved around 3D space without restrictions. We used four in order to light the scene as good as possible. But to make the light not too intensive we set the intensity (which is controlled by the multiplier) at 0.5 to 0.6. We also created a directional light. A directional light uses a cylinder of illumination, which means that the rays are parallel. This is used as sunlight. In this case we also have reduced the intensity to 0.9.

To recreate an environment, we created a sphere that included the whole model. We wanted to recreate the effect of a real sky, which shouldn't move when changing viewing point (at least for small time frames). By using 3D Studio's environment utility, a bitmap can be used as a sky. Figure 8 shows the relationship of the sphere to the rest of the model. By creating the sphere and assigning the bitmap as a map to it, we created a sky that didn't move. The only problem is that no matter what mapping technique you use (planar, spherical, shrink wrap etc) the bitmap is rectangular, which means that there cannot be a good fit. Distortion will occur and the projection center might appear unless it is below the horizon.

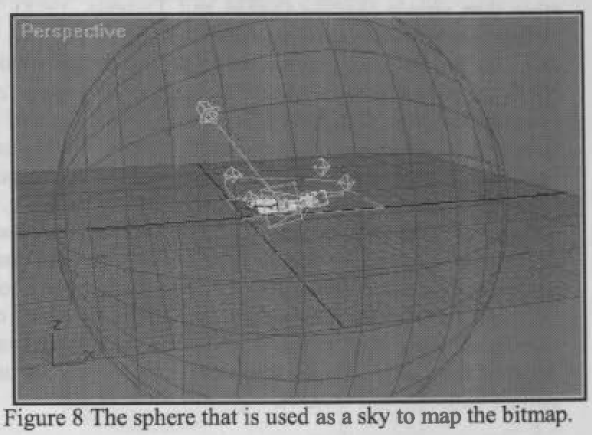
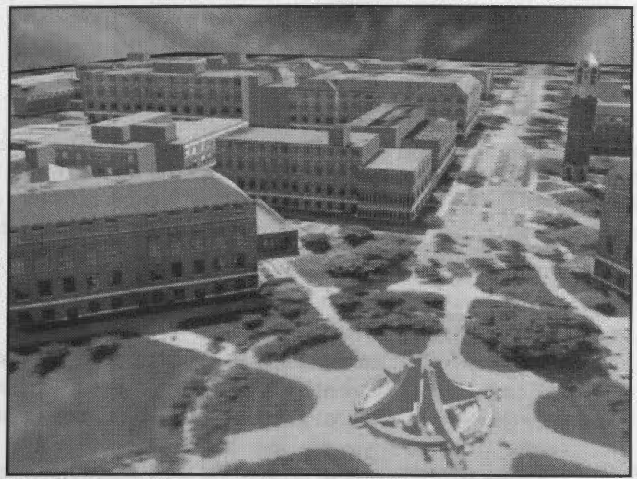


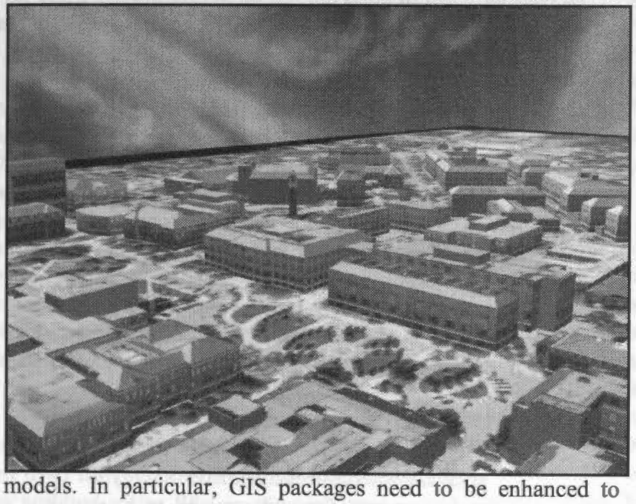
Figure 8 The sphere that is used as a sky to map the bitmap.

All of the above was implemented in creating a virtual model for the Purdue campus area. The final result is very realistic either on a building level especially when appropriate texture was used for the walls, or on a campus-wide level because the aerial image fits very nicely on top of the roofs, making up for lack of detail from the model. Rendering of the scene in a 640x480 frame took approximately 22 sec. We chose not to use the program's ray-tracing ability, which calculates shadows that are created from objects on other objects, for two reasons. One was that the aerial photograph already had shadows, and secondly because this option takes almost three times as long to render. We also decided not to have any shadows created on the buildings at all, since we believe that they would make some of the faces too dark to see. This is one more reason why we set up so many lights, to help in not having shadows anywhere. Different views are presented in Figure 9. A VRML model was also generated as shown in Figure 10.

5. CONCLUDING REMARKS

Urban environment can be modelled in different ways. With GIS, buildings can be created, connected to a table that contains thematic information about them, and then when needed viewed in three dimensions by extruding the polygons. TIN modelling is able to combine building models with other geographic data like DEM and features. However, these current GIS-supported approaches have difficulty to handle complicated building models and map textures to each geometric primitives. As a contrast, CAD tools can model highly complicated buildings and provide sufficient details, however, at a cost of timely modelling process and its inability to handle different forms of geographic features. Photorealistic visualization requires building textures acquired separately from ground and aerial photographs. Mapping textures to the geometric primitives can be a very tedious process. The study shows at current level of technology, photorealistic visualization requires an integration of different tools, including GIS to handle different types of geographic data, CAD to model complex buildings, and visualization package to generate realistic views and render the models. In particular, GIS packages need to be enhanced to accommodate CAD models and render complex views in a photorealistic manner. Our results and analyses justify the need for a GIS package to further integrate advanced 3-D modeling and animation functions to more effectively model and visualize complex urban environment.





a. View 1
b. View 2
Figure 9. Photorealistic view

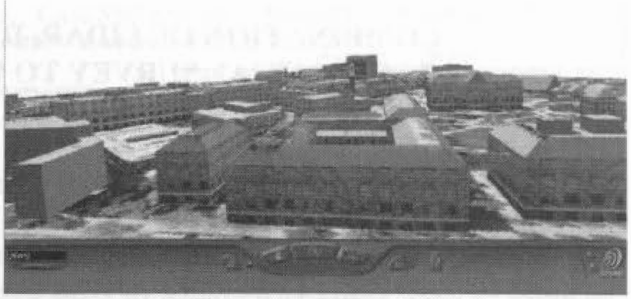


Figure 10. VRML model

REFERENCES

Braun C., Kolbe T.H., Lang F., Shickler W., Steinhage V., Cremers A.B., Förstner W., Plümer L. (1995). "Models for Photogrammetric Building Reconstruction.". Computer & Graphics, Vol. 19, No. 1, pp. 109-118.

Brenner C. and Haala N. (1999), "On the production of 3D city models", Geoinformatics, (July/August issue).

Förstner W., (1999). "3D City Models: Automatic and Semiautomatic Acquisition Methods.". D. Fritsch and R. Spiller (eds), Photogrammetric week 99, Wichmann Verlag, pp. 291-303.

Gruen A. and Wang X. (1998). "CC-MODELER: A topology generator for 3D city models". Proceedings of IAPRS, Volume: XXXII, Part 4. Stuttgart, pp. 188-196.

Ioannidis C., Potsiou C., Soile S., Badekas J. (2000). "Detailed 3D Representation of archaeological sites.". IAPRS, Vol. XXXIII, Amsterdam.

Lammi J. (1996). "3D modeling of buildings from digital aerial imagery.". International Archives of Photogrammetry & Remote Sensing. Vol. XXXI, Part B2, Vienna.

Matossian M. (2001). "3DS MAX for Windows.". Visual quick start guide. Peachpit Press.

Nakos V. and Tzelepis N. (1998). "Three dimensional representation of urban areas using photorealism.". 5th National Cartography Convention, Thessaloniki Greece.

Peterson M. (1998). "3D Studio MAX2 Guide". Giourdas Press.

Sinning-Meister M., Gruen A., Dan H. (1996). "3D City models for CAAD-supported analysis and design of urban areas.". ISPRS Journal of Photogrammetry & Remote Sensing 51, pp. 196-208.

Ware C. (1994). "Advanced 3D Visualization Techniques.". The 6thCanadian GIS Conference, Ottawa, pp. 1694-1698.

Zlatanova S. and Gruber M. (1998). "3D Urban GIS on the WEB: Data Structuring and Visualization.". IAPRS, Vol. 32, Part 4 "GIS between Visions and Applications", Stuttgart.

COMBINATION OF LIDAR, DIGITAL PHOTOGRAMMETRY AND TERRESTRIAL SURVEY TO GENERATE HIGH-QUALITY DEMS

Peter Reiss

Bayer. Landesvermessungsamt, Munich, Germany - Peter.Reiss@blva.bayern.de

Commission IV, WG IV/6

KEY WORDS: LIDAR, laser scanning, DEM, digital elevation models, DTM, digital terrain models, data fusion, digital photogrammetry, photogrammetric feature collection, terrestrial survey, tacheometry

ABSTRACT:

The paper describes the work-flow that has been developed at the State Surveying & Mapping Agency of Bavaria which consists of a combination of LIDAR (in Europe laser scanning") data collection and additional photogrammetric and terrestrial measurements for checking and complementing the mass data derived from LIDAR. This workflow is based upon procedures and recommendations for LIDAR-projects put together by a German user group that consisted of members of several German State Mapping Agencies (SMA's). Finally some experiences and practical results are presented.

1. INTRODUCTION

1.1 Situation in Germany

In 1994 first discussions had been started about LIDAR (or airborne „laser scanning" as it is commonly referred to in Europe) as a new technique for data collection in order to generate high-quality digital terrain models and subsequently contours for digital mapping. Several State Surveying and Mapping Agencies (SMA's) of the Federal States of Germany carried out tests and evaluated the results (e.g. Hoss, 1996; Petzold & Knabenschuh, 1999; Reiche, Schönemeier & Washausen, 1997).

Since these results were very promising a user group was formed in 1996 to come up with standard procedures for tenders and contracts as well as for verification and handling of the delivered data. An internal report was published to provide guidelines for other public organisations interested in the practical application of the LIDAR technique (Petzold, Reiss & Stössel, 1999). The following text contains some general results of this report as well as the work-flow developed at the "Bayerisches Landesvermessungsamt" (BLVA; i.e. *State Surveying & Mapping Agency of Bavaria* aka *Bavarian Land Survey Office*) and some experiences and practical results. Besides the 4 SMA's represented in the above mentioned user group meanwhile other SMA's in Germany have applied LIDAR for their topographical surveys as well.

1.2 Situation in Bavaria

The state of Bavaria is the largest of the 16 Federal States of Germany covering an area of approx. 70,550 km². Towards the end of the 1950's photogrammetry was introduced for topographical mapping of the Alpine regions and later on for the rest of the state as well. After analytical photogrammetric plotters replaced the analogue instruments in 1990 map compilation changed from pure graphical methods to object-coded data collection. Very flat terrain in the open fields as well as housing areas and forests still had to be completed using terrestrial survey (tacheometry) – a very tedious and time consuming task.

As of the early 90's approx. 65% of the state area were covered by contour maps of good but not up to date quality while for the remaining part only maps of inferior accuracy were available going back as far as 1872-1920.

In 1994 a working group at the BLVA produced an internal report containing recommendations how a statewide detailed DTM could be generated with special emphasis on the 35% covered only by very old and insufficient contour map sheets. A combination of photogrammetric image correlation for open areas, stereo-compilation of building areas and laserscannig (LIDAR) for forests were recommended because LIDAR was the only technique to receive terrain data on the ground more or less automatically within wooded areas. A first test using laser scanner data was carried out in 1996 which clearly showed several results:

- The overall accuracy of LIDAR was as good as a very accurate terrestrial tacheometric survey for open fields including very flat ones as well as for building areas and even large parts of wooded areas (except for dense coniferous forests).
- Restricting LIDAR to forest areas was not economical because smaller irregularly shaped project areas would cause the same costs as large rectangular areas. This meant that there was no practical reason to apply image correlation in addition or alternatively to LIDAR.
- The automatic filtering process to eliminate points that did not hit the ground was not 100% reliable, so careful inspection and editing had to be done to remove remaining data that did not define the bare ground. This happened especially in the middle of

larger flat roofs (e.g. of factory halls) and in the middle of wider areas covered by young and densely standing conifers.

- The Digital Photogrammetric Workstation (DPW) that was used in the beginning only for testing purposes was found to be a very efficient tool for checking and editing and therefore remained in constant use ever since.
- Depending upon the choice of parameters and the character of the terrain automatic filtering not only removed points that hit roofs or vegetation but also good measurements along ridges with steep shoulders (Figure 6).
- Although the point density was increased through the years (from an overall average of 4.5 m to 1.5 m) morphological features such as breaklines along dams or cuts, ridge or valley lines were not as precisely defined as desired.

Therefore since 1997/98 a combined method is in operational use consisting of

- laser scanning as the main input

supported by

- digital photogrammetry for checking and editing of the data gained from the scanning laser,
- analytical and/or digital photogrammetry for additional stereo compilation of features and geomorphological structures and
- terrestrial tacheometry to fill data gaps within densely wooded areas (especially young coniferous forests).

(Petzold, Reiss & Stössel, 1999; Reiss, 2001a; Reiss, 2001b).

While in the past the primary product were contours compiled at stereoplotters in order to produce contour maps with DTM data as a by-product now a bare-earth DTM derived from LIDAR / laserscanning – edited, enhanced and completed by photogrammetry and terrestrial tacheometry – is the main result with contours being automatically derived as a follow-up product from the DTM.

2. PLANNING PHASE

The planning / contracting phase for a project using LIDAR for a topographical survey includes

- definition of the project area (preferably of long rectangular shape)
- preparation of maps for the contractor
- supply of coordinates (project boundaries, GPS-reference stations, control points for datum transformation from WGS84 to local, geoid undulations)
- definition of one or more check areas for the contractor (to check the datum transformation from WGS84 to local; not to be used for adjusting the orientation in height; Figure 3)
- definition of several check areas for the SMA (for an initial overall check of orientation / transformation of the delivered mass points and – as far as applicable – individual strips; Figure 3).

3. LASERSCANNER FLIGHT, PHOTO FLIGHT, PREPARATORY FIELD WORK

Laserscanner flights for the BLVA usually have been carried out so far under the following conditions:

- flying height approx. 900 m
- strip width approx. 650 m
- side lap approx. 250 m
- resulting point density approx. 1.5 m (in the beginning about 4.5 m)

Best results are achieved if the LIDAR flights are carried out with "leaves off", i.e. between late autumn and early spring, as long as there is no snow coverage.

Parameters of the additional photo flights:

- image scale 1 : 14,000
- color film (diapositive)
- signalized control points (coordinated using terrestrial DGPS)

Image flights are made in early spring as soon as the snow is gone and before the leaves come out, i.e. from the middle of March until the middle of April.

Field work supporting LIDAR and photo flight resp.:

- signalization and measurement of control points
- set up and handling of at least 2 GPS reference stations within the project area during the LIDAR data collection
- survey of check areas to check the height accuracy (grid of about 50 m x 50 m with 10 m spacing within plane areas with little or no vegetation – usually sports / soccer fields; Figure 3)
- survey of several breaklines to check the horizontal positioning of the LIDAR data delivered

Outdoor work usually starts before the photoflight (middle of March).

4. CHECKS (ACCEPTANCE OF DATA)

Besides a report, plots (flight plan) and GPS-coordinates (of flight lines) several files of laser data have to be delivered by the company under contract (which is also responsible for the filtering process):

- file of registered points classified as ground points
- file of points classified as vegetation / buildings
- file of highest points (to define a raw surface model; optional)
- file of height differences (to define relative height of vegetation; optional)

As a first check point plots covering the whole project area are produced and inspected to validate coverage and point density (Figures 1 and 2).

Another very important check regarding accuracy and reliability consists of the comparison of the laser scanner measurements against the check areas from outdoor work. The laser data are accepted if the height difference compared with the grid surface is within |0.3 m| for at least 95% of the laser points classified as ground points (Figures 3 and 4).

Further initial checks to look for outliers / blunders can be carried out through inspection of contours and / or shaded reliefs derived from a very dense DTM-grid based on the raw laser data as delivered by the contractor (Figures 7 and 8).

5. DATA PROCESSING (EVALUATION, EDITING, ADDITIONAL DATA COLLECTION)

To enhance the results received from laser data geomorphological structures are compiled interactively at stereoplotters. Typical features are boundaries of streams and lakes, natural or artificial breaklines (especially at dams), ridge lines, valley lines, edges of cuts and fills as well as local extrema (saddle points, hills and hollows). See Figure 2, 7 and 10.

Later on the data are checked and edited in detail at several DPW's which give a stereo-scopic 3D-view (of the laser data plus the additional features) superimposed upon the stereomodels from the image flight (a 2D simulation is given in Figure 5). A superimposition of contours can be helpful as well; these are derived from an intermediate DTM out of the edited LIDAR points plus morphological features.

Areas that are not covered completely have to be filled by field survey parties using tacheometry. This mainly concerns forest areas covered by a dense population of young conifers. Outdoor data collection mainly consists of the measurement of ridge and valley lines because these are especially important for modelling the flow of water (dividers and collectors). Since 2001 outdoor data collection is done using registrating theodolites and laptop computers which results in a full digital data flow from field to office. The ArcView® software is installed on these laptops in order to visualize for the surveyor the raw contours as well as the line features in the field.

6. EXAMPLES OF PRACTICAL RESULTS

Figures 2 and 7 through 10 show sample plots out of one project covering rather rough terrain to visualize the effect of additional photogrammetrically and terrestrially collected morphological features.

7. CONCLUDING REMARKS

Since the introduction of laser scanning as the primary input for topographical survey the percentage of areas that have to be completed outdoors by a terrestrial survey went down from approx. 40% to less than 10% (large parts of Bavaria are covered by forests mainly consisting of conifers).

Traditional photogrammetric stereo-compilation has been reduced as well. This led to an increase in throughput of factor 4-5 so far which does not mark the end yet (primarily because the advantage of a complete data flow from outdoor work to office did not show in those figures yet).

8. REFERENCES

Hoss, H., 1996. DTM Derivation with Laser Scanner Data. *Geomatics Information Magazine (GIM)*, 10(10), pp. 28–31.

Petzold, B., Knabenschuh, M., 1999. Data post-processing of Laser Scan data for countywide DTM production. In: *Fritsch/Spiller (Eds.): Photogrammetric Week '99*, Heidelberg, Germany, pp. 233-240

Petzold, B., Reiss, P., Stössel, W., 1999. Laser scanning – surveying and mapping agencies are using a new technique for the derivation of digital terrain models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), pp. 95-104

Reiche, A., Schönemeier, P., Washausen, M., 1997. Der Einsatz des Laserscanner-Verfahrens beim Aufbau des ATKIS-DGM. *Nachrichten aus der Niedersächsischen Vermessungs- und Katasterverwaltung*, 47(2), pp. 68-87 (in German)

Reiss, P., 2001a. Neue Verfahren zur Gewinnung von Geobasisinformationen aus der Sicht der Landesvermessung - Teil 2: Einsatz von Laserscanning und Digitaler Photogrammetrie für die Topographische Geländeaufnahme. *Mitteilungsblatt des DVW-Bayern*, 53(1), pp. 55-73 (in German)

Reiss, P., 2001b. Topographical Survey using Laserscanning at the State Surveying & Mapping Authority of Bavaria. In: *Torlegard, K., Nelson, J. (eds.), 2001. Proceedings of OEEPE-Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Elevation Models, Stockholm, 1-3 March 2001.* OEEPE – Official Publication No. 40, Bundesamt für Kartographie und Geodäsie, Frankfurt a.M., Germany, pp. 135-137 (CD-ROM)

9. ACKNOWLEDGEMENTS

The author thanks Mr. Johannes Scherer-Herz for providing the examples of practical results and Mr. Wolfgang Stoessel for the preparation of all other illustrations.

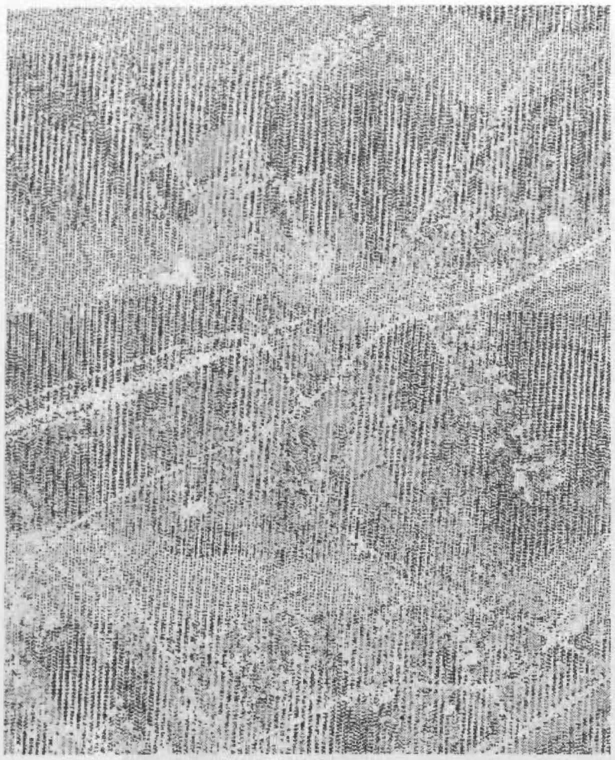


Figure 1. Point plot of LIDAR points classified as ground (black) and non-ground (i.e. vegetation / buildings; green)

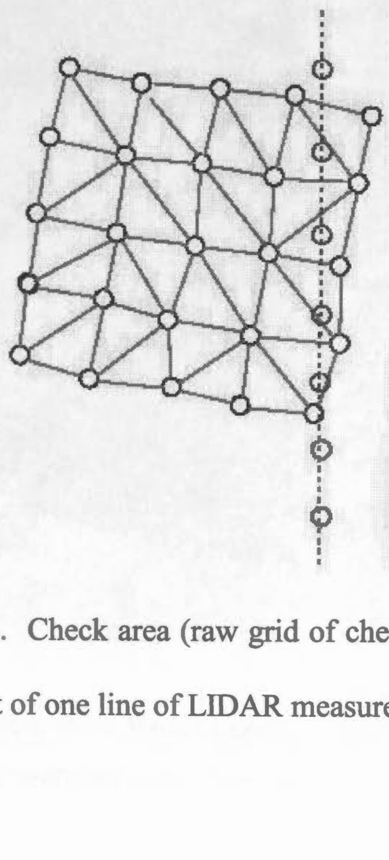


Figure 3. Check area (raw grid of check points and footprint of one line of LIDAR measurements)

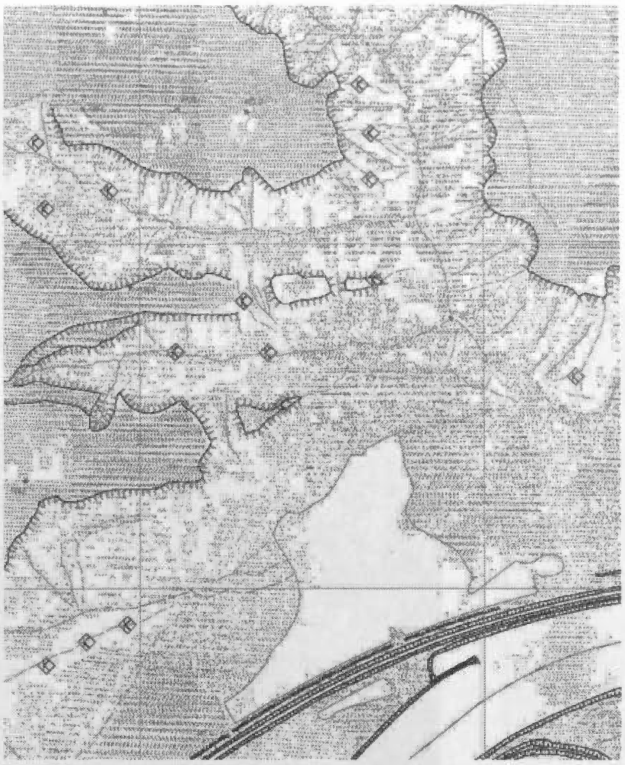


Figure 2. Point plot of LIDAR points overlayed with morphological features and markers for field check

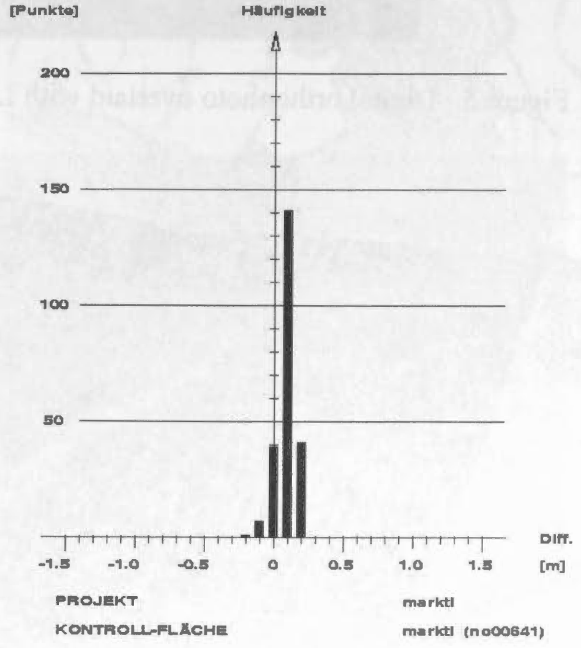


Figure 4. Histogram of height residuals within check area

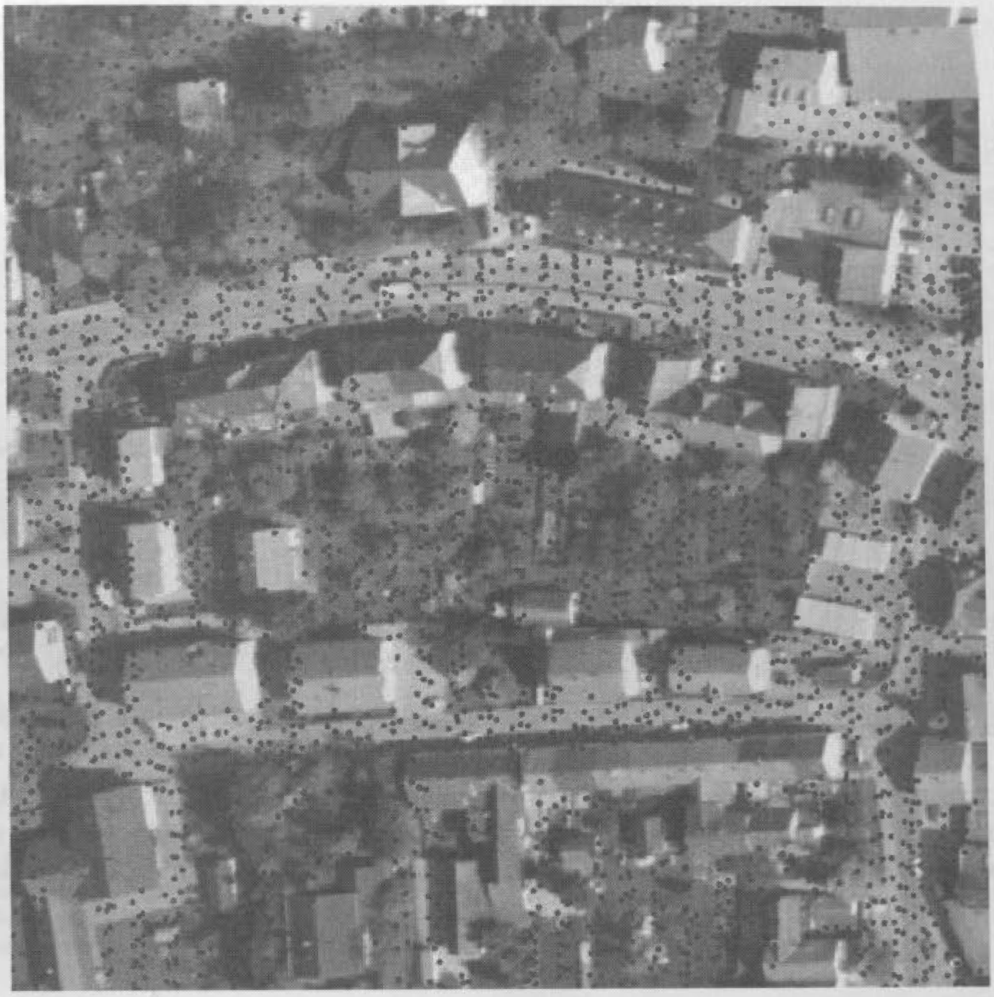


Figure 5. Digital orthophoto overlaid with LIDAR measurements

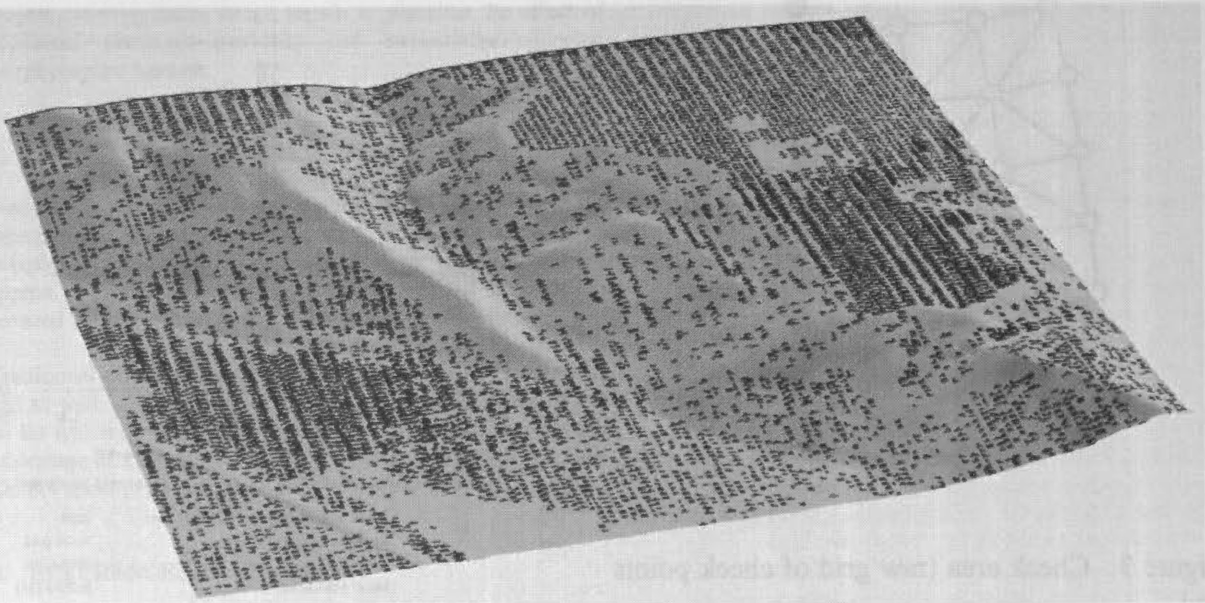


Figure 6. Shaded relief (perspective view) showing filtering of good LIDAR measurements along steep ridges

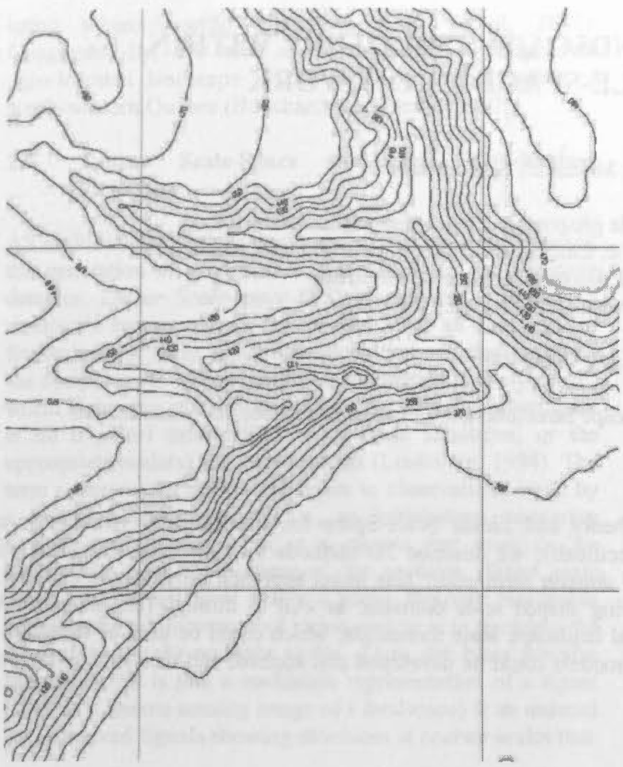


Figure 8. Contours derived from LIDAR data only

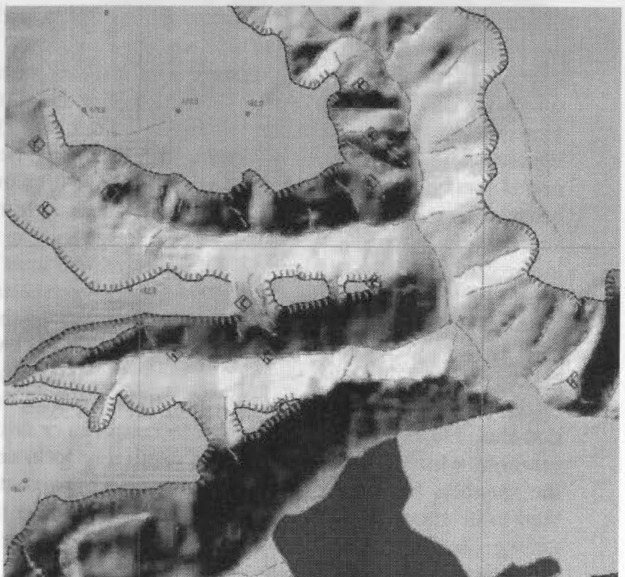


Figure 7. Shaded relief from raw LIDAR data superimposed with photogrammetrically compiled morphological features

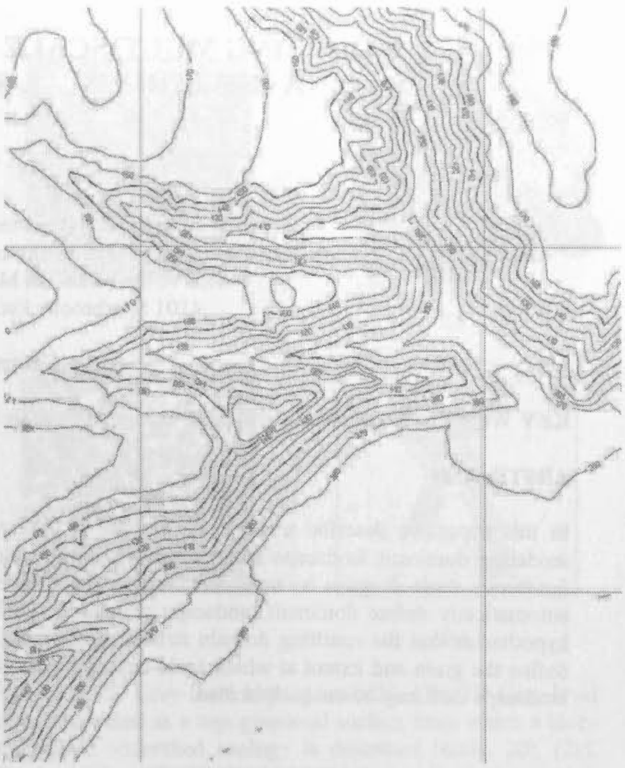


Figure 9. Contours derived from LIDAR data plus morphological features from stereo compilation

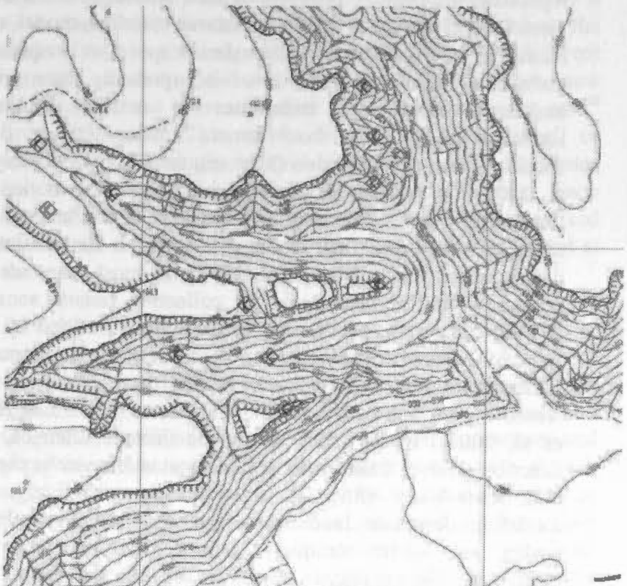


Figure 10. Overlay of contours derived from LIDAR data only and contours including morphological features

MODELING MULTISCALE LANDSCAPE STRUCTURE WITHIN A HIERARCHICAL SCALE-SPACE FRAMEWORK

Geoffrey J. Hay[a, *], D. J. Marceau[a], A. Bouchard[b]

[a] Université de Montréal, Département de géographie, Geocomputing Laboratory,
C.P. 6128, succursale Centre-Ville, Montréal, Qué, Canada, H3C 3J7. *Email: ghay@sympatico.ca
[b] IRBV, Université de Montréal, Jardin Botanique de Montréal,
4101 Sherbrooke Est, Montréal, Qué. Canada. H1X 2B2

Commission IV, WG IV/6

KEY WORDS: Hierarchy theory, Scale-Space, Multiscale Landscape Structure, Blob-Feature Detection, 3D-Visualization

ABSTRACT:

In this paper we describe a novel integration of Hierarchy Theory and Linear Scale-Space for automatically visualizing, and modeling dominant landscape structures at multiple scales. Specifically, we describe 3D methods for modelling and visualizing *landscape scale-domains* by using scale-space events as *critical domain thresholds*. This novel approach provides the capacity to automatically define dominant landscape structures within varying shaped scale domains, as well as through (all) domains. We hypothesize that the resulting domain structures represent critical landscape scale thresholds; which could be used as templates to define the grain and extent at which scale-dependent ecological models could be developed and applied, and the limits over which landscape data may be uniquely scaled.

1. INTRODUCTION

Landscapes are complex systems that are composed of a large number of heterogeneous components that interact in a non-linear way, are hierarchically structured, and scale dependent (Wu and Marceau, 2002). Remote sensing technology represents the primary provider of such landscape-sized data. Therefore, if we are to fully understand, monitor, model, and manage our interaction within landscapes, we require a multiscale framework capable of incorporating appropriate multiscale-theory and techniques to extract dominant landscape components from remote sensing data at their specific scales of expression (Marceau and Hay, 1999; Hay et al, 2002b). However, this is no trivial task. The notion of 'landscape' varies between users, there is no 'science of scale', nor are there fixed scaling laws or rules for translating data between scales (Hay et al, 1997). In most cases we do not know the 'correct' scale for collecting remote sensing data at, but rather, are limited to resolutions defined by the state of sensor technology (i.e., spatial, spectral, temporal, radiometric). Furthermore, once we do acquire data, how and where we scale to and from is often arbitrarily defined (Hay et al, 2001). In an effort to reduce these challenges, we describe a novel framework that integrates *Hierarchy theory* and *Scale-Space* (SS) for automatically visualizing and modeling dominant landscape structures through multiple scales, and within uniquely defined scale domains. In particular, the primary objective of our study is to automatically link structures at unique scales in scale-space, to higher-order objects called *'scale-space blobs'*, and to extract significant features based on their appearance and persistence through all scales. Blob-like structures, which persist in scale-space, are likely candidates to correspond to significant structures in the image, and thus in the landscape. In addition, by considering scale-space events as critical domain thresholds, we are able to three-dimensionally model and visualize multiple 'landscape scale-domains'. This novel approach provides a new capacity to automatically define dominant landscape structures within varying shaped scale domains, as well as through (all) domains. We hypothesize that these domain structures represent critical landscape scale thresholds; thus they may be used as templates to define the grain and extent at which scale-dependent ecological models could be developed and applied, and the limits over which landscape data can be scaled.

In general terms, a *hierarchy* may be defined as 'a partial ordering of entities'; thus hierarchies are composed of interrelated subsystems, each of which are made of smaller subsystems until a lowest level is reached. Within the formal framework of Hierarchy theory (Allen and Starr, 1982), a hierarchically organized entity can be seen as a three-tiered nested system in which levels corresponding to slower behavior are at the top (Level +1), while those reflecting successively faster behavior are seen as a lower level in the hierarchy (Level -1). The level of interest is referred to as the Focal level (Level 0). From a Landscape Ecology perspective, Hierarchy theory predicts that complex ecological systems, such as landscapes, are composed of relatively isolated levels (*scale domains)*, where each level operates at relatively distinct time and space scales. *Scale thresholds* separate these domains, and represent relatively sharp transitions or critical locations where a shift occurs in the relative importance of the variables influencing a process. In general, interactions tend to be stronger and more frequent *within* a domain than *among* domains. Conceptually, these ideas enable the perception and description of complex systems by decomposing them into their fundamental parts and interpreting their interactions.

2. METHODS

2.1 Study Site

Due to the computational demands required by SS processing, analysis was performed on a high-resolution (1.0 m) panchromatic IKONOS image (acquired in September, 2001) that was linearly contrast stretched from 11- to 8-bits. A 2.0-km^2 sub-area was then extracted and upscaled to 4.0 m using Object-Specific Upscaling (Hay et al, 2001). Geographically, this area represents a highly fragmented agro-forested landscape in the Haut St-Laurent region of south-western Québec (Bouchard and Domon. 1997).

2.2 Linear Scale-Space (SS) and Blob-Feature Detection

All multiscale analysis is composed of two main components: the generation of a multiscale representation, and a feature detector. *Linear Scale-space* (SS) is used for generating a multiscale representation. Essentially, SS is an uncommitted framework for early visual operations that was developed by the computer vision community to automatically analyze real-world structures at multiple scales – specifically, when there is no *a priori* information about these structures, or the appropriate scale(s) for their analysis (Lindeberg, 1994). The term *uncommitted framework* refers to observations made by a front-end vision system (i.e., an initial-stage measuring device) such as the retina or a camera that involves 'no knowledge', and 'no preference' for anything. When scale information is unknown within a scene, the only reasonable approach for an uncommitted vision system is to represent the input data at (all) multiple scales. Thus, the basic premise underlying SS is that a multiscale representation of a signal (such as a remote sensing image of a landscape) is an ordered set of derived signals showing structures at coarser scales that constitute simplifications of corresponding structures at finer scales. In practice, Gaussian filters are applied to an initial image at a range of kernel sizes resulting in a scale-space cube or 'stack' of progressively 'smoothed' image layers, where each new image layer represents convolution at an increased scale (Fig. 1). More explicitly, each 'smoothed' layer is created by convolving the nth-order *derivative of a Gaussian* (DOG) function with the original image, where the scale of each derived signal is defined by selecting a different standard deviation for the DOG function (at each new iteration). In the presented work we have only use the zeroth order derivative. This results in a 'scale-space cube', or 'stack' of increasingly 'smoothed' images, which illustrates the evolution of the original image through scale. Each hierarchical layer in a stack represents convolution at a fixed scale, with the smallest scale at the bottom, and the largest at the top.

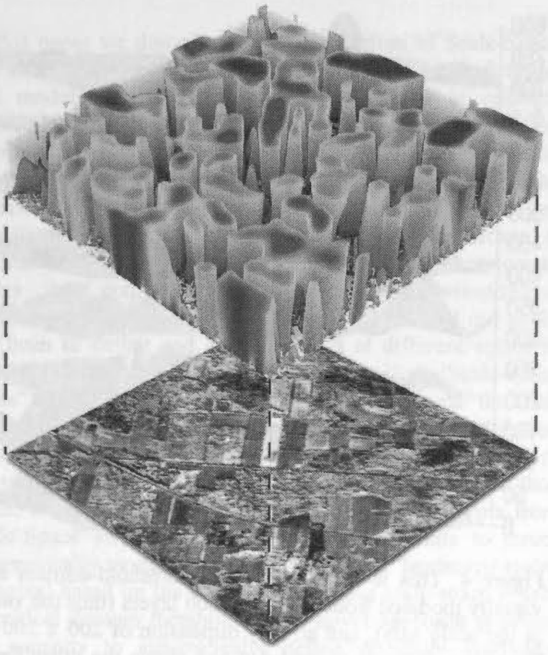
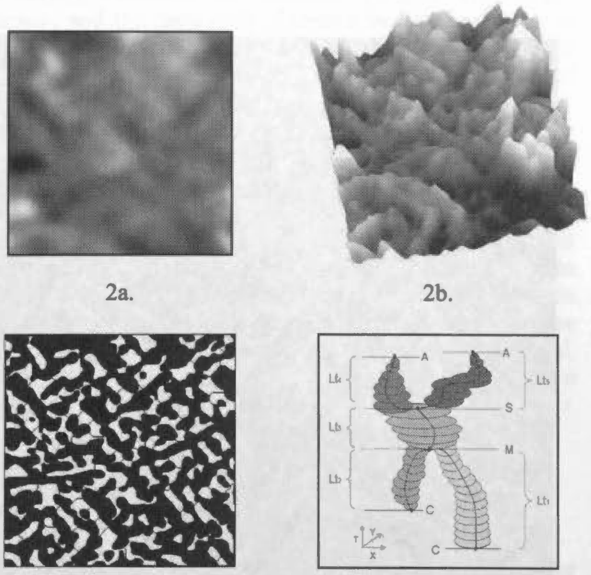


Figure 1. This illustrates a linear scale-space stack composed of 200 scales. The smallest scale is on the bottom; the largest is on the top. Note the diffusive patterns and persistence of scale-space objects through scale. For reference, the 500 x 500 pixel Ikonos image is also provided (bottom).



2a. 2b.

2c. 2d.

Figure 2a. 2D Grey-level blob at scale 20. (2b) 3D Grey-level blobs illustrated as a topographical surface from which a blob-delineation watershed analogy is described (scale, 20). (2c). Binary blob defined from 3a. (3d) Idealized hyper-blob illustrating four different SS-events: annihilations (A), creations (C), merges (M) and splits (S). The number of scales between SS-events represents the lifetime (Lt_n) of a SS-blob.

Blob-Feature Detection represents the second component of multiscale analysis. The primary objective of this non-linear approach is to link structures at different scales in scale-space, to higher-order objects called 'scale-space blobs', and to extract significant features based on their appearance and persistence over scales. The main features that arise at each scale within a stack are smooth regions, which are brighter or darker than the background and which stand out from their surrounding. These regions are referred to as 'grey-level blobs' (Fig. 2a). When blobs are evaluated as a volumetric structure within a stack, it becomes apparent that some structures visually persist through scale, while others disappear (Fig. 1). Therefore, an important premise of SS is that blob-like structures which persist in scale-space are likely candidates to correspond to significant structures in the image, and thus in the landscape. In simple terms, grey-level blobs at each scale in the stack are treated as objects with extent both in 2D space (x, y) and in grey-level (z-axis) – thus 3D. Grey-level blob delineation may best be defined with a watershed analogy.

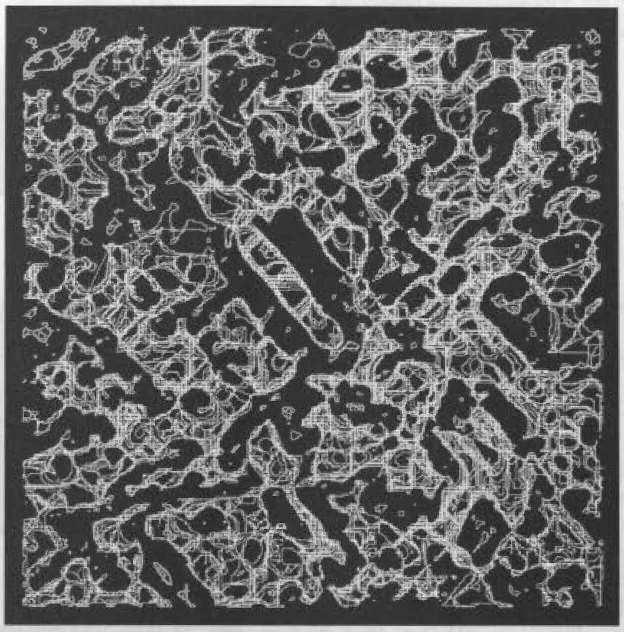


Figure 3. Ranked blobs converted to individual queriable vectors. Note how polygons from many different scales appear to overlay each other making analysis non-trivial.

At each scale in the stack, the image function of all blobs may be considered as a flooded 3D landscape (i.e., a watershed see Fig. 2b). As the water level gradually sinks, peaks appear. At some instance, two different peaks become connected. The corresponding 'connected' elevation levels are called the 'base level' of the blob. They are used for delimiting the 2D spatial extent or 'region of support' of each blob, which is defined as a binary blob (Fig. 2c). 2D binary blobs at all scales are then combined within a new stack to create 3D hyper-blobs. Within a single hyper-blob there are four primary types of visible structures or 'bifurcation events': *annihilations* (A), *merges* (M), *splits* (S), and *creations* (C) (Fig. 2d). The ability to define these SS-events is a critical component of SS, as scales between bifurcations are linked together forming the lifetime (Lt_n) and topological structure of individual SS-blobs. Next, the integrated normalized (4D) volume (x, y, z, t) of each individual SS-blobs is defined.

As blob behavior is strongly dependent upon image structure, it is possible that an *expected* image behavior may exist. Thus statistics are extracted from a large number of stacks resulting from random images[1]. These statistics describe how random noise blobs can be expected to behave in scale-space, and are used to generate a normalized 4D SS volume for each SS-blob.

These resulting normalized volumes are then ranked, and an arbitrary number of significant SS-blobs are defined, from which the scale (t) representing the maximum 3D grey-level blob volume (x, y, z) of each hyper-blob is extracted. From these layers the 2D spatial support (i.e., binary blob) is identified and related back to the corresponding structures in the image for further examination (Fig. 3). Thus, based on the underlying initial premise, 4D scale-space blobs are simplified to 3D grey-level blobs, which are further simplified to their 2D support region (x, y), and then to their corresponding real-world object in the original image. For a more detailed non-mathematical description of SS and Blob-Feature Detection, see Hay et al., 2002a.

2.3 Integrating Hierarch Theory and Scale-Space

A limitation of SS is that within a SS-cube a significant amount of redundant data results in large stack sizes, which in our research range from 200 MB to 980 MB each. In order to reduce the memory requirements when defining SS-blob topology, we have integrated a three tier approach from Hierarchy theory with the programming capability of IDL (interactive data language) to 'parallel-process' multidimensional array structures. Thus, instead of loading the entire stack into memory, we only need to load three scales of a SS-cube into memory at a time. From a Hierarchy theory perspective, we evaluate the blob locations at the 'focal' scale, and establish links with blobs in the scale above and with those in the scale below. We then shift up an additional scale in the cube, while dropping the bottom scale. Always keeping only three scales in memory at once. We then repeat this procedure until the last scale has been processed.

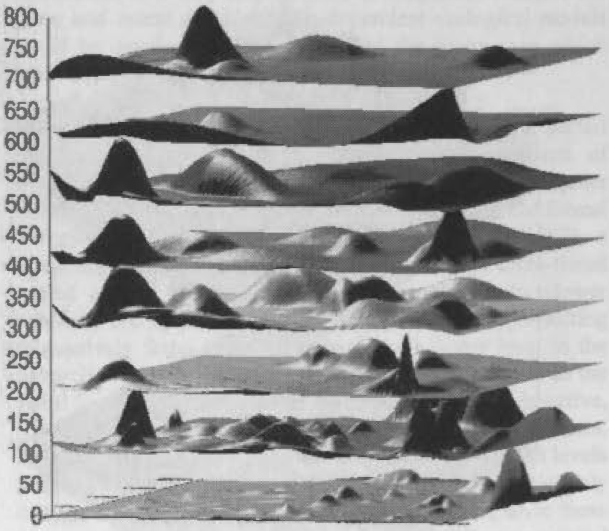


Figure 4. This is an example of 8 threshold-domain surfaces visually modeled from a stack of 100 layers (thus the value 800 in the scale axis), and an x, y dimension of 200 x 200 pixels. Each domain layer is modeled one above the other for visual interpretation. Each domain surface stacks exactly upon the surface underneath it, with no peak protruding into the upper or lower surface. Peak locations represent the bifurcation point of each scale-space blob defined within a single hyper-blob.

In order to overcome evaluation problems resulting from the large number of ranked SS blobs that visually obscure each other when overlaid on the study area (Fig. 3), we suggest that SS-events represent critical thresholds within a hyper-blob, where fundamentally different geometric structures exists both in scale and the landscape. Thus from an ecological perspective, the lifetime of a SS-blob may be considered as levels within a specific scale-domain. To define this domain, each hyper-blob is topologically registered as a unique entity, and its corresponding SS-events are isolated. That is, the first SS-event of all hyper-blobs are geometrically defined regardless of where, and what scale they exist within the stack (i.e., x, y, t). Then the second, third, and nth-events of each hyper-blob are isolated until the last possible event is defined. These event values are then considered as 'scale domain attributes' and are assigned to their corresponding ranked blobs. This domain attribute provides a unique way to query, partition, and evaluate the resulting multiscale 'domain' surface structures, as many blobs can and do exist within a single domain, but no more than one blob can exist within the same 'x, y, z, domain' space. Thus the problem of overlapping ranked blobs is resolved (Fig 3.) and it allows us to evaluate the resulting multiscale surface structures in terms of *critical scale-specific thresholds*.

---

[1] In our processing we generated 100 individual stacks resulting from different random images the same size as the original 500 x 500 pixel IKONOS image. Each random SS stack was composed of 200 layers with a scale increment of one.

In addition, by integrating these hierarchical concepts with geostatistics and 3D visualization techniques, domains can be visually modeled as 'scale-domain manifolds'. To visualize these domain structures, we define the center pixel of each bifurcation blob, apply Delaunay triangulation to all points, and then interpolate with a Quintic polynomial function to generate a smooth surface [Fig. 4 - see Hay et al, (2002b)]. Furthermore, we suggest that this structure correspond to the 'scaling ladder' as conceptualized by Wu (1999) in his description of the Hierarchical Patch Dynamics Paradigm.

3. CONCLUSION

In this paper we describe a novel integration of Scale-Space and Hierarchy Theory for automatically visualizing, defining and modeling dominant landscape structures at multiple scales. Scale-space originates from the computer vision community, where it was developed to analyze real-world structures with no *a priori* information about the scene being assessed. Its basic premise is that a multi-scale representation of a signal (such as a remote sensing image of a landscape) is an ordered set of derived signals showing structures at coarser scales that constitute simplifications of corresponding structures at finer scales. The primary objective of our study has been to define and link structures at different scales in scale-space to higher-order objects, called "scale-space blobs", and to extract significant features based on their appearance and persistence over all scales. Blob-like structures, which persist in scale-space, are likely candidates to correspond to significant structures in the image, and thus in the landscape. Furthermore, by integrating concepts from Scale-Space and Hierarchy Theory, we are able to three-dimensionally model and visualize multiple 'landscape scale-domains' based on the novel idea of using scale-space events as critical domain thresholds. This novel approach provides the capacity to automatically define dominant landscape structures within varying shaped scale-domains, as well as through (all) domains. Spatial statistics are used to describe these significant landscape structures, and 3-D tools have been developed to visualize and describe their multi-dimensional morphology. Our next objective is to ascertain relationships between the dominant patterns within each domain, and the (potential) processes that formed them, in order to better understand the multi-scale dynamics of this landscape, and to evaluate the efficacy of the integrated theory and techniques.

4. ACKNOWLEDGEMENTS

This work has been supported by a team grant from FCAR, Gouvernement du Québec, awarded to Dr. André Bouchard and Dr. Danielle Marceau, and by a series of scholarships awarded to Mr. Hay which include a Marc Bourgie Foundation Ph.D. Award of Excellence, a Biology graduate scholarship of excellence from the University of Montréal, and a GREFi Ph.D. scholarship (*Groupe de recherche en écologie forestière interuniversitaire*). We also thank Mr. Patrick Dubé for his assistance with scale-space coding.

5. REFERENCES

Allen, T.F.H. and Starr, T.B. 1982. Hierarchy Perspective for Ecological Complexity. University of Chicago Press, Chicago, 310 pages.

Bouchard, A. and G. Domon. 1997. The transformation of the natural landscapes of the Haut-Saint-Laurent (Québec) and its implication on future resources management. Landscape and Urban Planning (Elsevier Science Publ.) 37 (1-2): 99-107.

Hay, G. J., P. Dubé, A. Bouchard, and D. J. Marceau. 2002a. A Scale-Space Primer for Exploring and Quantifying Complex Landscapes. Ecological Modelling. In Press.

Hay, G. J., T. Blaschke, D. J. Marceau, and A. Bouchard, 2002b. Fractal net evolution, Scale-Space, and Object-Specific Analysis: A Comparison of Three Image-Object Methods for the Multiscale Analysis of Landscape Structure. ISPRS Journal of Photogrammetry and Remote Sensing. In Review.

Hay, G. J., D. J. Marceau, P. Dubé, and A. Bouchard, 2001. A Multiscale Framework for Landscape Analysis: Object-Specific Analysis and Upscaling. Landscape Ecology (Vol.16, No.6: 471 - 490).

Hay, G. J., K. O. Niemann, and D. G. Goodenough, 1997. Spatial Thresholds, Image-Objects and Upscaling: A Multi-Scale Evaluation. Remote Sensing of Environment, 62: 1-19.

Lindeberg, T. 1994. Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Dordrecht, Netherlands. 423

Marceau, D.J. and Hay, G.J. 1999. Contributions of remote sensing to the scale issues. Canadian Journal of Remote Sensing 25: 357-366.

Wu, J and D. Marceau. 2002. Modelling complex ecological systems: An introduction. Ecological Modelling. In Press.

Wu, J. 1999. Hierarchy and Scaling: Extrapolating Information Along A Scaling Ladder. Canadian Journal of Remote Sensing 25: 367-380.

ASPECTS OF DIGITAL ELEVATION MODEL DETERMINATION

John C. Trinder, Ping Wang, Yi Lu, ZhijunWang, Eric D. Cheng

School of Surveying and Spatial Information Systems
The University of New South Wales, Sydney, Australia
j.trinder@unsw.edu.au

Invited Paper – Commission IV

KEY WORDS: Digital Elevation Models, Kalman filtering, Wavelets, Image Matching, Image Processing, Interferometry.

ABSTRACT:

The paper describes aspects of research being undertaken on the determination and processing of digital elevation models. A 2-D Kalman filter has been developed based on dynamic and functional models for a grid DEM, to produce optimal estimates of terrain attributes of a DEM point using the relevant observations, and predictions derived from its orthogonal neighbouring points. A further method of filtering DEMs based on Wavelets has been developed to compare it with the efficiency of the Kalman filtering method. An intelligent DEM procedure under development comprises several components using image processing, image understanding and terrain modelling. These include the recognition of some buildings on the terrain and the correction of errors in the elevations caused by these buildings, recognition of areas of vegetation on the terrain surface, which will lead to poor quality image matching and hence erroneous elevations. In addition, the geomorphology of the terrain surface, based on the drainage patterns are being derived from the computed DEMs. Further, ERS1/2 Tandem SAR data has been tested as an appropriate data set for upgrading the national digital elevation database over Australia.

1. INTRODUCTION

Digital Elevation Models (DEM) are essential for many aspects of terrain and environmental modelling. They are available in a number of different forms, either as contours on hardcopy maps, digitized contours, and digitized regular or irregular grids. The types of DEMs required by users are often specifically related to their applications. For example, the spacing and accuracy of a user of DEMs who wishes to investigate the natural drainage system of a suburban area requires a different resolution and accuracy DEM than a person who is interested in studying the likelihood frosts in an area. Therefore, users of DEMs often find that the accuracy of available DEMs are inadequate for their purposes and hence they are required to acquire new DEMs at considerable cost, or adapt those available, even though they may be inadequate for their purposes. In addition, users often acquire data from various sources, such as digitized contours without adequate records of the lineage of the data, or indeed even an accuracy statement on the data. Using this data often results in what geographers call 'sinks or depressions' and 'mounds' in the DEM data. This data usually requires editing or smoothing to enable them to be useful.

Photogrammetrists derive DEMs from analogue or digital images and have a good knowledge of the accuracy of the data they derive. They usually do not considered such processes as terrain modelling and drainage pattern delineation within their area of interest, yet the design of theoretical sound methods of DEM processing should be an important task for the people who derive the data. People acquiring DEM data should also be aware of the needs of the users of their data, such as geographers, engineers and environmentalists. Since available DEMs are often inadequate for applications of users, there is a need to develop methods of data acquisition that are cheaper and faster, as well as methods of processing the data, and methods of quality assessment and assurance of DEMs. Digital photogrammetric methods of DEM acquisition have been developed over the last 10 years because of improvements of computer and display technologies. The methods are rapid in terms of the number of points that can be acquired per minute, but they suffer from significant errors caused by such aspects as poor correlation between the images, and buildings and trees on the terrain surface. The time taken to correct the errors in the data often means that the digital methods are only marginally more efficient than the manual methods, or indeed they may even be less efficient if the editing required is lengthy. Considering these issues related to DEMs, this paper describes research on processing and acquisition of DEMs, including procedures for noise reduction by Kalman filtering and Wavelets, the correction of digitally derived DEMs for errors caused by buildings and trees, the extraction of geomorphological parameters, such as drainage patterns as a process of ensuring the quality of determination of digitally derived DEMs, and the determination of elevations from ESA Tandem Mission SAR data by interferometry (InSAR).

2. DEM FILTERING

The 2-D Kalman filter is based on dynamic and functional models for a grid DEM to produce optimal estimates of terrain attributes of a DEM point using the relevant observations, and predictions derived from its orthogonal neighbouring points. It was developed to filter discrepancies in DEMs that may occur due to trees and similar objects on the terrain, thereby generating accurate terrain topographic attributes in gird DEMs (Wang 1998). The output of the process are the estimates of elevation, and the first order of partial derivatives of elevation (ie slopes in E and N directions).

2.1 2-D Kalman Filtering of DEMs

This technique uses the neighbouring DEM points, $p(i-1,j)$ and $p(i,j-1)$, to model any arbitrary point $p(i,j)$. For example, in X direction,

$$H(i,j)=H(i-1,j)+H_x(i-1,j)dx+v_{H_b}(i,j) \tag{1}$$

$$H_x(i,j)=H_x(i-1,j)+v_{H_{x_b}}(i,j) \tag{2}$$

$$H_y(i,j)=H_y(i-1,j)+v_{H_{y_b}}(i,j) \tag{3}$$

Where $H(i,j)$, $H_x(i,j)$, and $H_y(i,j)$ are the elevation, the first partial derivatives of elevation along the X and Y directions of point (i,j) respectively. $H(i-1,j)$, $H_x(i-1,j)$, and $H_y(i-1,j)$ are the elevation, the first partial derivatives of elevation along the X and Y direction of point $(i-1,j)$ respectively. dx is the sampling intervals of the DEM in X direction. v_{H_b}, $v_{H_{x_b}}$, and $v_{H_{y_b}}$ are assumed as white noise sequences, which are to be filtered. An additional three equations were formed in the Y direction.

The 2-D recursive process developed, based on these equations, has shown to be suitable for handling DEM random noise during the terrain modelling process. A DEM outlier detection and removal is an extra function that is added to the 2-D Kalman filter developed in this research. It uses the dispersions between the predicted estimates of elevation and the relevant elevation observations to detect outliers. In addition, a 2-D Kalman smoother is a linear combination of four 2-D Kalman filtering results, derived on the basis of different orientations, that further improves the accuracy of the 2-D Kalman filtering process. The methods have been applied experimentally on a simulated and a real world DEM and proved to be more efficient than existing methods.

2.2 DEM Filtering by Wavelet Transform

Wavelet transforms (WT) have furnished new approaches for describing and modelling the data. Using wavelets, a signal can be split into two or more bands. There are many 2D wavelet transform algorithms available. The most frequently used methods are two fast computation algorithms Mallat algorithm and á trous algorithm. Mallat's algorithm is based on the principle of reducing the redundancy of the information in the transformed data. The processing of decimating in the Mallat algorithm induces the phase distortion which results in edge shift, so it should not be employed to remove noise in DEMs. Á trous algorithm is shift-invariant, which can decompose the original signal into its approximation and its wavelet components at the same scale, and hence can deal with DEM noise removal. Its application to DEM error filtering and its influence on drainage network extraction will be considered in this paper.

2.2.1 Á Trous algorithm Wavelet transform based filtering strategy concerns both the wavelet transform algorithm used, and the processing carried out on the wavelet coefficients. Different strategies can be employed to remove noise from DEMs. Starck et al (2002) compared four wavelet transforms algorithms and eight different strategies for the processing carried out on the wavelet coefficients. Among those transforms, Á trous algorithm does not create artifacts when thresholding and results are significantly better. As opposed to the standard WT method, this transform is isotropic and performs better on isotropic structures.

Á trous algorithm was introduced to decompose the image into wavelet planes [Wang Zhijun, 2000]. Given an image *p*, we can construct the sequence of approximations: $F_1(p)=p_1, F_2(p)=p_2 \ldots$, The wavelet planes are computed as the differences between two consecutive approximations:

$$w_l = p_{l-1} - p_l, (l = 1 \ldots n), p_0 = p \tag{4}$$

we can write the reconstruction formula as:

$$p = p_r + \sum_{l=1}^{r} w_l \tag{5}$$

r refers to the decomposition level.

The discrete filter h is derived from the scale function $\phi(x)$. In our calculations, we select the filter h as a Lagrange Á trous filter, which is in one-to-one correspondence with the convolutional squares of the Daubechies filters for orthogonal wavelets of compact support (Shensa 1992). It is derived by the convolution of an analysis lowpass filter and a reconstruction lowpass filter of Daubechies filter.

2.3. Drainage Network Extraction

Drainage patterns can be used as a means of checking the quality of the derived DEMs and correction of errors that have occurred in the extraction process. There are a number of algorithms available for extracting drainage network from grid based DEM. Basically, all the methods can be classified into two groups: terrain surface geometrical analysis and terrain surface water-flow analysis. In our project, we have used the terrain surface water-flow analysis algorithm based on ARC/INFO grid model. It determines the water flow direction and flow accumulation of each grid segment in the DEM. The flow direction is the steepest slope in each grid segment. Accumulated flow can be computed from the weighted sum of the flow into each down-slope grid segment. Grid segments with a high flow accumulation are areas of concentrated flow and may be used to identify stream network, while segments with a flow accumulation of zero are local topographic highs and may be used to identify ridges. Small depressions have to be smoothed out as they can be seen as errors in the elevations, and can cause serious problems in flow routing. They should be removed by noise reduction methods before the computation of flow direction. Depressions or sinks cannot be simply removed by noise removal, ARC/INFO provides the tools to remove depressions and produce a 'depressionless' DEM.

2.4. Experiments on Drainage Network Extraction

In our research, we have investigated a strategy to remove the DEM noise by wavelet transform, and extracted the drainage networks from the original DEM and the smoothed DEM. We extracted the DEM dataset automatically by L-H Systems digital photogrammetric workstation with a grid space of 5m, a total of 228 x 283 = 64524 points. Figure 1 is the 3D perspective view of the area.

After applying á trous algorithm on the original DEM, the smoothed DEM is shown in Figure 2. Profiles of the original and smoothed DEMs and the overlay of both are shown in Figure 3, from which it can be concluded that á trous algorithm gives a good smoothed result. The original and smoothed DEMs were then transferred into ARC/INFO GRID module, using functions *flowdirection* and *flowaccumulation*.

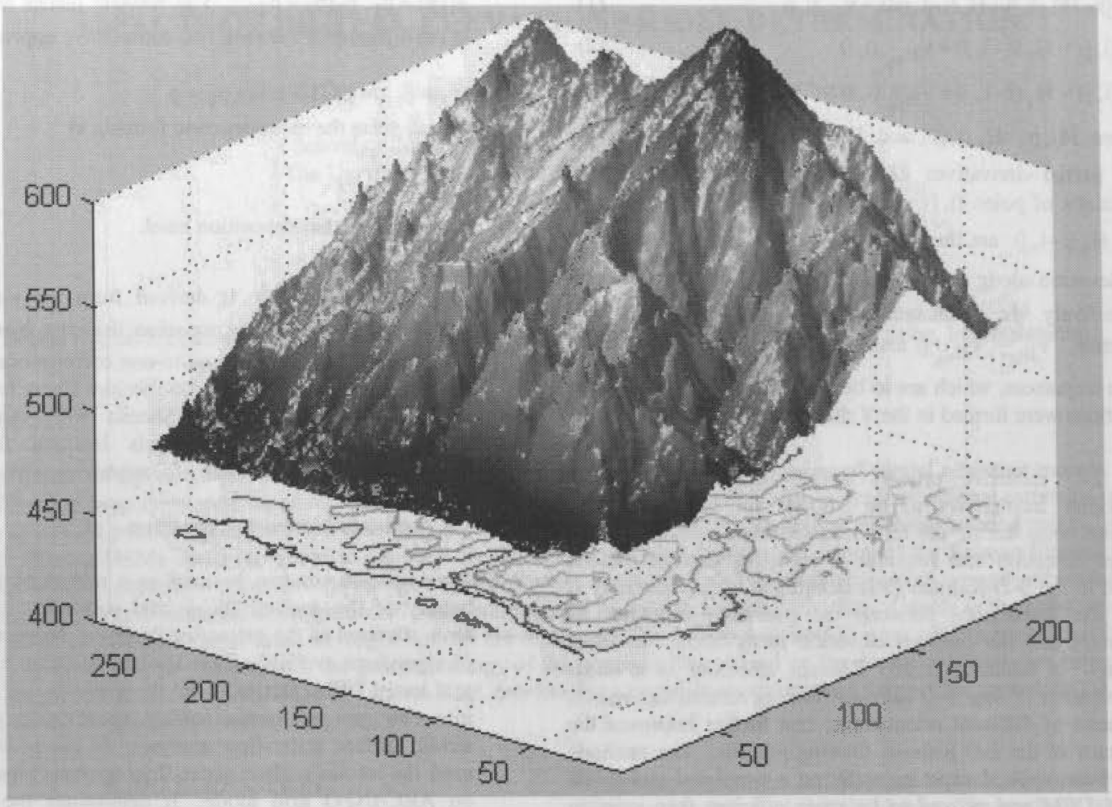

600
550
500
450
400
250
200
150
100
50
200
150
100
50

Figure 1. 3D perspective view of original DEM

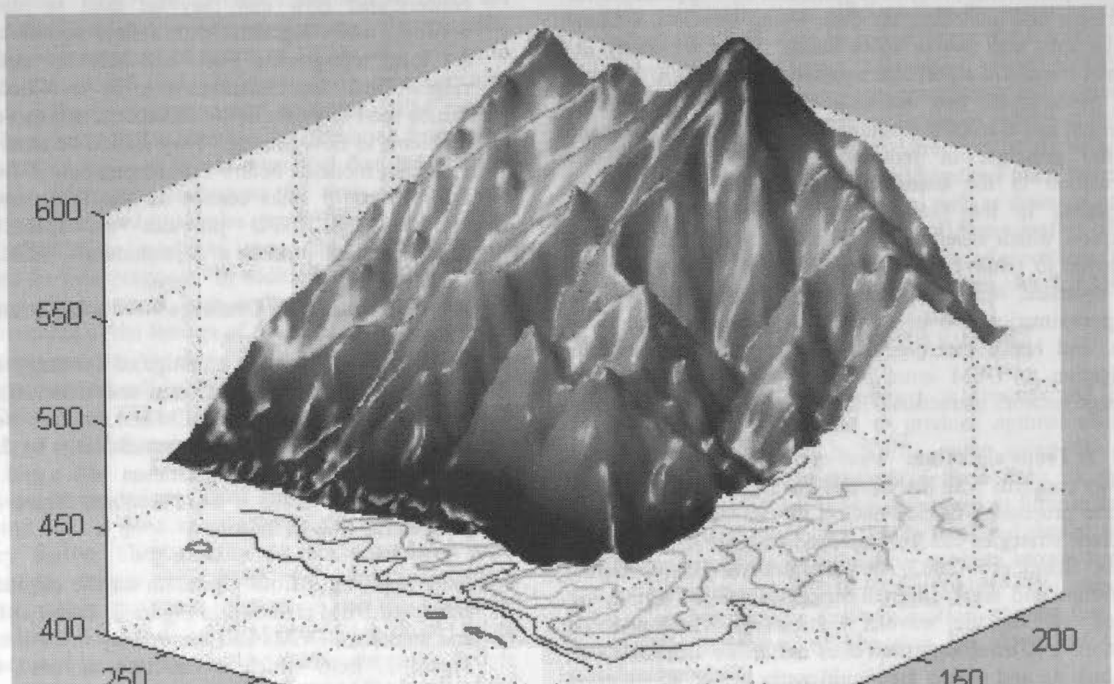

600
550
500
450
400
250
200
150
100
50
200
150
100
50

Figure 2. 3D perspective view of Wavelet transform smoothed DEM

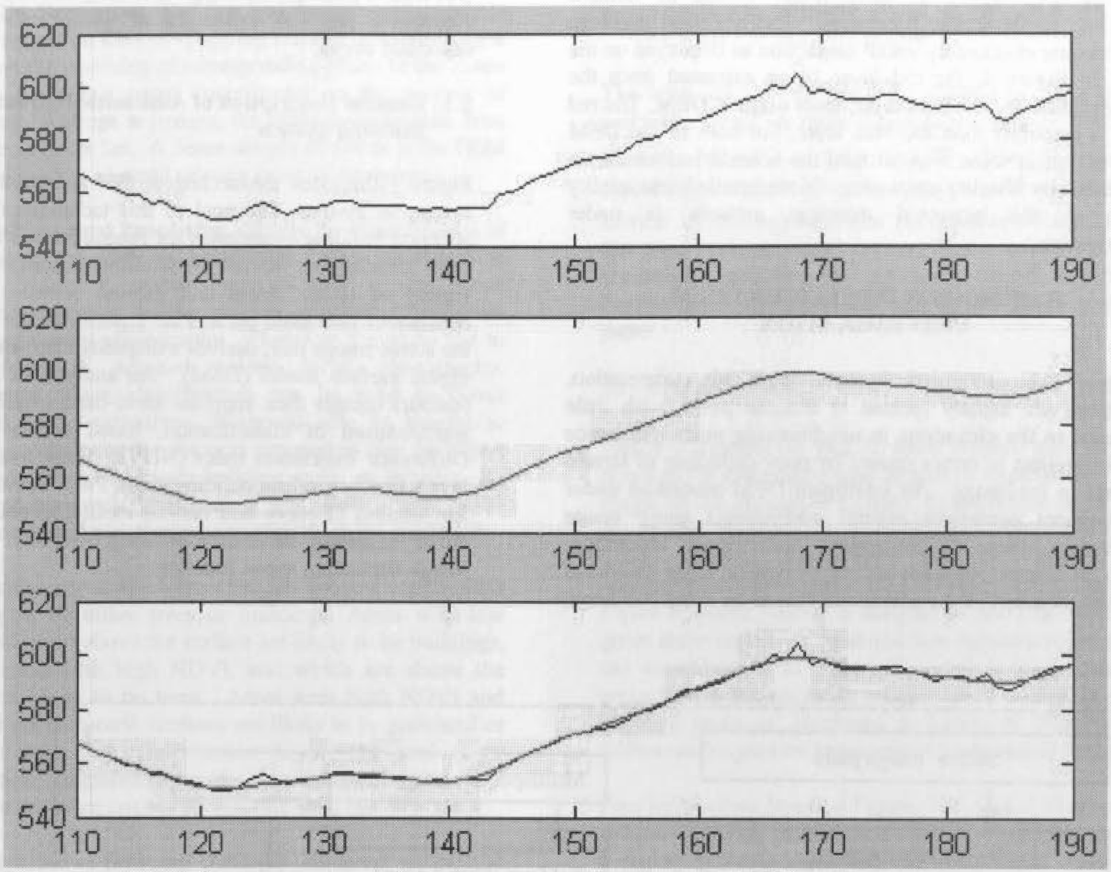


Figure 3: Profiles of the original DEM, Wavelet Transform smoothed DEM and its overlay

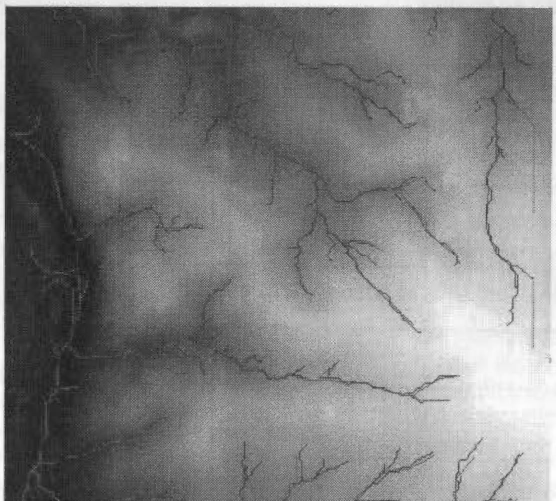


Figure 4 : Overlay of extracted drainage network on original DEM
Red: Drainage network from smoothed DEM
Blue: Drainage network from original DEM

The result of the drainage network extraction is dominated by the threshold, which we have found should be defined according to the terrain roughness. The extracted drainage networks are checked by visual inspection as displayed on the DEM in Figure 4, the red layer being extracted from the smooth DEM and the blue layer, from original DEM. The red layer is smoother than the blue layer, but both fit the DEM well, except in some areas around the boundaries, which are influenced by filtering processing. More detailed data quality check of the extracted drainage network is under investigation.

3. TOWARDS INTELLIGENT DEM DETERMINATION

In digital photogrammetric systems for DEMs computation, elevations are usually located at discrete posts, with little reference to the elevations in neighbouring posts and hence may be subject to errors caused by poor definition of terrain features in the image. An intelligent DEM procedure under development comprises several components using image processing, image understanding and terrain modelling. These processes are designed to recognize some buildings and correct errors in the elevations caused by these buildings, to recognize areas of vegetation on the terrain surface which will lead to poor quality image matching and hence erroneous elevations, and the extraction of terrain morphology, as described above.

3.1 General Description of Automatic Extraction of Building System

Figure 5 illustrates the architecture of an automatic building extraction system. The goal of this technique is to achieve accurate building boundaries for reconstruction of elevations from overlapping aerial or satellite images over a wide variety of terrain types and ground cover. The system consists of two main parts. Part 1 performs the matching of the stereo image pair, derives a disparity map, and produces a digital surface model (DSM). An analysis of multispectral (colour) images then supplies multi-band classification, and segmentation of classification, based on the Normalised Difference Vegetation Index (NDVI). These four information layers finally produce building areas. Part 2 applies a level set formulation of curve and surface motion to obtain an initial curve indicating the desired building boundary, driven by an image-dependent speed function.

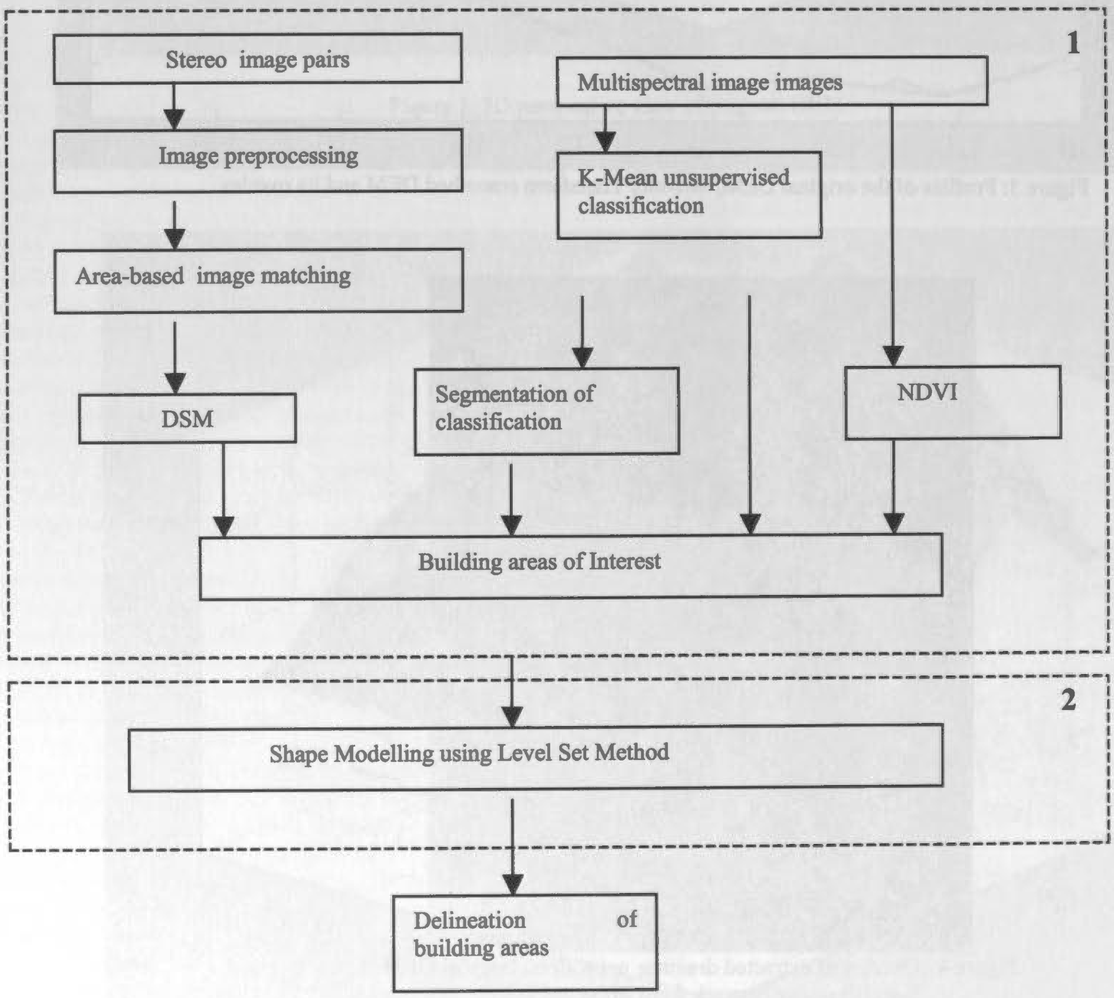

1
Stereo image pairs
Multispectral image images
Image preprocessing
K-Mean unsupervised classification
Area-based image matching
DSM
Segmentation of classification
NDVI
Building areas of Interest
2
Shape Modelling using Level Set Method
Delineation of building areas

Figure 5 Architecture of the building extraction system

3.2 Processing of Stereo and Colour Images

3.3.1 DSM Derivation The first step in the recovery of 3D terrain information from overlapping aerial or satellite images is based on the matching of corresponding pixels in the stereo images. Since this paper concentrates on the process of recognizing buildings in images, the DSM was obtained from L-H Systems Socet Set. A dense sample of points in the DSM is required in order to avoid missing some of structures.

3.3.2 Multispectral Processing Ideally the classification of terrain cover should be undertaken with multispectral imagery. However, most multi-spectral images have insufficient resolution for the location of buildings and trees that occur in aerial photography. Therefore the study had to be restricted to scanned colour aerial photography. Multispectral image classification can be used to detect individual object primitives. It should also be helpful to reduce the complexity for the next processing step for object structure detection. To locate building areas, an iterative K-Means unsupervised classification was used. The NDVI (Vegetation Index) was then used to identify green vegetation.

Simplistically, areas which have heights above a certain limit are likely to be either trees or buildings. Areas with low NDVI, which are above the surface are likely to be buildings, whereas areas with high NDVI, and which are above the surface are likely to be trees. Areas with high NDVI and located at the bare earth surfaces are likely to be grassland or cultivated areas. Four information layers, the land cover classification, segmentation of classification, DSM and NDVI, are therefore combined to locate final building areas.

3.3.3 Location of Building Outlines The level set method for propagating interfaces was introduced by Osher and Sethian (1988). It is based on mathematical and numerical work of curve and surface motion by Sethian (1985) and offers a highly robust and accurate method for tracking interfaces moving under complex motions. Consider a closed curve moving in the plane. Let $\gamma(0)$ be a smooth, closed initial curve in Euclidean plane R^2, and let $\gamma(t)$ be the one-parameter family of curves generated by moving $\gamma(0)$ along its normal vector field with speed $F(K)$, K is a given scalar function of the curvature. Let $x(s,t)$ be the position vector which parameterises $\gamma(t)$ by s, $0 \le s \le S$.

One numerical approach to this problem is to take the Lagrangian description of the problem. One can produce equations of motion for the position vector $x(s,t)$, and then determine the parameterisation with a set of discrete marker particles lying on the moving front. These discrete markers are updated in time by approximating the spatial derivatives in the equations of motion, and advancing their positions. The detailed equations can be found in (Sethian 1999). However, there are several problems with this approach.

- Small errors in the computed particle positions are highly amplified by the curvature term, and calculations become unstable, unless an extremely small time step is employed.
- In the absence of a smoothing curvature term, singularities develop in the propagating front, and an entropy condition must be observed to extract the correct weak solution.
- The topological changes are difficult to manage as the evolving interface breaks and merges.
- Significant problems occur in the extension of this technique to three dimensions.

The level set method has been applied to solve the above problems. Level set method represents the front $\gamma(t)$ as the level set $\{\Phi = 0\}$ of a function Φ. Thus, given a moving closed hypersurface $\gamma(0)$, we produce a formulation for the motion of the hypersurface propagating along its normal direction with speed F. F can be a function of various arguments, including the curvature, normal direction, etc. Further developments of this method will not be given in this paper.

3.2 Test and Results

Figure 6 illustrates a pair of aerial images with 580 × 560 pixels in the row and column directions respectively. The focal length of the camera is 153mm, flying height is 3070m, and the ground resolution is 0.3m. Figure 7 illustrates the DSM image obtained from stereo images. Results of the unsupervised classification using K-Means method are given in Figure 8, where yellow is assigned to building areas, red and green show vegetation areas and blue indicates roads. Based on the classification of Figure 8, using a post classification procedure, a segmentation image can be created to show the building areas as illustrated in Figure 9. The Normalized Difference Vegetation Index values are shown in Figure 10.

Four information layers of Figures 7, 8, 9, and 10 are combined to locate the final building areas are shown in Figure 11. Using a region growing algorithm, the small areas which do not belong to buildings can be deleted from the Figure 11. Further processing is then undertaken using the routine based on the level set to define the outlines of the buildings. The building outlines are then combined with the building areas in Figure 11 to accurately define the buildings. Once the building areas are located, they can be deleted from the DSM to derive the correct DEM of the terrain surface.

4. DEMS BY INTERFEROMETRY FROM TANDEM MISSION DATA

In April 1995 the European Space Agency (ESA) launched its second earth observation satellite, ERS2, which like the first satellite, ERS1, included a C-band Synthetic Aperture Radar (SAR) remote sensing imaging system. The 35 day orbits of the two satellites were programmed for the Tandem Mission, so that there was a 1-day interval between the times ERS1 and ERS2 looked at the same area on the Earth. The ERS Tandem Mission commenced on the August, 1995 and operated until May1996. It provided vast amounts of data that can be used for a number of applications in the geosciences industry, based on the principle of interferometry, (InSAR). InSAR techniques provide a new method of deriving DEMs with little manual input over almost all of the global land surface (including Antarctica). The main objective of the Tandem Mission, was to acquire a unique and valuable set of SAR data that could be used for InSAR purposes. ERS1/2 Tandem SAR data is considered an appropriate data set for upgrading the elevation data over Australia, and hence tests were carried out on the quality of the elevations that can be derived from this dataset.

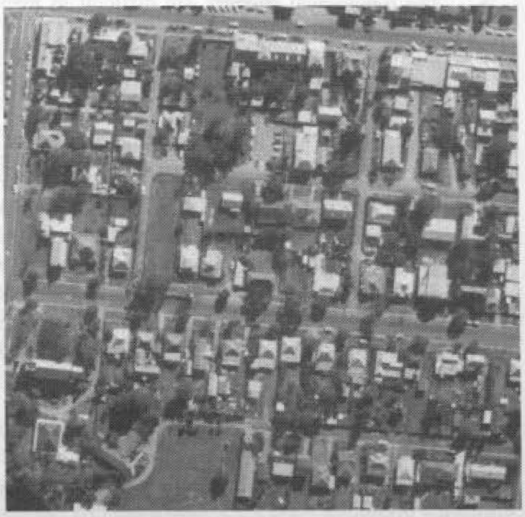


Figure 6 Stereo aerial images

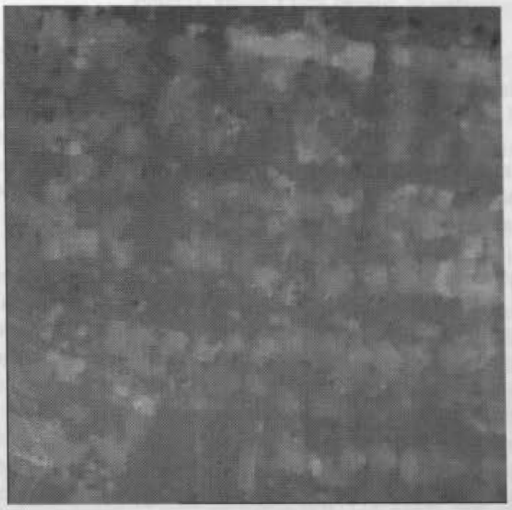


Figure 7 DSM from stereo image matching

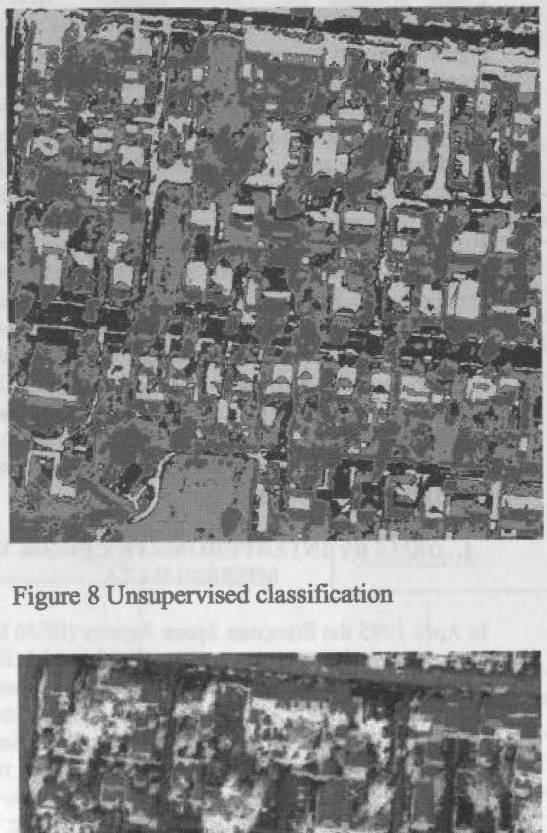


Figure 8 Unsupervised classification

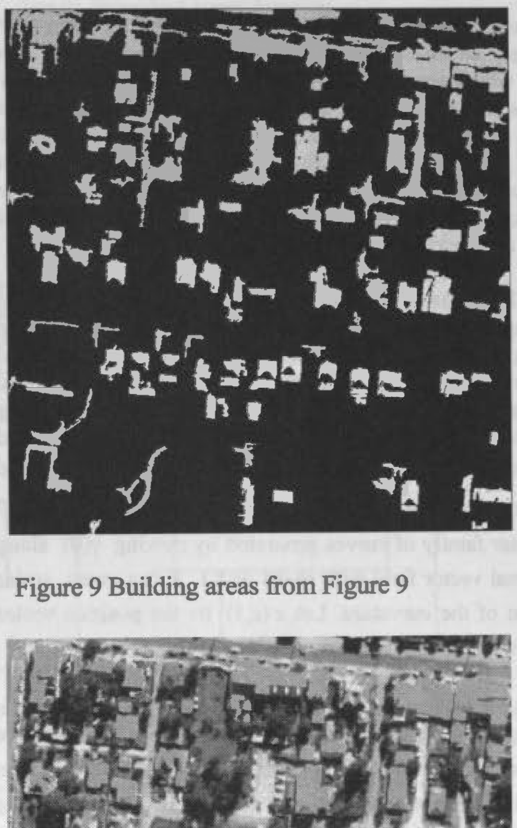


Figure 9 Building areas from Figure 9

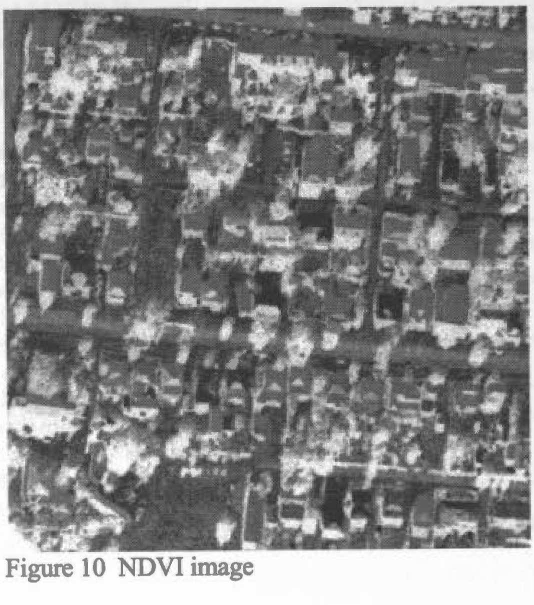


Figure 10 NDVI image

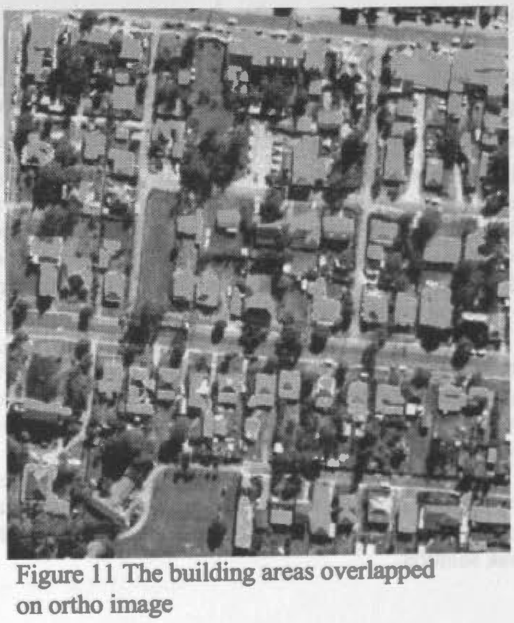


Figure 11 The building areas overlapped on ortho image

4.1 Description of Tests in InSAR Data

The aims of this project were to test the suitability of the method of InSAR, using the ERS1/2 Tandem Mission data for achieving greater efficiencies in upgrading the digital elevation model (DEM) data in Australia. In addition, the objectives were to develop practical methods of using InSAR techniques on the Tandem Mission data in a production environment. The study has involved the use of two software packages, Atlantis EarthView™ and Phoenix Systems Pulsar™, for processing the Tandem ERS1/2 data. A third package, Vexcel is also available in Geoscience Australia. Tests were carried out on 13 image pairs of Tandem Mission data, some of which overlap with each other. The procedure for computation of DEMs involved the determination of the coherence between the pairs of images, referred to as master and slave images, the computation of the interferograms from the phase information of the pixels, phase unwrapping to convert the interferogram values to elevations. In the case of the Atlantis EarthView™ software it was then necessary to locate Ground Control Points (GCPs) to reduce the elevations to the height datum. The maps available for the identification of the GCPs varied from 1:25,000 to 1:100,000, which meant that the elevations of the GCPs had accuracies varying from 3m to 8m. The selection of GCPs was a difficult task, since the identification of map features on the radar images, where their appearance depended on the reflectance characteristics of the terrain surface, presented problems. A maximum of 100-200 GCPs were usually identifiable on the maps and an image pair. PulSAR software did not require the use of GCPs, since it uses the satellite ephemeris to locate the DEM on the terrain surface. If required the precise ephemeris can also be obtained from ESA to improve the positioning of the DEM. PulSAR also uses a coarse DEM of the area to assist in the phase unwrapping and provide the initial values of the DEM.

The mean errors and estimated accuracy of DEMs in this study were determined by comparing the computed elevations of the GCPs and check points located manually on existing topographic maps, with their known values, and also by overlaying the derived DEM onto a known DEM. The known DEM available is the AUSLIG 9" DEM national database of Australia with an estimated accuracy of about 10m. In all cases in these tests, accuracies were quoted as the RMS of the differences between the elevations derived from topographic maps of the area, or the known DEM of the area, and the computed elevations by interferometry.

4.2 Achieved Accuracy of DEMs by InSAR

Conditions found necessary for determination of DEMs from InSAR are as follows:

- The perpendicular baseline between the orbits of the two data passes should be less than 300m, preferably 100-200m
- Ground control points should be available for Atlantis software with an accuracy compatible with the expected accuracy of the computed elevations
- Good coherence between the pairs of images is required, otherwise large elevation errors will occur
- The terrain conditions for the two acquisitions should not vary significantly, otherwise temporal decorrelation will render the images unsuitable for InSAR applications.

As a result of the tests using PulSAR software on full scenes (110kmx40km), accuracies of elevations of the order 10m were achievable with the InSAR technique from ERS1/2 Tandem Mission data, based on the mean of elevations obtained from at least two pairs of images of the same area. Individual pairs of images resulted in accuracies of the order of 15m, but on some occasions poorer accuracies occurred. The accuracies achieved are comparable with similar tests in Europe, based on the use of ERS1/2 Tandem data. In some cases large errors existed in the computed DEM, but these were always in areas where coherence between the two images was low and the elevations of many points in the same region could not be determined satisfactorily.

Altantis software resulted in significantly lower accuracy elevations, of the order of 30m for a full scene and 10-15m for areas of ¼ of a scene. However, the need to use GCPs reduced the accuracy of the elevations and slowed the process very significantly. For many parts of Australia, a sufficient number of accurate GCPs would not be available from existing maps to provide control for Atlantis software.

5. FINAL REMARKS

This paper has given a review of research being undertaken on DEMs at UNSW. Given the importance of elevation data as an layer in GIS, to many activities involving a study of terrain form, such as civil engineering, environmental monitoring and terrain modelling, efficient methods of data acquisition, as well as data quality assurance, require investigation and development. Some aspects of these issues have been presented in this paper, including, filtering DEMs, determination of terrain cover and terrain morphology to increase the intelligence of the DEM acquisition process, and the computation of elevations from a readily available data source, the ESA ERS1/2 Tandem Mission SAR data by interferometry.

6. REFERENCES

Osher S. & Sethian J.A. (1988) "Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulation", *J. of Computational Phhysics*, Vol. 79, pp.12-49.

Sethian J. A. (1985) "Curvature and the evolution of fronts", *Comm. In Math. Phys. 54,* pp.87-499.

Sethian J. A. (1999) " Level set methods and fast marching methods", Cambridge University Press.

Shensa M. J., The discrete Wavelet Transfrom: Wedding the Á trous and Mallat algorithms, IEEE TRANS. On Siginal Processing, Vol 40(10):2464-2482, 1992.

Starck J.-L., Murtagh F., (1988) Image filtering by combining multiple vision models, http://www.mathsoft.com/signal_processing.htm

Wang, P., (1998). Applying Two Dimensional Kalman Filtering on Digital Elevation Models, *International Archives of Photogrammetry and Remote Sensing,* Vol.32, Part 4, pp649-656.

Wang Zhijun, (2000) Wavelet Transform Based Multi-sensor Remote Sensing Image Fusion, PhD thesis, Wuhan University, P.R.China.

(2) A [illegible] full scenes [illegible] under 10m [illegible] from ERS1/2 [illegible] observations [illegible] same area. [illegible] of the order [illegible] occurred, [illegible] these tests in [illegible] In some [illegible] fact there [illegible] the two [illegible] shown in the [illegible]

[illegible] accuracy [illegible] around 10-15m [illegible] the GCPs reduced the accuracy of the elevations and slowed the process very significantly. For many parts of Australia, a sufficient number of accurate GCPs would fail to be available thus requiring users to provide control for Atlantis software.

[illegible] undertaken [illegible] data [illegible] a study of [illegible] commercial [illegible] of data [illegible] supply [illegible] more [illegible] DEMs, [illegible] technology to [illegible] and [illegible] data [illegible] data for [illegible]

Osher, S. & Sethian J.A. (1988). "Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulation", J. of Computational Physics, Vol. 79, pp.12-49.

[illegible] of Radar",

[illegible] mapping,

[illegible] Wrapping the [illegible] the Digital

[illegible] by modeling

[illegible] the Optimal Kalman [illegible] Archives [illegible] Vol. 32, Part 4,

[illegible] Fields versus [illegible] Wales

4.1 De[illegible]

The aim [illegible] purpose of [illegible] for selecting [illegible] elevations [illegible] objectives [illegible] techniques [illegible] environment [illegible] packages [illegible] Fusion™ [illegible] package [illegible] Tests were [illegible] data, such as [illegible] for comparison [illegible] obtaining [illegible] and stereo [illegible] the phase [illegible] convert the interferogram values to elevations. In the case of the Atlantis EarthView™ software it was then necessary to locate Ground Control Points (GCPs) to reduce the elevations to the height datum. The same materials for the identification [illegible] which [illegible] varying [illegible] difficult [illegible] radar [illegible] reflectors [illegible] problems [illegible] identifiable [illegible] did not [illegible] spheres [illegible] required [illegible] ESA to [illegible] uses a [illegible] unwrapping [illegible]

The [illegible] study [illegible] elevation [illegible] existing [illegible] by overlaying the derived DEM onto a known DEM. The known DEM available is the AUSLIG 9" DEM national database of Australia with an indicated accuracy of some 10m. In all cases in these tests, accuracies were quoted as the RMS of [illegible] and the [illegible]

4.2 Adv[illegible]

Conclusions [illegible] InSAR and [illegible]

- The [illegible] data, [illegible] 200[illegible]
- Due [illegible] software [illegible] areas [illegible]
- Care [illegible] obtaining [illegible]
- The [illegible] vary [illegible] render the images unusable for DEM generation.

WG IV / 7:

DATA INTEGRATION AND DIGITAL MAPPING

3D MODELS FOR HIGH RESOLUTION IMAGES: EXAMPLES WITH QUICKBIRD, IKONOS AND EROS

Th. Toutin [a], *, R. Chénier [ab], Y. Carbonneau [ab]

[a] Natural Resources Canada, Canada Centre for Remote Sensing, 588Booth Street, Ottawa, Ontario, K1A 0Y7 Canada - thierry.toutin@ccrs.nrcan.gc.ca
b Under contract with Consultants TGIS inc., 7667 Curé Clermont, Anjou, Québec, H1K 1X2 Canada – (rene.chenier; yves.carbonneau)@ccrs.nrcan.gc.ca

Commission IV, WG IV/7

KEY WORDS: 3D models, Parametric and non-parametric, High resolution, QuickBird, IKONOS, EROS

ABSTRACT:

EROS, IKONOS and Quickbird are the three civilian satellites, which presently provide panchromatic images with the highest spatial resolution: 2-m, 1-m and 0.6-m, respectively. They also have off-nadir viewing up-to-60° in any azimuth depending on the sensor, which enables stereo images (along and across track) to be acquired. However, the image acquisition system produces different geometric distortions, which need to be accurately corrected. The paper reviews the image distortions and the different 3D models (non-parametric and parametric) for the geometric processing with their applicability to high-resolution images depending on the type of images and their pre-processing level. Results are then presented with stereo images for the three sensors using different 3D models. In general, the 3D parametric models achieve more consistent results.

RÉSUMÉ :

EROS, IKONOS et Quickbird sont les trois satellites civils, qui fournissent des images panchromatiques avec la meilleure résolution spatiale : 2 m, 1 m et 0,6 m, respectivement. Ils peuvent aussi dépointer jusqu'à 60° dans tous les azimuts pour acquérir des images stéréoscopiques (dans le sens de l'orbite ou perpendiculairement). Par contre, le système d'acquisition des images crée différentes distorsions géométriques, qui doivent être corrigées avec précision. L'article présente les distorsions géométriques et les différents modèles 3D (non-paramétrique et paramétrique) pour le traitement géométrique, ainsi que leur applicabilité aux images de haute résolution suivant le type d'images et leur niveau de prétraitement. Des résultats sont alors donnés avec des images stéréoscopiques des trois capteurs avec les différents modèles 3D. En général, les modèles paramétriques 3D donnent des résultats plus cohérents.

1. INTRODUCTION

The generation of high-resolution imagery using previously-proven defence technology provides an interesting source of data for digital topographic mapping as well as thematic applications such as agriculture, forestry, and emergency response (Kaufmann und Sulzer, 1997; Konecny, 2000). Instead of using aerial photos, highly detailed maps of entire countries can be frequently and easily updated using these data. Petrie (2002) gives a review and comparisons of the characteristics of this new generation of high-resolution satellites.

However, high-resolution images usually contain so significant geometric distortions that they cannot be used directly with map base products into a geographic information system (GIS). Consequently, multi-source data integration (raster and vector) for mapping applications requires geometric and radiometric models and processing adapted to the nature and characteristics of the data in order to recover the original information from each image in the composite geocoded image.

1.1 Image Distortions

The source of distortions can be related to two general categories: **the Observer** or the acquisition system (platform, and imaging sensor) and **the Observed** (atmosphere and Earth). In addition to these distortions, the deformations related to the map projection have to be taken into account because the terrain and most of GIS end-user applications are generally represented and performed in referenced topographic maps.

The distortions caused by the platform are mainly related to the variation of the elliptic movement around the Earth, caused at first order by the variations of the Earth gravity (Escobal, 1965). Depending of the acquisition time and the size of the image, the variation of the elliptic movement has a various impact on the image distortion. Some effects include:

- the altitude variations in combination with the focal length, the flatness and the relief of the Earth change the pixel spacing;
- the attitude variations in the roll, pitch and yaw axes change the orientation and the shape of high resolution images;
- the velocity variations change the line spacing or create line gaps/overlaps in the images.

The distortions caused by the imaging sensor are:

- the calibration parameters, such as the focal length and the instantaneous field of view (IFOV);
- the panoramic distortion in combination with the oblique-viewing system, the Earth curvature and the topographic

relief changes the ground pixel sampling along the column.

The distortions caused by the Earth are:

- the rotation generates lateral displacements in the column direction between image lines depending of the latitude;
- the curvature creates variation in the image pixel spacing;
- the topographic relief generates a parallax in the scanning azimuth.

The deformations caused by the map projection are:

- the approximation of the geoid by a reference ellipsoid;
- the projection of the reference ellipsoid on a tangent plane.

1.2 Image Correction Models

All these geometric distortions require models and mathematical functions to perform the geometric corrections of an image: either with 2D/3D non-parametric models or with rigorous 3D parametric models. With 3D parametric models, the geometric correction can be corrected step-by-step with a mathematical function for each distortion or all together with a "combined" mathematical function.

Since the 2D polynomial functions do not reflect the sources of distortion during the image formation and for the relief, they are limited to images with few or small distortions, such as nadir-viewing images, small images, systematically-corrected images, flat terrain. They also are very sensitive to input errors. The 2D polynomial functions were mainly used in the 70's and 80's on images, whose systematic distortions, excluding the relief, where already corrected by the image providers. Since it is well known that 2D polynomial functions are not suitable for accurately correcting remote sensing images, especially high-resolution images, only 3D models are addressed in this paper. Finally, some results on ortho-images generation and 3D data extraction will be presented with these high-resolution images.

2. 3D NON-PARAMETRIC MODELS

The 3D non-parametric models can be used when the parameters of the acquisition systems or a rigorous 3D parametric model are not available. Since these models do not require *a priori* information on any component of the total system (platform, sensor, Earth, map projection), they do not reflect the source of distortions described previously. Consequently, due to the lack of physical meaning, the interpretation of the parameters are difficult (Madani, 1999).

These non-parametric models are based on different XYZ mathematical functions:

- For the 3D polynomial functions, P_{3D}:

$$P_{3D}(XYZ) = \sum_{i=o}^{m} \sum_{j=o}^{n} \sum_{k=o}^{p} a_{ijk} X^i Y^j Z^k \qquad (1)$$

- For the 3D rational functions, R_{3D}:

$$R_{3D}(XYZ) = \frac{\sum_{i=o}^{m} \sum_{j=o}^{n} \sum_{k=o}^{p} a_{ijk} X^i Y^j Z^k}{\sum_{i=o}^{m} \sum_{j=o}^{n} \sum_{k=o}^{p} b_{ijk} X^i Y^j Z^k} \qquad (2)$$

where: X, Y, Z are the terrain or cartographic coordinates;
i, j, k are integer increments;
m, n and p are integer values; and
m+n+p is the order of the polynomial functions.

The order of the polynomial functions is generally less than three because higher orders do not improve the results. The 3D 1st, 2nd and 3rd order polynomial functions will then have 4, 10 and 20-unknown terms. The 3D 1st, 2nd and 3rd order rational functions will then have 7, 19 and 39-term unknowns. In some conditions, specific terms, such as XZ, YZ^2 or Z^3, could be dropped of the polynomial functions, when these terms can not be related to any physical element of the image acquisition geometry; it thus reduces potential correlations between terms.

2.1 3D Polynomial Models

The 3D polynomial functions are an extension of the 2D polynomial function by adding Z-terms related to the third dimension of the terrain. However, they are subjected to the same problems related to 2D non-parametric functions: application to small images, need lot of ground control points (GCPs) regularly distributed, correct locally at GCPs, very sensitive to input errors, lack of robustness, etc. Their use should be thus limited to small images or to systematically-corrected images, where all systematic distortions except the relief were corrected. They were applied with georeferenced images, such as SPOT-HRV (level 1B) (Palà and Pons, 1995) and IKONOS Geo-products (Hanley and Fraser, 2001).

The terms related to terrain elevation in the 3D polynomial functions could be reduced to a_iZ for high-resolution images, whatever the order of the polynomial functions used.

2.2 3D Rational Models

These 3D rational functions have recently drawn interest in the civilian photogrammetric and remote sensing community due to the launch of new civilian high-resolution sensors. The major reason of their recent interest is that some image vendors, such as Space Imaging do not release information on the satellite and the sensor. The 3D rational functions can be used with two approaches:

1) To approximate an already-solved existing 3D parametric model; and
2) To determine by least-squares adjustment the coefficients of the polynomial functions (equation 2) with GCPs.

The first approach is performed in two steps. A 3D regular grid of the imaged terrain is first defined and the image coordinates of the 3D grid ground points are computed using the already-solved existing 3D parametric model. These grid points and their 3D ground and 2D image coordinates are then used as GCPs to resolve the 3D rational functions and compute the unknown terms of polynomial functions.

This approach has been proven adequate for aerial photographs or satellite images (Tao and Hu, 2001). However, they found that the results are sensitive to GCP distribution with satellite images. When the image is too large, the image itself has to be subdivided and separate 3D rational functions are required for each subdivided image. It sometimes results in "less user-friendly" processing than a direct 3D parametric model. Image vendors or government agencies, which do not want to deliver satellite/sensor information with the image, utilize this approach. They thus provide with the image all the parameters of 3D rational functions. Consequently, the end-users can directly process the images for generating ortho-images or DEM and also post-process these products to improve their final accuracy. However, this approach is useless for the end-users because it requires the knowledge of a 3D parametric model. In this condition, the end-users can directly apply the 3D parametric model. Furthermore, the approximation of a 3D parametric model will not generally be as precise than a 3D parametric model by itself, depending on the image, its distortions and its processing level.

The second approach can be performed by the end users with the same processing method than with polynomial functions. Since there are 38 to 78 parameters (two equations (2) for column and for line) for the four 2nd and 3rd order polynomial functions, a minimum of 19 and 39 GCPs respectively are

polynomial functions, rational functions do not model the physical reality of the image acquisition geometry and are sensitive to input errors. Consequently, much more GCPs are needed to reduce error propagation in operational environment.

Such as the 3D polynomial functions, the rational functions mainly correct locally at the GCPs, and errors and inconsistencies between GCPs can be found (Davies and Wang, 2001). They should not be used with raw and large-size images but only with small-size or georeferenced/geocoded images. Otherwise, a piecewise approach as described previously should be used for large raw images, and the number of GCPs should be increased proportional to the number of sub-images. However, the 3D rational functions are certainly the best selection among the non-parametric functions, when 3D parametric solution is not available.

3. 3D PARAMETRIC MODELS

Although each sensor has its own specificity, one can drawn generalities for the development of 3D parametric functions, in order to fully correct all distortions described previously. The 3D parametric functions should model the distortions of the platform (position, velocity, and attitude), the sensor (viewing angles, panoramic effect), the Earth (ellipsoid and relief) and the cartographic projection. The geometric correction process can address each distortion one by one and step by step or all together. In fact, it is better to consider the total geometry of viewing: platform + sensor + Earth + map, because some of the distortions are correlated (Toutin, 1995). It is theoretically more precise to compute one "combined" parameter than each individual component of this "combined" parameter, separately.

As examples of combined parameters, we have:

- the orientation of the image is a combination of the platform heading due to orbital inclination, the yaw of the platform, the convergence of the meridian;
- the scale factor in along-track direction is a combination of the velocity, the altitude and the pitch of the platform, the detection signal time of the sensor, the component of the Earth rotation in the along-track direction; and
- the levelling angle in the across-track direction is a combination of platform roll, the viewing angle, the orientation of the sensor, the Earth curvature.

The general starting points of these research studies to derive the 3D parametric functions are generally the well-known collinearity condition and equations (Light *et al.*, 1980), which are only valid for a scanline acquisition. However, the parameters of neighbouring scanlines of scanners are highly correlated, it is thus possible to link the exposure centres and the rotation angles of the different scanlines with supplemental information. Ephemeris and attitude data can be integrated using 2nd order polynomial functions (Konecny *et al.*, 1986) or benefiting from theoretical work in celestial mechanics (Toutin, 1995).

4. APPLICATIONS TO HIGH RESOLUTION IMAGES

4.1 QuickBird

Raw images with detailed metadata are provided to the end users (DigitalGlobe, 2002). The adaptation and application of 3D parametric models already developed for push-broom scanner (Konecny *et al.*, 1986; Toutin, 1995) can be easily done. On the other hand, it is not advised to use 3D non-parametric models, which do not reflect the geometry of viewing and are sensitive to GCP errors and distribution.

DigitalGlobe provided a test image (16 x 17 km) acquired over Reno, USA for testing the CCRS 3D parametric model capability (Toutin, 1995). The image is raw-type with 61-cm pixel spacing. However, the sensor resolution seems better. As an example, Figures 1 & 2 show a sub-image (61-cm pixel spacing) over an urban area and a sub-image (10-cm pixel spacing) resampled six times, respectively. The quality and the details of Figure 2 give an idea of the sensor resolution and easily demonstrate the high mapping potential of this data.

The image is presently processed at CCRS and preliminary results using 22 10-cm accurate GPS GCPs with rational and CCRS-parametric models are only given. 3D polynomial models were not used on raw images with large distortions. The results (Table 1) show the adaptability and the superiority of CCRS 3D parametric model for QuickBird. The parametric model was not sensitive to the number and distribution of GCPs, while the rational model was. Since the main errors come from GCP definition and plotting (around 1-2 pixels), sub-metre accuracy could thus be achieved with better-defined GCPs.

Correction Method	RMS (m)		Maximum (m)	
	X	Y	X	Y
Rational 1st order	4.0	2.1	9.5	4.3
Parametric	1.4	1.3	2.5	2.8

Table 1. Comparison of RMS and maximum errors over 12 ICPs from 1st order rational model and Toutin's parametric model computations with 10 GCPs

An in-track stereo pair has just been acquired over an area North of Quebec City, Canada with good control data and laser DTM. It is a residential and semi-rural environment with a hilly topography (500-m elevation range). More results on 3D parametric and non-parametric models, ortho-image generation and 3D extraction (DTM, canopy and building heights) will be shown at the Symposium. Due to the raw image type and its size (16 x 17 km), a piecewise approach should be used with the rational models to improve the previous results (4-m accuracy in Table 1), but leading to the increase of GCP number.

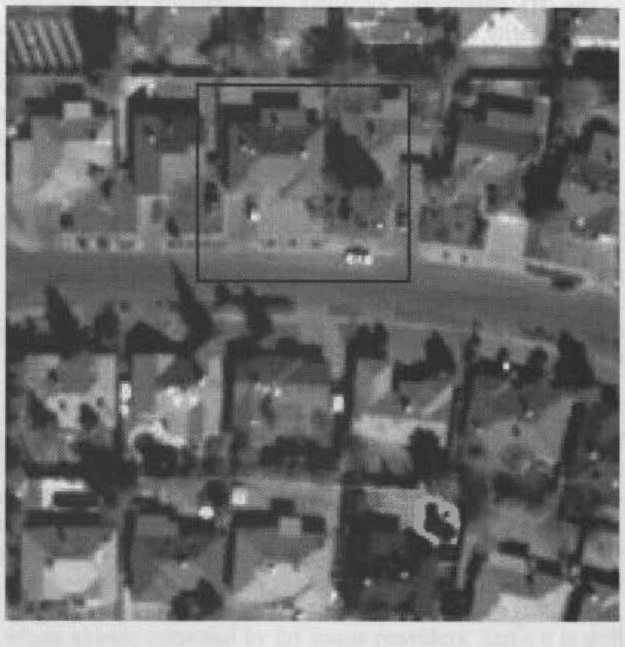


Figure 1: Sub-image (200 x 200 pixel; 61-cm pixel spacing) of Quickbird image acquired over Reno, USA. Courtesy and copyright of DigitalGlobe, 2002

4.2 IKONOS

The raw images are not available and the basic product provided by Space Imaging is the map-oriented Geo product. It has an accuracy of 50-m CE90, which means that any point within the image is within 50 meters horizontally of its true position on the earth's surface 90% of the time (Space Imaging, 2002). Accuracy becomes worse in mountainous areas if the images are acquired with off-nadir viewing, which is quite common for the IKONOS data. Hence, the product will only meet the geometric requirements of mapping scale at 1:100,000.

The fact remains that detailed sensor information for the IKONOS satellite has not yet been released. Despite this, the integrated and unified 3D parametric model (Toutin, 1995) developed at CCRS, Natural Resources Canada has successfully been adapted to IKONOS Geo product using basic information from the metadata and image files (Toutin, 2001). The main reason is that this CCRS model integrated the correlated parameters into a reduced number of "combined" parameters.

A 10x10-km Geo image (1-m pixel) acquired over semi-flat area has been processed with the three 3D models (parametric and non-parametric), using 30 50-cm accurate GCPs. A 3D polynomial model can be used because IKONOS images are already georeferenced. Table 2 shows the RMS and maximum errors over the 23 ICPs of the three models computed with only seven GCPs. The errors are smaller with the parametric model than with the polynomial or rational models. Furthermore, the maximum errors are much greater with non-parametric models. This shows that the parametric model is both stable and robust over the full image without generating local errors while the non-parametric models mainly cancel the errors locally at the GCPs but can generate large local errors elsewhere.

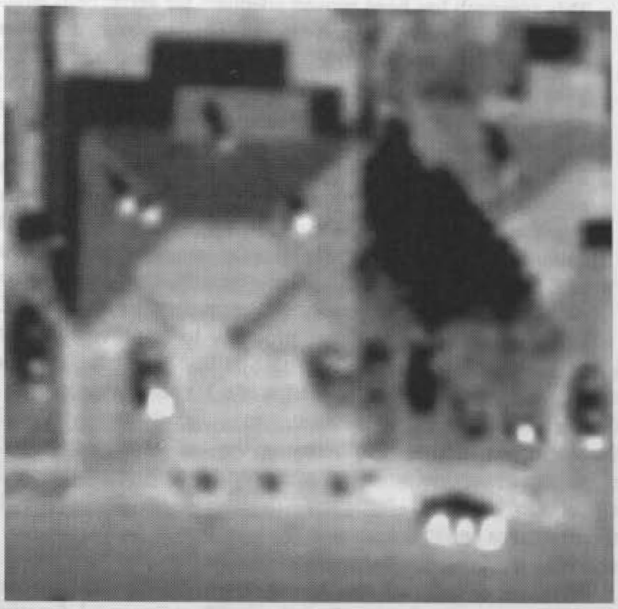


Figure 2: Partial resampling of Figure 1 using 16pt. sin(x)/x kernel (400 x 400 pixels; 10-cm pixel spacing). Courtesy and copyright of DigitalGlobe, 2002

Correction Method	RMS (m)		Maximum (m)	
	X	Y	X	Y
Polynomial 2nd order	1.8	2.4	4.1	7.9
Rational 1st order	2.2	5.2	5.1	10.4
Parametric	1.3	1.3	3.0	3.0

Table 2. Comparison of RMS and maximum errors over 23 ICPs of 2nd order polynomial, 1st order rational and Toutin's parametric model computations with 7 GCPs

4.3 EROS

The EROS-A imaging concept, by which the space scanlines are acquired, derived from military technology to follow a target. Since the satellite is moving "too fast" according to the ground sampled distance, the orientation of the linear array sensor is continuously backward pitching to acquire a continuous image without gap between the scanlines (ImageSat Intl., 2002). Consequently, the distortions due to attitude will be much higher than with the normal push-broom scanners. Therefore, more precise attitude model is requested to integrate linear and non-linear variations of the rotation angles in the 3D model. This point already suggests than non-parametric models, which do not reflect the image geometry and attitude distortions, will not be able to accurately correct such large attitude distortions, and especially their high-frequency variations.

A raw sub-image (3x3 km; 2-m pixel) acquired over a flat area was processed with the CCRS 3D generalized sensor model, using 24 50-cm accurate GCPs. A small image size was voluntary chosen to see the adaptability of rational functions to this type of image. Table 3 gives the RMS and maximum errors over the 14 ICPs of the two models computed with 10 GCPs. The RMS and maximum errors are much smaller with the parametric model than with the rational model. Even applied on a small image size and over a flat area, the resulting accuracy (4-6 pixels) of the 1st order rational model is poor. In addition, a great variability in the results of this rational model and local errors were noticed when varying the GCP number and distribution, while the parametric model was stable.

Correction Method	RMS (m)		Maximum (m)	
	X	Y	X	Y
Rational 1st order	8.0	13.2	20	23
Parametric	3.9	3.5	6.2	6.0

Table 3. Comparison of RMS and maximum errors over 14 ICPs of 1st order rational model and Toutin's parametric model computations with 10 GCPs

5. CONCLUSIONS

3D models (parametric or non-parametric) were used to correct geometric distortions of new high-resolution images. The non-parametric models (polynomial or rational) can be applied on small-size images when the systematic distortions are already corrected, such as IKONOS Geo product. The rational models can be applied on raw images, such as QuickBird or EROS. Parametric models can be applied to any image type or size.

Among the non-parametric models, the rational models are certainly the most appropriate, only if parametric models are not available. However, results over three different high-resolution images (level of pre-processing, pixel spacing, size) show that the 3D non-parametric models are less precise than the 3D parametric model, such as Toutin's model developed at CCRS. While non-parametric models could have given similar results than parametric models in specific and limited conditions and in a well-controlled environment where input errors are limited, these present results confirm that they can be non-consistent, unstable and sensitive to GCP number and distribution (see also Madani, 1999; Davies and Wang, 2001). In operational environment where input errors are common, the parametric models should be primarily used because they will insure, in addition of better accuracy, more robustness and consistency whatever the data and its level of pre-processing.

ACKNOWLEDGEMENTS

The authors thank DigitalGlobe and CORE Technologies for providing the Quickbird and EROS images, respectively to adapt and test 3D Toutin's parametric model developed at CCRS. They also thank Dr. Philip Cheng of PCI Geomatics for integrating 3D Toutin's parametric model and algorithms for high-resolution images into Geomatica OrthoEngine^{SE} of PCI.

REFERENCES

Davies C. H., and X. Wang, 2001. Planimetric Accuracy of IKONOS 1-m Panchromatic Image Products, *Proceedings of the 2001 ASPRS Annual Conference*, St. Louis, MI, USA, April 23 - 27, CD-ROM.

DigitalGlobe, 2002. http://www.digitalglobe.com/ (accessed 06 March 2002).

Escobal, P.R., 1965. *Methods of orbit determination*, Krieger Publishing Company, Malabar, USA, 479 pages.

Hanley H.B. and Fraser C.S., 2001. Geopositioning accuracy of IKONOS imagery: Indications from two-dimensional transformations, *Photogrammetric Record*, 17(98), pp. 317-329.

ImageSat Intl., 2002. (accessed 06 March 2002), http://www.imagesatintl.com/1024/company/eros.html

Kaufmann, V., und W. Sulzer, 1997. Über die Nutzungsmöglichkeit hochauflösender amerikanischer Spionage-Satellitenbilder (1960-1972). *Vermessung und Geoinformation*, 3, pp.166-173.

Konecny G., 2000. Mapping from Space, *Remote Sensing for Environmental Data in Albania: A Strategy for Integrated Management*, NATO Science Series, Vol. 72, Kluwer Academic Publishers, pp. 41-58.

Konecny, G., E. Kruck and P. Lohmann, 1986. Ein universeller Ansatz für die geometrische Auswertung von CCD-Zeilenabtasteraufnahmen, *Bildmessung und Luftbildwesen*, 54(4), pp. 139-146.

Light, D.L., D. Brown, A. Colvocoresses, F. Doyle, M. Davies, A. Ellasal, J. Junkins, J. Manent, A. McKenney, R. Undrejka and G. Wood, 1980. Satellite Photogrammetry, in *Manual of Photogrammetry 4th Edition, Chapitre XVII*, Editor in chief: C.C. Slama, American Society of Photogrammetry Publishers, Falls Church, USA, pp. 883-977.

Madani, M., 1999. Real-Time Sensor-Independent Positioning by Rational Functions, *Proceedings of ISPRS Workshop on Direct Versus Indirect Methods of Sensor Orientation*, Barcelona, Spain, November 25-26, pp. 64-75.

Palà V. and X. Pons, 1995. Incorporation of Relief in Polynomial-Based Geometric Corrections, *Photogrammetric Engineering & Remote Sensing*, 61(7), pp. 935-944.

Petrie, G., 2002. Optical Imagery from Airborne & Spaceborne Platforms. *GeoInforamtics*, 5(11), pp. 28-35.

Space Imaging, 2002. (accessed 06 March 2002), http://www.spaceimaging.com/carterra/geo/

Tao V. and Y. Hu, 2001. 3-D Reconstruction Algorithms with the Rational Function Model and their Applications for IKONOS Stereo Imagery, *Proceedings of ISPRS Joint Workshop "High Resolution Mapping from Space"*, Hannover, Germany, September 19-21, CD-ROM, pp. 253-263.

Toutin, Th., 1995. Multisource data fusion with an integrated and unified geometric modelelling. *EARSeL Journal Advances in Remote Sensing*, 4(2):118-129. http://www.ccrs.nrcan.gc.ca/ccrs/eduref/ref/bibpdf/1223.pdf

Toutin Th., 2001, Geometric processing of IKONOS Geo images with DEM, *Proceedings of ISPRS Joint Workshop "High Resolution Mapping from Space"*, Hannover, Germany, September 19-21, CD-ROM, pp. 264-271. http://www.ccrs.nrcan.gc.ca/ccrs/eduref/ref/bibpdf/13116.pdf

IMAGE FUSION USING WAVELET TRANSFORM

Zhu Shu-long [a, b]

[a] Department of Remote Sensing Information Engineering, Zhengzhou Institute of Surveying & Mapping, Zhengzhou, China, Postcode: 450052 -zsl1964@public2.zz.ha.cn

[b] LIESMARS of Wuhan University, Wuhan, China, Postcode: 430079

Commission IV, Working Group IV/7

KEY WORDS: Multi-source images, Fusion, Wavelet transform, Edge processing, Wavelet based fusion, Fused Images

ABSTRACT:

Extracting more information from multi-source images is an attractive thing in remotely sensed image processing, which is recently called the image fusion. In the paper, the image fusion algorithm based on wavelet transform is proposed to improve the geometric resolution of the images, in which two images to be processed are firstly decomposed into sub-images with different frequency, and then the information fusion is performed using these images under the certain criterion, and finally these sub-images are reconstructed into the result image with plentiful information.

1. THE INTRODUCTION

For the remotely-sensed images, some have good spectral information, and the others have high geometric resolution, how to integrate the information of these two kinds of images into one kinds of images is a very attractive thing in image processing, which is also called the image fusion. For the purpose of realization of this task, we often need some algorithms to fuse the information of these two kinds of images, and however we find few such algorithms. In recent ten years, a new mathematical theory called "Wavelet theory" (Charles K. Chui, 1992) has gradually been used in the fields of graphics and imagery, and been proved to be an effective tool to process the signals in multi-scale spaces (Thierry Ranchin et al.). In this paper , the image fusion algorithm based on wavelet transform is proposed to improve the geometric resolution of the images, in which two images to be processed are firstly decomposed into sub-images with the same resolution at the same levels and different resolution among different levels, and then the information fusion is performed using high-frequency sub-images under the "gradient" criterion, and finally these sub-images are reconstructed into the result image with plentiful information. Since the geometric resolution of the image depends on the high-frequency information in it, therefore this image fusion algorithm can acquire good results.

2. THE ALGORITHM OF WAVELET TRANSFORM

2.1 Mallat Algorithm of Wavelet Transform

Suppose that $\{V_j, j \in Z\}$ is a multi-resolution Analysis in $L^2(R)$, $\varphi(x)$ is the scaling function of subspace V_0 , W_j is the orthogonal complement of V_j with respect to V_{j+1} , i.e. $V_{j+1} = V_j + W_j$, $\psi(x)$ is wavelet function of subspace W_0 . If a signal $f(x) \in V_{j+1}$, then it can be expressed as :

$$f(x) = \sum_n c_n^{j+1} \varphi_{j+1,n} \tag{1}$$

since $V_{j+1} = V_j + W_j$, then $f(x)$ can also be expressed as:

$$f(x) = \sum c_n^j \varphi_{j,n} + \sum d_n^j \psi_{j,n} \tag{2}$$

Combining formula (1) with (2), we get ：

$$\sum_n c_n^{j+1} \varphi_{j+1,n} = \sum_n c_n^j \varphi_{j,n} + \sum d_n^j \psi_{j,n} \tag{3}$$

since $\varphi_{j,k}$ is orthogonal with respect to different j and k, if two sides of formula (3) are multiplied by $\varphi_{j,k}$ and then integrated with respect to x, we can obtained :

$$c_k^j = \sum_n c_n^{j+1} \left\langle \varphi_{j+1,n}, \varphi_{j,k} \right\rangle = \frac{1}{\sqrt{2}} \sum_n h_{n-2k} c_n^{j+1} \tag{4}$$

Using the same method, we also have:

$$d_k^j = \frac{1}{\sqrt{2}} \sum_n g_{n-2k} c_n^{j+1} \tag{5}$$

The formulas (4) and (5) are the decomposition formula of signal, where c^j is an approximation of c^{j+1}, and d^j is the detailed part of c^{j+1}.

If having the decomposed signals $\{c_n^j, n \in Z\}$ and $\{d_n^j, n \in Z\}$, then two sides of formula (3) are multiplied by $\varphi_{j+1,k}$, and then integrated with respect to x, we can obtained :

$$\begin{aligned} c_n^{j+1} &= \sum_k c_k^j \left\langle \varphi_{j,k}, \varphi_{j+1,n} \right\rangle + \sum_k d_k^j \left\langle \psi_{j,k}, \varphi_{j+1,n} \right\rangle \\ &= \frac{1}{\sqrt{2}} \sum_n c_k^j h_{n-2k} + \frac{1}{\sqrt{2}} d_k^j g_{n-2k} \\ &= \frac{1}{\sqrt{2}} \left(\sum_k c_k^j \tilde{h}_{2k-n} + \sum d_k^j \tilde{g}_{2k-n} \right) \end{aligned} \tag{6}$$

Where $\tilde{h}_n = h_{-n}$, $\tilde{g}_n = g_{-n}$, when the signal is finite, $\tilde{h}_n = h_{1-n}, \tilde{g}_n = g_{1-n}$. The formula (6) are the reconstruction formula of signal. The formulas (4)、(5) and (6) are called Mallat algorithm (S. Mallat, 1989).

For two-dimensional signal c^{j+1}, its decomposition formulas are :

$$\begin{aligned} c_{m,n}^j &= \frac{1}{2} \sum_{k,l \in Z} c_{k,l}^{j+1} h_{k-2m} h_{l-2n} \\ d_{m,n}^{j1} &= \frac{1}{2} \sum_{k,l \in Z} c_{k,l}^{j+1} h_{k-2m} g_{l-2n} \\ d_{m,n}^{j2} &= \frac{1}{2} \sum_{k,l \in Z} c_{k,l}^{j+1} g_{k-2m} h_{l-2n} \\ d_{m,n}^{j3} &= \frac{1}{2} \sum_{k,l \in Z} c_{k,l}^{j+1} g_{k-2m} g_{l-2n} \end{aligned} \tag{7}$$

and its reconstruction formula is ：

$$c_{m,n}^{j+1} = \frac{1}{2} \left(\sum_{k,l \in Z} c_{k,l}^j \tilde{h}_{2k-m} \tilde{h}_{2l-n} + \sum_{k,l \in Z} d_{k,l}^{j1} \tilde{h}_{2k-m} \tilde{g}_{2l-n} + \sum_{k,l \in Z} d_{k,l}^{j2} \tilde{g}_{2k-m} \tilde{h}_{2l-n} + \sum_{K,l \in z} d_{k,l}^{j3} \tilde{g}_{2k-m} \tilde{g}_{2l} \right. \tag{8}$$

where c^j is an approximation of c^{j+1}, d^{j1}、d ^{j2} and d^{j3} are the detailed parts of c^{j+1}. Fig. 1 illustrates the Mallat algorithm of 2D images intuitively.

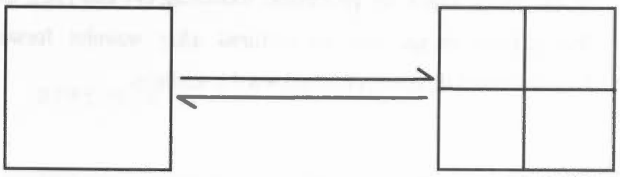


Figure 1. The Mallat algorithm of 2D image

2.2 Edge Processing for Wavelet Transform

Since our image fusion algorithm need to use the forward and the inverse algorithm of wavelet transform, in order to reconstruct the signals precisely, the image edges must be processed under the process of wavelet transform, otherwise some information will be lost in this operation.

Assuming that the low-pass filter and the corresponding high-pass filter are four coefficient Daubechies' (I. Daubechies, 1988), i.e. $H=(h_0,h_1,h_2,h_3)$ and $G=(g_{-2},g_{-1},g_0,g_1)$ respectively, the samples of the original discrete signal are $C^0=(c_0^0,c_1^0,\ldots,c_{M-1}^0)$, then the image edges are processed according the following precise formulas (Zhu Shu-long and Geng zhe-Xun, 1999) in our experiments:

For left image edges, we have:

$$\begin{cases} c_{-1}^0 = c_{-2}^0 = 0 \\ c_{-1}^1 = \dfrac{h_2}{h_1} d_0^1 \end{cases}$$

(9)

For right image edges, we have:

$$\begin{cases} c_M^0 = 0, \quad d_{M/2}^1 = 0 \\ c_{M+1}^0 = -\dfrac{h_2}{h_0} c_{M-1}^0 + \dfrac{h_3}{h_0} c_{M-2}^2 \end{cases}$$

(10)

If the image edges are processed according (9) and (10), then the original image can be restored after wavelet forward transform and then wavelet backward transform.

3. IMAGE FUSION ALGORITHM BASED WAVELET TRANSFORM

The main idea of our algorithm is that: (1) the two images to be processed are resampled to the one with the same size; and (2) they are respectively decomposed into the subimages using forward wavelet transform, which have the same resolution at the same levels and different resolution among different levels; and (3) information fusion is performed based on the high-frequency subimages of decomposed images; and finally the result image is obtained using inverse wavelet transform.

Let $A(x,y)$ and $B(x,y)$ be the images to be fused, the decomposed low-frequency sub-images of $A(x,y)$ and $B(x,y)$ be respectively $lA_J(x,y)$ and $lB_J(x,y)$(J is the parameter of resolution),the decomposed high-frequency sub-images of $A(x,y)$ and $B(x,y)$ be respectively $hA_j^k(x,y)$ and $hB_j^k(x,y)$ (j is the parameter of resolution and $j=1,2,\cdots,J$. for every j, $k=1,2,3$), the gradient image generated from $hA_j^k(x,y)$ and $hB_j^k(x,y)$ be respectively $GA_j^k(x,y)$ and $GB_j^k(x,y)$, then the fused high-frequency sub-images $F_j^k(x,y)$ are :

$$\text{if } GA_j^k(x,y) > GB_j^k(x,y)$$

$$F_j^k(x,y) = hA_j^k(x,y),$$

$$\text{if } GA_j^k(x,y) < GB_j^k(x,y)$$

$$F_j^k(x,y) = hB_j^k(x,y),$$

and the fused low-frequency sub-image $F_J(x,y)$ are :

$$F_J(x,y) = k1 \cdot lA_J(x,y) + k2 \cdot lB_J(x,y)$$

where $k1$ and $k2$ are given parameters, if image B is fused into image A, then $k1 > k2$. If the decomposed high-frequency sub-images $hA_j^k(x,y)$ of $A(x,y)$ is replaced by be respectively $F_j^k(x,y)$, the decomposed low-frequency sub-image $lA_J(x,y)$ of $A(x,y)$ is replaced by be respectively $F_J(x,y)$, and $F_J(x,y)$ and $F_j^k(x,y)$ are used to reconstruct the image $A'(x,y)$ using the inverse wavelet transform, then $A'(x,y)$ not only has the low-frequency information of $A(x,y)$ and $B(x,y)$, but also has the high-frequency information of $A(x,y)$ and $B(x,y)$ at the same time. This algorithm shows that the high-frequency information fusion between two images is completed under the "gradient" criterion.

For multi-spectral images, we can use the same method proposed above to process every band images respectively. For color images, we can convert the color images from RGB (i.e.

Red, Green, Blue) space to HIS (i.e. Hue, Intensity, Saturation) space, and then only process the intensity component of HIS space using the same method proposed above, and finally acquired the fused the images.

4. THE ASSESSMENT FOR THE FUSED IMAGES

The index "Average gradient" is applied to measure the detailed information in the images. The formula of average gradient "g" can be denoted as follows:

$$g=\frac{1}{(M-1)\cdot(N-1)}\sum_{x=1}^{M-1}\sum_{y=1}^{N-1}\sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2+\left(\frac{\partial f(x,y)}{\partial y}\right)^2} \quad (11)$$

Where $f(x,y)$ is image function, M and N are the length and width of the image respectively. Generally, the larger g is, the clearer the image is. For color images, average gradient "g" can be calculated from the "intensity" images.

5. THE EXPERIMENTS

In our experiments, two images are used to verify our algorithm, the first is the image composed of SPOT multi-spectral images with the resolution of 20 m (as shown in Fig.2), the second is SPOT panchromatic image with the resolution 10 m (as shown in Fig.3). In order to fully utilize the spectral information of the former and geometric information of the latter, the image fusion algorithm proposed above is used to introduce the details of the latter into the former so that the result image (as shown in Fig.4) has not only good spectral information but also good geometric information. The table 1 illustrates the average gradient "g" in these three images quantitatively, in which the average gradient of the result image is obviously greater than that of the original image 1.

Now this algorithm has been a practicable function in our Image Processing System (IPS). The further research is to adapt this algorithm to the large-data images.

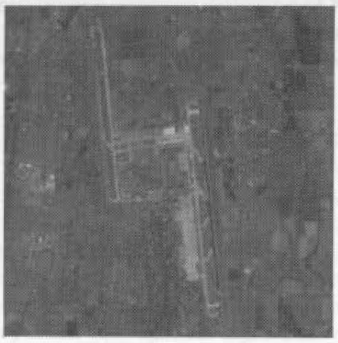
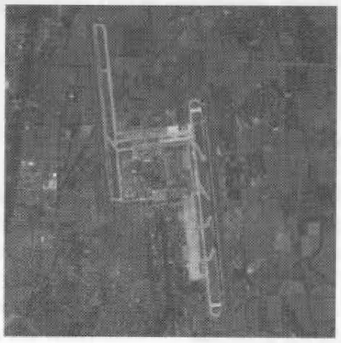


Fig. 2. The original image 1 with the resolution 20 m



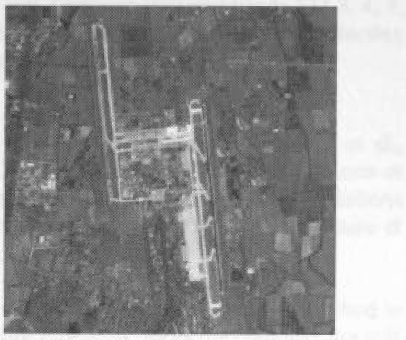


Table 1 the image and its average gradient

Image Name	The average gradient
the original image 1 with the resolution 20 m	4.78
the original image 2 with the resolution 10 m	12.33
the result image generated by our algorithm	11.16

REFERENCES:

Thierry Ranchin et al., Efficient data fusion using wavelet transform: the case of SPOT satellite images, in Mathematical imaging: Wavelet Application in Signal and Image Processing, Andrew F. Laine, Editor, Proceeding of SPIE 2034, pp171-179.

Charles K. Chui, 1992, An Introduction to Wavelets, Academic Press, Inc., America.

S. Mallat, 1989, A theory for multiresolution signal decomposition: The wavelet representation, IEEE Transaction on PAMI, Vol.11, No.7, pp.674-693.

I. Daubechies, 1988, Orthonormal bases of compactly supported wavelets, Communication on Pure and Applied Mathematics, Vol. XLI, pp.909-996.

Zhu Shu-long and Geng zhe-Xun, 1999, The Wavelets and Its Applications to Graphics and Imagery, The PLA Press, P.R.C.

LINEAR SPECTRAL MIXTURE ANALYSIS OF TM DATA FOR LAND-USE AND LAND-COVER CLASSIFICATION IN RONDÔNIA, BRAZILIAN AMAZON

Dengsheng Lu,[1] Mateus Batistella,[2,3] and Emilio Moran[1,2]

[1]Center for the Study of Institutions, Population, and Environmental Change (CIPEC)
Indiana University, Bloomington, Indiana, USA
Telephone: 812-856-5767, Fax: 812-855-2634
E-mail: dlu@indiana.edu

[2]Anthropological Center for Training and Research on Global Environmental Change (ACT)
Indiana University, Bloomington, Indiana, USA

[3]Brazilian Agricultural Research Corporation, EMBRAPA Satellite Monitoring
Campinas, São Paulo, Brazil

Commission IV

KEY WORDS: linear spectral mixture analysis, land use, land cover, classification, Thematic Mapper, Brazilian Amazon

ABSTRACT:

The mixed pixels in remotely sensed data are one of the main error sources resulting in poor classification accuracy using traditional classification methods. In order to improve classification accuracy, linear spectral mixture analysis (LSMA) has been used to handle the mixed pixel problems. This paper aims to achieve an appropriate processing routine of LSMA through the comparison of classification results derived from different processing methods, i.e., constrained and unconstrained least-square solutions, different numbers of endmembers, different image bands used, and minimum noise fraction (MNF) transformation. A Landsat Thematic Mapper (TM) image of June 18, 1998, was used, and field data were collected in Rondônia, Brazilian Amazon. Seven classes are defined: mature forest, intermediate secondary succession (SS2), initial secondary succession (SS1), pasture, agriculture, water, and bare land (including urban areas, roads, and bare soil for cultivation). This study indicates that using constrained or unconstrained least-square solutions, atmospherically corrected or raw TM images in LSMA do not produce significant difference in the overall classification accuracy. However, reducing correction between image bands used in LSMA is useful in improving fraction quality and classification accuracy. Selection of four endmembers (green vegetation, shade, bright soil, and dark soil) and bands TM 3, 4, 5, and 7 provided the best classification accuracy. The overall classification accuracy reached 86%. This study shows that selecting appropriate endmembers and image bands is crucial for developing high quality fraction images using LSMA.

INTRODUCTION

Mixed pixels are common in remotely sensed data because of limitations in spatial resolution of the data and the heterogeneity of features on the ground. The mixture spectra are often generated when the pixel covers more than one land-cover class. This problem often produces poor classification accuracy when conventional algorithms such as the maximum likelihood classifier (MLC) are used. In moist tropical regions, classification of stages of secondary succession (SS) is especially difficult because of its variations in vegetation stand structure, species composition, and biomass (Mausel et al., 1993). However, it is possible to obtain better results if the mixed pixels are decomposed into different proportions of selected components. In order to solve the mixed pixel problem, scientists have developed different models to unmix the pixels into different proportions of the endmembers (Ichoku and Karnieli, 1996). Linear spectral mixture analysis (LSMA) is one of the most often used methods for handling the spectral mixture problem. It assumes that the spectrum measured by a sensor is a linear combination of the spectra of all components within the pixel (Roberts et al., 1998; Ustin et al., 1998). LSMA is a physically based image analysis process that supports repeatable and accurate extraction of quantitative subpixel information (Adams et al., 1986; Smith et al., 1990; Roberts et al., 1998; Ustin et al., 1999). It has been used for vegetation or land-cover classification (Ustin et al., 1996; Cochrane and Souza, 1998; Aguiar et al., 1999; Petrou, 1999; Ustin et al., 1999; DeFries et al., 2000; Ustin and Xiao, 2001; Theseira et al., 2002) and for change detection (Adams et al., 1995; Roberts et al., 1997; Roberts et al., 1998; Ustin et al., 1998; Elmore et al., 2000; Rogan et al., 2002).

Although LSMA has been recognized as an effective method in handling spectral mixture problems, some uncertainties are still not fully understood. For example, which unmixing solutions (e.g., constrained and unconstrained least-square solutions) can improve the quality of fraction images? How many endmembers are suitable for land-use and land-cover (LULC) classification? Can image transformation (e.g., minimum noise fraction or MNF) improve the fraction quality? Is atmospheric correction of remotely sensed data required before using LSMA? This paper contributes to these debates through comparisons of LULC classification results based on fraction images, which are developed from different image processing routines in Rondônia, Brazilian Amazon.

STUDY AREA AND FIELD DATA COLLECTION

Rondônia has had the highest deforestation rate in the Brazilian Amazon during the last twenty years (Dale and Pearson, 1997). Following the national strategy of regional occupation and development, colonization projects initiated by the Brazilian government in the 1970s played a major role in this process (Schmink and Wood, 1992). Most colonization projects in the state were designed to settle landless migrants. Settlement began in this area in the mid-1980s, and the immigrants transformed the landscape into a mosaic of forest remnants, cultivated crops, pastures, and fallow land. The data used in this study were collected in Machadinho d'Oeste in northeastern Rondônia (Figure 1).The climate in Machadinho d'Oeste is classified as equatorial hot and humid, with tropical transition. The well-defined dry season lasts from June to August, and the annual average precipitation is 2,016.6 mm (Rondônia, 1998). The annual average temperature is 25.5°C, and monthly averages for air moisture range from 80 to 85 percent. The terrain is undulated, ranging from 100 m to 450 m above sea level. Several soil types were identified, mainly alfisols, oxisols, ultisols, alluvial soils, and other less spatially represented associations (Miranda and Mattos, 1993). Settlers, rubber tappers, and loggers inhabit the area, transforming the landscape through their economic activities and use of resources.

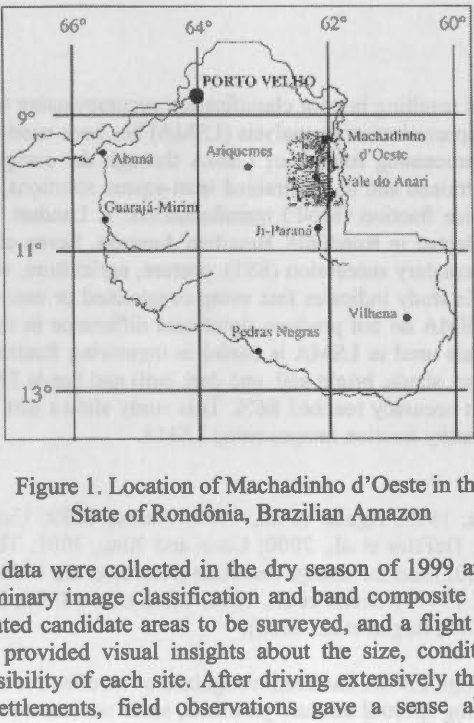


Figure 1. Location of Machadinho d'Oeste in the State of Rondônia, Brazilian Amazon

Field data were collected in the dry season of 1999 and 2000. Preliminary image classification and band composite printouts indicated candidate areas to be surveyed, and a flight over the areas provided visual insights about the size, condition, and accessibility of each site. After driving extensively throughout the settlements, field observations gave a sense about the structure of regrowth stages, mainly regarding total height and ground cover of dominant species. Indicator species, such as *Cecropia sp.*, *Vismia sp.*, palms, grassy vegetation, and lianas also helped to assign the SS stages. Every plot was registered with a Global Positioning System (GPS) device to allow integration with spatial data in Geographic Information Systems (GIS) and image processing systems. Detailed information regarding data gathering can be found in Batistella (2001). The field data were randomly separated into two groups. One group was used for training data for supervised classification, and another group was used for accuracy assessment.

LULC CLASSIFICATION USING LSMA

The application of LSMA to LULC classification involves image preprocessing, endmember selection, constrained or unconstrained least-square unmixing solutions, and classification. TM imagery for June 18, 1998, was radiometrically and atmospherically calibrated into surface reflectance using an improved image-based DOS model (Lu et al., in press). The image was geometrically rectified based on control points taken from topographic sheets at 1:100,000 scale (UTM south 20 zone). Nearest-neighbor resampling technique was used. The root-mean-square (RMS) error was smaller than 0.5 pixel.

A variety of methods were used to determine endmembers (Smith et al., 1990; Quarmby et al., 1992; Boardman, 1993; Roberts et al., 1993; Settle and Drake, 1993; Boardman et al., 1995; Bateson and Curtiss, 1996). Some previous literature has also discussed and summarized the methods of endmember selection (Adams et al., 1993; Tompkins et al., 1997; Mustard and Sunshine, 1999). For many applications of LSMA, image endmembers are often used because they can be obtained easily, representing spectra measured at the same scale as the image data (Roberts et al., 1998). Image endmembers were derived from the extremes of the image feature space, assumed to represent the purest pixels in the images (Roberts et al., 1998; Mustard and Sunshine, 1999). Three endmembers (shade, soil, and green vegetation or GV) were identified from the scattergram of bands TM 3 and TM 4 and scattergram of bands TM 4 and TM 5. An average of 4 to 10 pixels of these vertices was calculated. When selecting the endmembers, cautions need be taken to identify outliers.

In order to develop high-quality fraction images, different image transformations can be used (Cochrane and Souza, 1998; Van der Meer and de Jong, 2000). The MNF transformation is one of the often used methods for reducing redundancy of information between image bands and assisting selection of endmembers. In this paper, the main objective is to find a proper image processing method for LULC classification using LSMA. Therefore, six different processing methods were tested and their classification results were compared. The following routines were defined:

1. Raw-c: using constrained LSMA method with three endmembers (GV, shade, and soil) on raw six-band TM image;
2. Ref-c: using constrained LSMA method with three endmembers on six-band TM reflectance image;
3. Ref-uc: using unconstrained LSMA method with three endmembers on six-band TM reflectance image;
4. MNF-c: using constrained LSMA method with three endmembers on the first four MNF components;
5. Subset-c: using constrained LSMA method with three endmembers on image band subset (bands TM 3, 4, 5, and 7); and
6. Subset-4c: using constrained LSMA method with four endmembers (GV, shade, bright soil, and dark soil) on image band subset (bands TM 3, 4, 5, and 7).

Seven LULC classes—mature forest (MF), intermediate secondary succession (SS2), initial secondary succession (SS1), pasture (including cultivated and degraded pastures), agriculture (including coffee and cacao plantations and other crops), water, and bare land (including urban areas, roads, and bare soil for cultivation)—were defined according to the LULC characteristics of the study area. The classification was conducted on the fraction images using a maximum likelihood classifier.

Error matrices are often used to assess classification accuracy by comparing the relationships between ground-truth data (reference data) and classified results (Congalton, 1991; Janssen and Van der Wel, 1994; Jensen, 1996; Smits et al., 1999). Producer's accuracy (a measure of omission error) and user's accuracy (a measure of commission error) were calculated based on error matrices. The reference data covering different LULC classes were collected during fieldwork. These data were linked to the TM image, using the IMAGINE AOI (area of interest) tool to create test samples. Each AOI was associated with an integer number corresponding to a given LULC class. Then, a reference data image was created and compared to the classified image pixel by pixel to create the error matrix table for accuracy assessment.

RESULTS AND DISCUSSION

The endmember fractions were developed using LSMA based on different image processing routines. Figure 2 gives an example illustrating the fraction characteristics of some typical LULC classes in the study area. In the soil fraction, bare land has significantly higher values, while different successional and mature forests have very small fraction values. Pastures and agriculture such as coffee plantations have relative higher fraction values than those of successional and mature forests. In the GV fraction, water and bare land have small fraction values. Mature forest has the lowest GV fraction values (after bare land and water), and SS1 has the highest GV fraction values. In the shade fraction, water has the highest value and bare land the lowest value. Mature forest has significantly higher fraction value, and SS1 and pastures have lower fraction values. SS2 and agriculture have lower values than mature forest but higher values than pasture and SS1. The error fraction indicates that high quality fraction images are obtained and the results are reliable because the error value for all LULC classes is very small. This figure indicates that water and bare land are the two classes that have the most different characteristics in the fraction image compared with any other LULC classes. The distinction between vegetation types (SS and mature forest) and pasture or agricultural land is better in the soil fraction, but the distinction between different SS stages and mature forest or between pasture and agriculture is better in the GV and shade fractions. Other fraction images derived from different image processing routines have similar trends in distinguishing LULC classes.

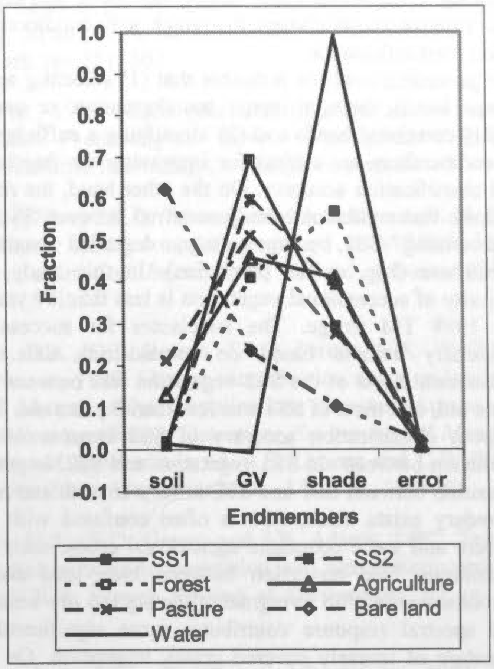


Figure 2. Fraction values of typical LULC classes in Rondônia

Table 1 summarizes the classification accuracies using different processing routines. Concerning atmospheric correction or not, the RAW-c produced 0.9% higher overall accuracy than Ref-c. This implies that atmospheric correction of TM images cannot improve classification or such correction is not required for LSMA application to LULC classification when image endmembers are used. The RAW-c method provides a slightly better accuracy for pasture and bare land, but the Ref-c method provides a little better accuracy for successional classes.

Table 1. Comparison of classification accuracy derived from different processing routines

LULC Classes	Different Image Processing Routines											
	Raw-c		Ref-c		Ref-uc		MNF-c		Subset-c		Subset-4c	
	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA
Forest	96.04	99.65	95.82	99.53	95.71	99.53	95.67	98.59	95.98	99.41	95.73	98.77
SS2	35.89	52.58	35.66	57.74	35.83	57.10	37.42	52.73	36.04	61.61	35.02	62.58
SS1	85.08	64.30	94.15	62.51	93.87	64.00	89.55	69.12	92.52	64.68	91.43	70.95
Pasture	78.41	88.68	79.47	83.53	80.52	79.38	86.48	86.37	81.17	89.08	84.96	89.02
Agriculture	85.03	76.03	77.23	80.17	77.09	82.31	68.24	78.84	82.78	77.85	87.71	84.96
Bare land	94.72	93.49	73.45	96.08	65.61	97.40	86.67	94.79	86.87	95.04	98.22	86.68
Water	100.00	80.91	100.00	74.55	100.00	71.82	97.75	79.09	100.00	73.64	100.00	92.73
OA	83.26		82.34		81.83		83.93		84.11		85.90	

Note: UA = user's accuracy, PA = producer's accuracy, and OA = overall accuracy in percent.

With a constrained or unconstrained least-square solution, the Ref-c and Ref-uc did not produce significantly different overall classification accuracy, although Ref-c has 0.5% higher accuracy than Ref-uc. The Ref-c provided a little better accuracy for bare land, but the accuracy for other classes between Ref-c and Ref-uc is very similar. Concerning image transformation or not, MNF-c produces 1.6% higher overall classification accuracy than that of Ref-c. This implies that image transformation is helpful in improving the separability of LULC classes. MNF transformation mainly improves the pasture and bare land classification accuracy. For image bands used in LSMA, Subset-c provides 1.8% better overall accuracy than Ref-c. This implies than removal of highly correlated image bands such as TM 1 and TM 2 in LSMA improved the quality of fraction images. The Subset-c method mainly improves pasture, SS2, and bare land accuracy. Concerning how many endmembers are used, the Subset-4c provides 1.8% better overall classification accuracy than Subset-c. The Subset-

4c method mainly improves SS1, pasture, and agriculture accuracy. This implies that adding one more soil endmember can improve those classes for which soil conditions directly affect their reflectance.

The preceding analysis indicates that (1) selecting appropriate image bands through image transformation or removal of highly correlated bands and (2) identifying a sufficient number of endmembers are critical for improving the fraction quality and classification accuracy. On the other hand, the results also indicate that confusion mainly occurred between SS1 and SS2 and among SS1, pasture (e.g., degraded pasture), and agriculture (e.g., coffee plantation). In this study area, the majority of successional vegetation is less than 10 years old on the 1998 TM image. The subclasses for succession were artificially defined based on ground-truth data and TM reflectance. Most of the SS2 vegetation was between 8 and 10 years old, and most of SS1 was less than 8 years old. This leads to poor classification accuracy of SS2 because of the high confusion between old SS1 vegetation and SS2. In practice, the transition between SS1 and SS2 is very smooth and no distinct boundary exists. Also, SS1 is often confused with degraded pasture and some economic agricultural crops, such as coffee plantations. The confusion between bare land and pasture becomes worse with overgrazing during the dry season, when soil spectral response contributes more significantly to the signature of sparsely covered grassy vegetation. On the other hand, degraded pasture in the process of vegetation recovery often has high densities of *Vismia sp.* and *Orbignya sp.*, increasing the confusion with the spectral response for SS1 or even perennial agriculture. Spectral responses for perennial agriculture can also be confused with SS1, mainly in areas of initial recovery of disturbed gallery vegetation.

In the moist tropical forest, vegetation stand structure and species composition are very complex. For an optical satellite sensor such as Landsat TM, the sensor mainly captures the leaves, wood, and shadowing information for a dense vegetation area. But for sparse vegetation, soil, and litter also affect the reflectance. In a large study area, soil conditions can be different, and the impacts of soil on reflectance can vary. Not all components selected are resolvable in a given image because of the particular mixture and their spectral contrasts. Different study purposes and different characteristics of selected study areas will require different endmembers. The selection of endmembers can be refined based on the analysis of the error fraction or unmodeled spectral variance. For TM images, selecting more than four endmembers is often difficult based on the image itself. Also, high correlation between TM bands limits the number of endmembers that can be used in LSMA.

CONCLUSIONS

This study indicated that LSMA is a promising approach for LULC classification in the Amazon. The soil fraction helps better distinguish successional and mature forests from pastures and agricultural lands. Green vegetation and shade fractions better distinguish mature forests from successional forests and pastures from agricultural lands. Based on the classification results, the following conclusions were achieved:

1. Constrained or unconstrained least-square unmixing solutions and atmospherically corrected or raw TM images do not produce significantly different overall classification accuracy. The overall accuracy was between 81.8% and 83.3%.
2. MNF transformation improved the overall accuracy, especially for pasture classes and bare land. An overall accuracy of 83.9% was achieved when using this technique, a 1.6% increase compared with the Ref-c method.
3. TM band subset images (bands TM 3, 4, 5, and 7) improved the accuracy for SS2, pasture, and bare land. An overall accuracy of 84.1% was achieved in this case.
4. Four endmembers (GV, shade, bright soil, and dark soil) with TM band subsets (TM 3, 4, 5, and 7) improved the accuracy for SS1, pasture, and agriculture. The overall accuracy reached 85.9% when using this approach.

In summary, selecting endmembers appropriately and reducing correlation between image bands used are two crucial aspects for developing high-quality fraction images when LSMA is used for LULC classification.

ACKNOWLEDGEMENTS

The authors wish to thank the National Science Foundation (grants 95-21918 and 99-06826), the National Aeronautics and Space Administration (grant N005-334), and Brazil's CAPES (Program for the Advancement of Education) for their support, which provided funds for the research that led to this paper. This project is part of the Large-Scale Biosphere-Atmosphere Experiment in Amazônia (LBA) program, LC-09, examining the human and physical dimensions of land-use and land-cover change. We also thank Indiana State University and Indiana University for facilities and support of our work and collaborators in Brazil, especially the LBA Program, EMBRAPA Satellite Monitoring, and the population of the study area, who made this work possible. We also appreciate the editing done by Joanna Broderick. None of the above funding organizations or individuals should be held responsible for the views presented in this paper. Sole responsibility for content lies with the authors.

REFERENCES

Adams, J.B., D.E. Sabol, V. Kapos, R.A. Filho, D.A. Roberts, M.O. Smith, and A.R. Gillespie, 1995. Classification of multispectral images based on fractions of endmembers: application to land cover change in the Brazilian Amazon. *Remote Sensing of Environment*, 52, pp. 137–154.

Adams, J.B., M.O. Smith, and A.R. Gillespie, 1993. Imaging spectroscopy: interpretation based on spectral mixture analysis. In: *Remote Geochemical Analysis, Topics in Remote Sensing 4*, ed. C.M. Pieters and P.A.J. Englert, Cambridge University Press, Cambridge, U.K., pp.145–166.

Adams, J.B., M.O. Smith, and P.E. Johnson, 1986. Spectral mixture modeling: a new analysis of rock and soil types at the Viking Landser 1 site. *Journal of Geophysical Research*, 91, pp. 8098–8112.

Aguiar, A.P.D., Y.E. Shimabukuro, and N.D.A. Mascarenhas, 1999. Use of synthetic bands derived from mixing models in the multispectral classification of remote sensing images. *International Journal of Remote Sensing*, 20, pp. 647–657.

Bateson, A., and B. Curtiss, 1996. A method for manual endmember selection and spectral unmixing. *Remote Sensing of Environment*, 55, pp. 229–243.

Batistella, M., 2001. *Landscape change and land-use/land-cover dynamics in Rondônia, Brazilian Amazon.* Ph.D. diss., Indiana University, Bloomington, Ind. USA, 399 pp.

Boardman, J.M., 1993. Automated spectral unmixing of AVIRIS data using convex geometry concepts. In: *Summaries of the Fourth JPL Airborne Geoscience Workshop,* JPL Publication 93-26, Jet Propulsion Laboratory, Calif., USA, pp.11–14.

Boardman, J.M., F.A. Kruse, and R.O. Green, 1995. Mapping target signature via partial unmixing of AVIRIS data. In: *Summaries of the Fifth JPL Airborne Earth Science Workshop,* JPL Publication 95-1, Jet Propulsion Laboratory, Calif., USA, pp.23–26.

Cochrane, M.A., and C.M. Souza Jr., 1998. Linear mixture model classification of burned forests in the eastern Amazon. *International Journal of Remote Sensing,* 19, pp. 3433–3440.

Congalton, R.G., 1991. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment,* 37, pp. 35–46.

Dale, V.H., and S.M. Pearson, 1997. Quantifying habitat fragmentation due to land-use change in Amazônia. In: *Tropical Forest Remnants: Ecology, Management, and Conservation of Fragmented Communities,* ed. W.F. Laurance and R.O. Bierregaard Jr., The University of Chicago Press, Chicago, Ill., USA, pp.400–409.

DeFries, R.S., M.C. Hansen, and J.R.G. Townshend, 2000. Global continuous fields of vegetation characteristics: a linear mixture model applied to multi-year 8km AVHRR data. *International Journal of Remote Sensing,* 21, pp. 1389–1414.

Elmore, A.J., J.F. Mustard, S.J. Manning, and D.B. Lobell, 2000. Quantifying vegetation change in semiarid environments: precision and accuracy of spectral mixture analysis and the normalized difference vegetation index. *Remote Sensing of Environment,* 73, pp. 87–102.

Ichoku, C., and A. Karnieli, 1996. A review of mixture modeling techniques for sub-pixel land cover estimation. *Remote Sensing Reviews,* 13, pp. 161–186.

Janssen, L.F.J., and F.J.M. Van der Wel, 1994. Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing,* 60, pp. 419–426.

Jensen, J. R., 1996. *Introduction Digital Image Processing: A Remote Sensing Perspective,* 2d ed. Prentice Hall, New York.

Lu, D.S., P. Mausel, E.S. Brondizio, and E. Moran, in press. Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research. *International Journal of Remote Sensing.*

Mausel, P., Y. Wu, Y. Li, E.F. Moran, and E.S. Brondizio, 1993. Spectral identification of succession stages following deforestation in the Amazon. *Geocarto International,* 8, pp. 61–72.

Miranda, E.E., and C. Mattos, 1993. *Machadinho d'Oeste: de colonos a munícipes na floresta tropical de Rondônia.* Ecoforça/Embrapa-NMA, Campinas, Brazil, 154 pp.

Mustard, J.F., and J.M. Sunshine, 1999. Spectral analysis for earth science: investigations using remote sensing data. In: *Remote Sensing for the Earth Sciences: Manual of Remote Sensing,* 3d ed., v. 3, ed. A.N. Rencz , John Wiley and Sons, New York, pp. 251–307.

Petrou, M., 1999. Mixed pixel classification: an overview. In: *Information Processing for Remote Sensing,* ed. C.H. Chen, World Scientific Publishing Co., Singapore, pp. 69–83.

Quarmby, N.A., J.R.G. Townshend, J.J. Settle, and K.H. White, 1992. Linear mixture modeling applied to AVHRR data for crop area estimation. *International Journal of Remote Sensing,* 13, pp. 415–425.

Roberts, D.A., G.T. Batista, J.L.G. Pereira, E.K. Waller, and B.W. Nelson, 1998. Change identification using multitemporal spectral mixture analysis: applications in eastern Amazonia. In: *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications,* ed. R.S. Lunetta and C.D. Elvidge, Ann Arbor Press, Ann Arbor Mich., USA, pp. 137–161.

Roberts, D.A., R.O. Green, and J.B. Adams, 1997. Temporal and spatial patterns in vegetation and atmospheric properties from AVIRIS. *Remote Sensing of Environment,* 62, pp. 223–240.

Roberts, D.A., M.O. Smith, and J.B. Adams, 1993. Green vegetation, non-photosynthetic vegetation, and soils in AVIRIS data. *Remote Sensing of Environment,* 44, pp. 255–269.

Rogan, J., J. Franklin, and D.A. Roberts, 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment,* 80, pp. 143–156.

Rondônia, 1998. *Diagnóstico sócio-econômico do Estado de Rondônia e assistência técnica para formulação da segunda aproximação do zoneamento sócio-econômico-ecológico – Climatologia,* v. 1. Governo de Rondônia/PLANAFLORO, Porto Velho, Brazil, 401 pp.

Schmink, M., and C.H. Wood, 1992. *Contested Frontiers in Amazônia,* Columbia University Press, New York, USA.

Settle, J.J., and N.A. Drake, 1993. Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing,* 14, pp. 1159–1177.

Smith, M.O., S.L. Ustin, J.B. Adams, and A.R. Gillespie, 1990. Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sensing of Environment,* 31, pp. 1–26.

Smits, P.C., S.G. Dellepiane, and R.A. Schowengerdt, 1999. Quality assessment of image classification algorithms for land-cover mapping: a review and a proposal for a cost-based approach. *International Journal of Remote Sensing,* 20, pp. 1461–1486.

Theseira, M.A., G. Thomas, and C.A.D. Sannier, 2002. An evaluation of spectral mixture modeling applied to a semi-arid environment. *International Journal of Remote Sensing,* 23, pp. 687–700.

Tompkins, S., J.F. Mustard, C.M. Pieters, and D.W. Forsyth, 1997. Optimization of endmembers for spectral mixture analysis. *Remote Sensing of Environment,* 59, pp. 472–489.

Ustin, S.L., Q.J. Hart, L. Duan, and G. Scheer, 1996. Vegetation mapping on hardwood rangelands in California. *International Journal of Remote Sensing*, 17, pp. 3015–3036.

Ustin, S.L., D.A. Roberts, and Q.J. Hart, 1998. Seasonal vegetation patterns in a California coastal savanna derived from Advanced Visible/Infrared Imaging Spectrometer (AVIRIS) data. In: *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, ed. R.S. Lunetta and C.D. Elvidge, Ann Arbor Press, Ann Arbor, Mich., USA, pp. 163–180.

Ustin, S.L., M.O. Smith, S. Jacquemoud, M. Verstraete, and Y. Govaerts, 1999. Geobotany: vegetation mapping for Earth sciences. In: *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, 3d ed., v. 3, ed. A.N. Rencz, John Wiley and Sons, New York, USA, pp. 189–233.

Ustin, S.L., and Q.F. Xiao, 2001. Mapping successional boreal forests in interior central Alaska. *International Journal of Remote Sensing*, 22, pp. 1779–1797.

Van der Meer, F., and S.M. de Jong, 2000. Improving the results of spectral unmixing of Landsat Thematic Mapper imagery by enhancing the orthogonality of end-members. *International Journal of Remote Sensing*, 21, pp. 2781–2797.

THE USE OF MULTI-BEAM RADARSAT-1 FOR TERRAIN MAPPING

E.C. Grunsky

Alberta Geological Survey, 4999 98th Ave., Edmonton, Alberta, CANADA T6B 2X3, Email: eric.grunsky@gov.ab.ca

KEY WORDS: RADARSAT-1 Standard Beam, S1, S7, Ascending, Descending, Principal Components Analysis, Multi-beam radar, terrain mapping, geomorphology

ABSTRACT

Multi-beam RADARSAT-1 satellite imagery has been used as part of a regional terrain mapping program in northern Alberta, Canada. Principal components analysis (PCA) has been applied to ascending and descending standard beam modes with incidence angles of 20-27° (S1) and 45-49° (S7). The resulting components yield imagery that highlights geomorphology, geologic structure, variation in vegetation and a measure of moisture balance in the study area. These features are highlighted due to the interaction of incidence angles and look directions within and between the two beam modes.

1. INTRODUCTION

The use of radar imagery has been commonly employed for mapping geomorphology and geologic structure (Lewis et. al, 1998; Singhroy et. al, 1993; Gupta, 1991, Singhroy and Saint-Jean, 1999, Smith et. al., 1999). The range of incidence angles, all-weather atmospheric penetration, and response to surface morphology give radar imagery significant advantages in measuring surface features relative to conventional fixed beam optical satellites. Radar imagery has been merged with optical and geophysical imagery (Harris et. al, 1994; Mustard, 1994) for the purposes of extracting information based on the synergies provided by the integration. Masuoka et. al (1988) applied the technique of principal components analysis on a radar image composite derived from Shuttle Imaging Radar (SIR-B) and Seasat. While most studies have focused on the use of radar for delineating geologic structure, it has also been used for mapping surficial geology (Graham and Grant, 1994).

RADARSAT-1 image characteristics are summarized by Luscombe et al., 1993. The RADARSAT-1 satellite operates at a single microwave frequency of 5.3 GHz (5.6 cm wavelength), generally known as C-band radar. The microwave transmission operates in H-H polarization mode. RADARSAT-1 Path Image (SGF) georeferenced images using a land based lookup table were provided by Radarsat International and the Canada Centre for Remote Sensing (CCRS). Each Standard Beam image is a composite of 4 looks (Raney, 1998, p. 73). This composite increases the signal to noise ratio at the expense of the spatial resolution. The imagery was provided at a nominal resolution of 12.5m (close to the single look spatial resolution) although the true spatial resolution of the averaged 4 look image is closer to 25m.

This study has focused on the use of principal components analysis (PCA) using RADARSAT-1 standard beam imagery as a tool for mapping surface features and geologic structures over large areas in Alberta, Canada (see Figure 1). Previous studies on the use of PCA with radar imagery (ERS-1 and CCRS C-SAR) found the technique useful for highlighting structural features in the Sudbury area of Ontario, Canada (Moon et. al, 1994 and Harris et. al, 1994). In the study by Moon et. al (1994), different incident angles, look direction, frequencies and polarizations were used to highlight geologic structure.

The detection of geologic structure is partially dependent on the look direction of the satellite (Harris, 1984; Lowman et. al, 1987). Studies have indicated (Harris, 1984) that linear features show up as distinctive lines in radar imagery when the feature is within 20 degrees of the perpendicular to the look direction of the radar sensor. In the case of using multi-beam imagery for RADARSAT-1 data, the identification of linear features is determined by a range of look directions that are different for each beam mode.

Singhroy and Saint-Jean (1999) have shown that variation in RADARSAT-1 incidence angles highlights ground features based on relief and surface texture. For this study, Standard Beam Modes S1 (20-27°) and S7 (45-49°) imagery was used to contrast the radar responses as a function of incidence angle and look direction (ascending-east looking /descending- west looking).

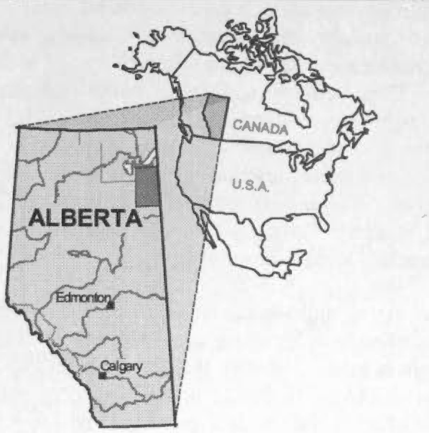


Figure 1.Location map of the study area filled by dark grey.

Figure 2 shows the configuration of beam modes and look direction for the image integration used for this study. Influences on the response are backscatter, which can be attributed to both volume and surface conditions, which in turn, are influenced by topography, vegetation and surface moisture (Raney, 1998). Lowman et. al (1987) describes the effect of incidence angle on the ability to detect differences in topography in which is expressed by shading controlled by local incidence angle variation. Incidence angles that are shallow (S1) are better suited for highlighting the differences in topography.

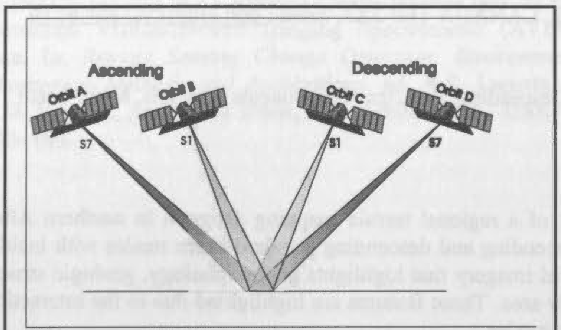


Figure 2. Multi-beam image assembly of RADARSAT-1. Each scene represents a distinct orbit pass. Effects of moisture content, and the state of the vegetation canopy can be different from one scene to the next due to temporal changes.

2. IMAGE ACQUISITION AND PROCESSING

The Alberta Geological Survey acquired RADARSAT-1 imagery for its regional mapping geologic program and for future applications in other areas of environmental monitoring and resource management (Grunsky, 2002). Two hundred and eighty Standard Beam modes S1 and S7 scenes were captured for both ascending and descending passes. Image acquisition was carried out over a ten week period from 01-Oct-1999 to 07-Dec-1999.

Autumn was chosen in order to minimize the effect of vegetation and to maximize microwave response from the ground surface. Although the effects of weather patterns are negligible on the radar response, factors that will affect the response include; the amount of surface moisture from significant rain events, and the loss of leaves in the deciduous foliage. These effects will cause some differences between scenes captured at different times. In this study, these effects have not been examined. However, it is unlikely that these effects would mask variation in terrain morphology or features associated with geologic structures. It is more likely that the loss of foliage will increase the contrast in features associated with structure and geomorphology.

For processing multi-beam radar imagery, the images must be orthorectified in order to minimize the effects of the shortening and layover associated with the difference in incidence angles. Orthorectification of the S1 and S7 imagery was carried out using digital elevation data provided by the Resource Data Division of Alberta Department of Sustainable Development. The data was provided in a grid form at 100 meter resolution.

A Gaussian smoothing filter was applied to the imagery to reduce speckle and highlight ground features. The filter implements a discrete 11x11 pixel (95% area under a normal distribution) Gaussian distribution applied over the image. The application of this filter improves the overall visual quality of the image and provides a closer approximation to a continuous tone surface, which permits identification of geologic structure, surface morphology, vegetation variation and cultural features.

Multi-channel remotely sensed imagery (multivariate data) contains responses that may reflect interactions between channels that are not visually obvious or are difficult to visualize from the individual source channels. A number of methods are suitable for contrasting differences in measurements over geographic areas. When two images are being compared, methods such as ratios or subtraction can be applied and can yield suitable contrasts. When three of more images are compared, contrasts become more complicated and multivariate techniques may be more appropriate. The method of PCA creates linear combinations of variables (channels) based on the covariance of the input channels and is commonly available with satellite image processing packages. The application of principal components is one method that can be used to discover interactions between the multi-channel data. Each successive principal component represents a linear combination of the input channels, which accounts for a portion of the overall variation of the data. The first principal component accounts for the most variation (by definition) followed by successively low order components which account for decreasing variation of the data. Each component can be viewed as a new variable (or channel) that describes relationships of the original variables in the form of linear combinations. Discussions on the use of PCA can be found in Richards (1986) and Gupta (1991).

3. STUDY AREA

In many parts of Canada the glacial surface deposits are characterized by till, sand and clay (Klassen, 1989). Glacial activity modified the surface leaving behind deposits of sand, till and clay along with glacial features such as drumlins, hummocks, outwash fans, and glacio-lacustrine deposits. These features have distinctive ground textures and have the potential to reflect differences in the radar return responses. In addition, the type of vegetation that grows over the glacial deposits is often characteristic to that type of terrain. As a result the variable backscatter associated with differences in vegetation can emphasize differences in the soil surface conditions.

The area chosen to apply the methodology of PCA on multi-beam radar imagery is located in northeast Alberta; National Topographic Series (NTS) Sheet 74L. The geology of the area is at the eastern edge of the Western Canada Sedimentary Basin where the sedimentary cover thins over the earlier Proterozoic Athabasca Basin and crystalline basement rocks. The area is covered with recent alluvial and organic deposits. Pleistocene age deposits include, aeolian sands, moraine, glacio-lacustrine and glacio-fluvial deposits. The surficial geology of the area was previously mapped by (Bayrock, 1971,1972).

4. RESULTS

The PCA methodology that was applied for the imagery used in this study generates eigenvalues and eigenvectors based on a correlation matrix generated from the four images (S1 Ascending/Descending and S7 Ascending/Descending). By definition, each successive component is orthogonal and in decreasing order. The results of the PCA on the correlation matrix are shown in Table 1. The table of R-loadings are the eigenvectors weighted by the contribution of the eigenvalues. Thus, the magnitude and sign of the R-loadings indicate the degree of significance and orientation of each variable (channel).

Table 1 shows that the first component contributes 57.2% of the overall variation in the imagery. The first principal component (not shown) is dominated by S7 Descending (S7D) , S7 Ascending (S7A) and S1 Ascending (S1A). S7A, S7D and S1A all contribute to the overall brightness of the first principal component. S1D makes a smaller contribution. Thus, the range of brightness value is contrasted between incidence angle differences. Look direction does not appear to be as significant. This could imply that the difference between responses between incidence angles is a function of the dielectric constant effect surface roughness, and vegetation differences (density and structure).

The second principal component (24.0% of the total variation) is dominated by S1 Descending imagery (S1D). The R-loading value of S1D shows that it has a magnitude of 0.899 and is at least twice the magnitude of the other channels. S1D is inversely associated with S7A and S7D for this component. This implies that scene brightness and variability is controlled by incidence angle and look direction. The R-loadings also indicate that the S7 images are opposite in sign to the S1 images. This inverse association implies that bright areas in S7 scenes are dark in S1 scenes, and vice-versa. In terms of the relative contributions, more than 80.8% of the variation of the S1D response is accounted for by the second principal component. A comparison of the image of this component (not shown) with the Alberta Land Cover Classification scheme (Alberta Dept. of Sustainable Development, 1999, shows that there is a good discrimination between forest canopy type (open/closed) and grasslands.

The third principal component (15.3% of the total variation) appears to be dominated by S1A. The magnitude of the R-loadings (Table 1) show that the S1D, S7A and S7D images are inversely associated with S1A. This supports the findings by Lowman et. al (1987) where steep incidence angles provide a better response for variation in topography. From the R-loadings shown in Table 1, S1D and S7A, S7D appear to have similar characteristics. The scene variability of the third principal component is controlled by the geomorphology. These relationships suggest that terrain features can be observed by contrasting look directions and is clearly shown in Figure 3. Fields of drumlins, sand dunes, eskers, embankments and other prominent topographic features can be clearly scene in this imagery.

The fourth principal component, which represents only 3.5% of the overall variation, shows little distinction between the channels. The imagery (not shown) shows differences in scene brightness and detailed features that include streams and small lakes. These features can be interpreted as being uncorrelated with other features in the image. The S7 Ascending and Descending channels are dominant over the S1 channels however these differences may be insignificant and might be artefacts of noise in the imagery. The variability of the fourth component is controlled by the contrast between look direction of S7A and S7D. S1A and S1D contribute little to the variability.

5. INTERPRETATION

The application of principal components is inherently dependent on the covariance (correlation) of the variables (imagery channels). Thus, the relative associations of scene response between different beam modes (incidence angles) and look direction, determines the correlations between the variables. Previous work by Singhroy and Saint-Jean (1999) has shown that Standard Beam Mode 1 is useful for identifying features in gently rolling or flat terrain whereas Standard Beam Mode 7 is useful for identifying features in areas of higher relief. The factors that affect these responses are; surface roughness, backscatter from the forest canopy and backscatter due to surface moisture.

Table 2. Eigenvalues, component contribution, and eigenvectors of the principal components analysis.

	λ1	λ2	λ3	λ4
Eigenvalues	2.287	0.958	0.613	0.141
% Contribution	57.190	23.960	15.330	3.530
Cumulative%	57.190	55.960	81.300	100.000

Eigenvectors	PC1	PC2	PC3	PC4
S1A	0.472	0.219	0.851	-0.068
S1D	0.285	0.871	-0.370	0.150
S7A	0.582	-0.382	-0.168	0.698
S7D	0.598	-0.216	-0.331	-0.697

R-Loadings	PC1	PC2	PC3	PC4
S1A	0.709	-0.209	-0.673	0.021
S1D	0.342	-0.899	0.270	-0.046
S7A	0.893	0.320	0.163	-0.272
S7D	0.907	0.188	0.264	0.268

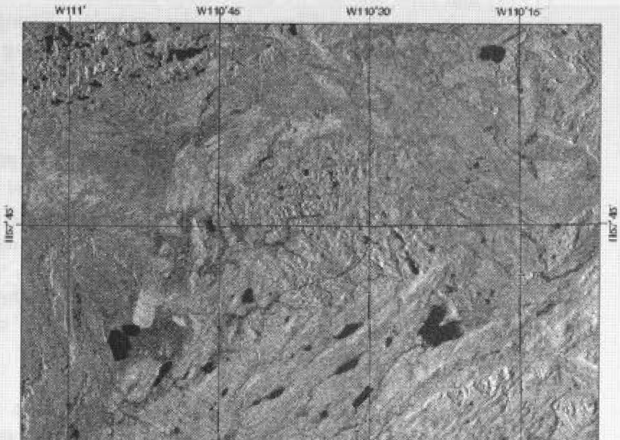


Figure 3. The third principal component of within NTS sheet 74E. Note the close match of glacial features between the geological map of Figure 4a.

Figure 4a shows a colour composite (RGB) image of the first three principal components. Features of geomorphology such as drumlins, ice crevasse fillings, sand dunes, etc. are highlighted through the contrast of the colours. Variation of vegetation shows up as varying amounts and combinations of red, green and blue. Areas of inferred increased moisture and forest canopy type have an overall darker intensity whereas areas that have little surface moisture display a brighter colour value. The colour composite clearly displays variation across the area that can be attributed to vegetation differences, surface roughness, surface moisture and topography. A study by Graham and Grant (1994) showed that X-band SAR imagery was useful for discriminating between different surficial deposits based on roughness and texture of the surface. The colour composite of the C-band RADARSAT-1 shows a range of colour and texture that outlines areas that match several of the surficial units mapped by Bayrock (1971) that is shown in Figure 4b. The areas of glacial outwash, ground moraine, aeolian, and ice contact deposits and fields of drumlins (noted as red ellipses) in Figure 4b stand out clearly in the images of Figures 3 and 4a.

6. CONCLUDING REMARKS

The application of principal components to multi-beam RADARSAT-1 imagery shows promise as a tool for identifying features associated with soil moisture, vegetation type, vegetation density and geomorphology.

As the principal components are based on the correlations of the input channels, the nature of the linear transformations of the original channels will be dependent on the variability of the terrain. In this study it appears that contrasts between S1 Ascending and S1 Descending imagery discriminates the features associated with the vegetation and geomorphology. This is consistent with the findings of Singhroy and St. Jean (1999) that state that the S1 beam mode is useful for evaluating features in gently rolling and flat terrain. Topographic enhancement was noted by Masuoka *et.* al (1988) using SIR-B and Seasat imagery. The topographic effect was the result of two opposite look directions of the Seasat images.

The features displayed by the imagery of the first principal component highlight the major differences between the S1 and S7 beam modes. From preliminary findings, it would appear that the contrast of these two beam modes highlight the difference between dry and wet surface conditions. Further investigations are required to test this interpretation. The use of a multi-beam radar image shows promise as a tool for surficial geological mapping as shown in Figure 3.

The methodology described here can be applied in various terrain types. The features extracted from the principal components will depend on the degree of relief in the area, the incident angles used (RADARSAT-1 beam modes) and look directions. From the work of Singhroy and St. Jean (1999) in areas of higher relief, S7 would likely play a more significant role in feature definition, while S1 would likely be less significant.

This methodology has been applied in the Buffalo Head Hills area of northern Alberta (NTS Sheet 84A/G) (Eccles et. al, 2002; Pagannelli et. al, 2001, 2002), where RADARSAT-1 S1 and S7 imagery were used to study the relationship of geologic structure with the presence of kimberlite intrusions.

The methodology of combining multi-beam RADARSAT-1 imagery using the method of principal components analysis shows promise as a tool for terrain mapping at a regional scale.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the assistance and support of the Department of Sustainable Development, Government of Alberta for providing the funding for the acquisition of the RADARSAT-1 imagery. Ken Dutchak and Bob Sleep provided the logistical and administrative assistance for the imagery capture. Bob Sleep also assisted in providing the author with a copy of the Alberta Land Classification scheme. Alberta Geological Survey colleagues, Kevin Parks and Laurence Andriashek, are thanked for their helpful discussions and comments on the interpretation of the imagery. The manuscript also benefited from reviews by Jeff Harris and Graeme Bonham-Carter.

The RADARSAT-1 imagery was supplied under a program administered by the Canadian Space Agency. RADARSAT-1 imagery was received by the Canada Centre for Remote Sensing. The imagery was processed and distributed by RADARSAT International

REFERENCES

Alberta Dept. of Sustainable Development, 1999. *Alberta Ground Cover Classification, NE Boreal Region*, Final Report, Dendron Resources Surveys Inc., February 1, 1999, 31p, CD-ROM

Bayrock, L.A., 1971. MAP 140 *Surficial geology of the Bitumount area*, Alberta, NTS 74E.. re-issued 2000

Bayrock, L.A., 1972. MAP 141 *Surficial geology of the Fort Chipewyan area*, Alberta, NTS 74L, Scale 1:250,000

Eccles, D.R., Grunsky, E.C., Grobe, M., and Weiss, J., 2001. Structural Emplacement Model for Kimberlitic Diatremes in Northern Alberta, Alberta Geological Survey, *Earth Sciences Report 2000-01*, 103pp.

Graham, D.F., and Grant, D.R., 1994. Airborne SAR for Surficial Geologic Mapping, *Canadian Journal of Remote Sensing*, Special Issue on Radar Geology, 20(3), pp.319-323.

Grunsky, E. C. in press. Regional Terrain Mapping in Northern Alberta using RADARSAT-1, Alberta Geological Survey *Earth Science Report.*

Gupta, R.P., 1991. *Remote Sensing Geology*, Springer-Verlag, New York, 356 p.

Harris, J., 1984. Lineament Mapping of Central Nova Scotia Using Landsat-MSS and SEASAT-SAR Imagery, *Proceedings of the 9th Canadian Symposium on Remote Sensing*, St. John's, Newfoundland, pp. 359-373.

Harris, J.R., Bowie, C., Rencz, A.N., and Graham, D., 1994. Computer Enhancement Techniques for the Integration of Remotely Sensed, Geophysical, and Thematic Data for the Geosciences, *Canadian Journal of Remote Sensing*, 20(3), p. 210-221.

Klassen, R.W., 1989. Quaternary Geology of the southern Canadian Interior Plains; in Chapter 2 of *Quaternary Geology of Canada and Greenland*, R.J. Fulton (ed.); Geological Survey of Canada, Geology of Canada, No. 1, pp.138-173.

Lewis, A.J., Henderson, F.M. and Holcomb, D.W. 1998. Radar Fundamentals: The Geoscience Perspective, in: *Manual of Remote Sensing : Principles and Applications of Imaging Radar*, 3rd edition Vol. 2, John Wiley & Sons, p 131-181.

Luscombe, A.P., Feguson, I., Shepherd, N., Zimcik, D.G., and Naraine, P., 1993. The RADARSAT Synthetic Aperture Radar Development, , Canadian Journal of Remote Sensing, Special Issue RADARSAT, 19(4), pp.298-310.

Lowman, P.D., Harris., J. Masuoka, P.M., Singhroy, V.H. and Slaney, V.R., 1987. Suttle Imaging Radar 9SIR-B) Investigations of the Canadian Shield: Initial Report, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-25, January, 1987, pp. 55-66.

Masuoka, P.M., Harris. J., Lowman, P.D. and Blodget, H.W., 1988. Digital Processing of Orbital Radar Data to Enhance Geologic Structure: Examples from the Canadian Shield, *Photogrammetric Engineering and Remote Sensing*, 54(5), May 1988, pp. 621-632.

Moon, W.M., Won, J.S., Singhryo, V., Lowman, P.D., 1994. ERS-1 and CCRS C-SAR Data Integration for Look-Direction Bias Correction using Wavelet Transform, *Canadian Journal of Remote Sensing*, Special Issue on Radar Geology, 20(3) ,pp. 280-285.

Mustard, J.F., 1994. Lithologic Mapping of Gabbro and Periodotite Sills in the Cape Smith Fold and Thrust Belt with Thematic Mapper and Airborne Radar Data, *Canadian Journal of Remote Sensing*, Special Issue on Radar Geology, 20(3)), pp.222-232.

Paganelli, F., Grunsky, E.C., and Richards, J.P., 2002. RADARSAT-1 Principal Component Imagery: a case study in the Buffalo Head Hills area, northern central Alberta, *Canadian Journal of Remote Sensing*, in press.

Paganelli, F., Grunsky, E.C., and Richards, J.P., 2001. Structural Interpretation of Radarsat and Landsat7 TM Images for Kimberlite Exploration in the Buffalo Head Hills Area, North-central Alberta,10th Annual Calgary Mining Forum - *Alberta Geological Survey Minerals Section Open House*, April 18-19, 2001, Calgary, Alberta.

Raney, K.R, 1998. Radar Fundamentals: Technical Perspective, in: *Manual of Remote Sensing: Principles and Applications of Imaging Radar*, 3rd edition Vol. 2, John Wiley & Sons, p 9-130.

Richards, J.A., 1986. *Remote Sensing Digital Image Analysis*, Springer-Verlag, New York, 281 p.

Singhroy, V., Slaney, R., Lowman, P, Harris, J. and Moon, W., 1993. Radarsat and Radars Geology in Canada, *Canadian Journal of Remote Sensing*, Special Issue RADARSAT, 19(4), pp. 338-351.

Singhroy, V., Saint-Jean, R., 1999. Effects of Relief on the Selection of RADARSAT-1 Incidence Angle for Geological Applications, *Canadian Journal of Remote Sensing*, (25)(3), pp. 211-217.

Smith S.K., Grieve, R.A.F., Harris, J.R. and Singhroy, V., 1999. The Utilization of RADARSAT-1 Imagery for the Characterization of Terrestrial Impact Landforms, *Canadian Journal or Remote Sensing*, 25(3), pp. 218-228.

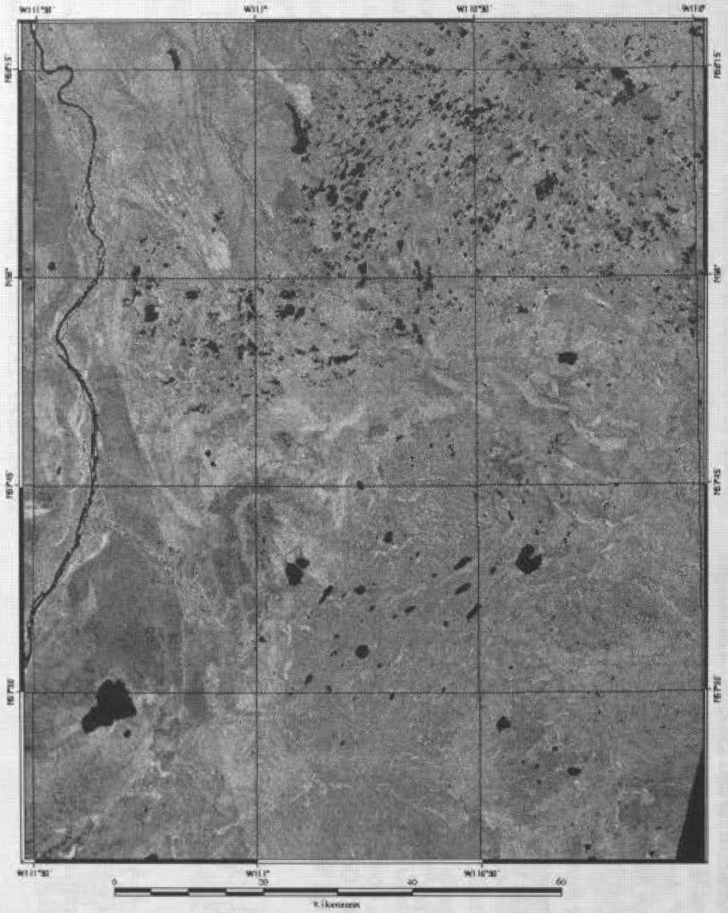

W111°30'
W111°
W110°30'
W110°
N58°15'
N58°
N57°45'
N57°30'
0
20
40
60
Kilometers

Figure 4a. A colour composite image of the first three principal components (Red – PC1, Green – PC2, Blue – PC3). The variability of colour and intensity displays features related to surface moisture, surface roughness, variation in vegetation and topography. Comparison with Figure 4a reveals features of the surficial geology in this image.match of glacial features between the geological map and the image. The image highlights features that were not noted in the map of the surficial geology.

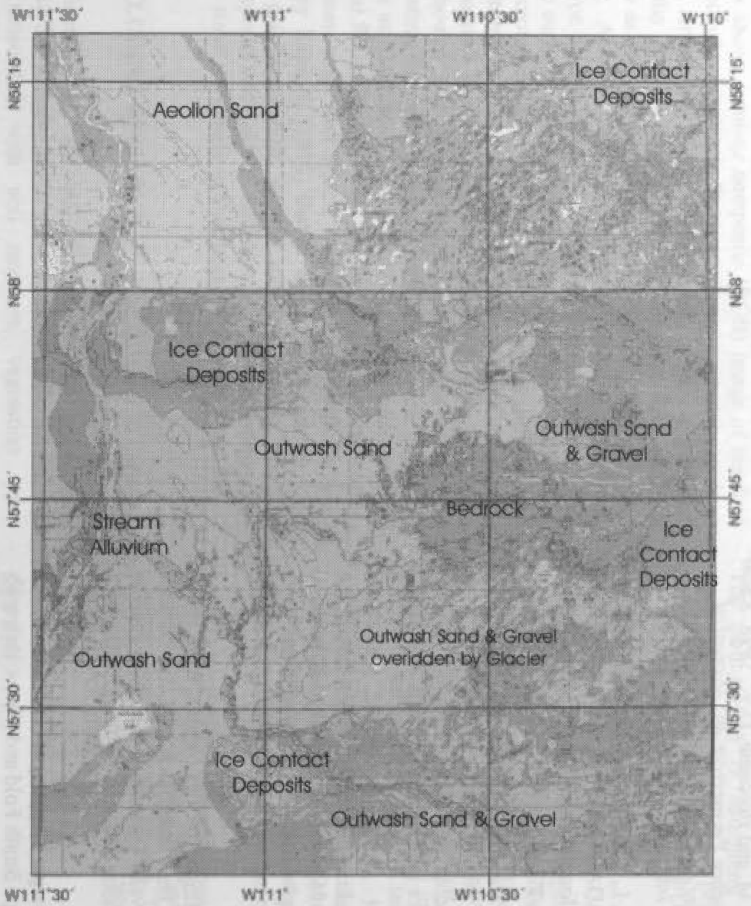

W111°30'
W111°
W110°30'
W110°
N58°15'
N58°
N57°45'
N57°30'
Aeolion Sand
Ice Contact Deposits
Ice Contact Deposits
Outwash Sand
Outwash Sand & Gravel
Bedrock
Stream Alluvium
Ice Contact Deposits
Outwash Sand & Gravel overidden by Glacier
Outwash Sand
Ice Contact Deposits
Outwash Sand & Gravel

Figure 4b. A portion of the surficial geological map by Bayrock (1971) showing glacial features and major surficial deposits. Drumlins are shown as red elliptical features.

AUTOMATIC EXTRACTION OF BUILDINGS FROM LIDAR DATA AND AERIAL IMAGES

F. Rottensteiner*, J. Jansa

Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Gußhausstraße 27-29,
A-1040 Vienna, Austria - (fr,jj)@tuwien.ac.at

Commission IV, WG IV/7

KEY WORDS: 3D building models, data acquisition, sensor integration, object-relational data bases

ABSTRACT:

This paper is focused on two topics: first, it deals with a technique for the automated generation of 3D building models from directly observed LIDAR point clouds and digital aerial images, and second, it describes an object-relational technique for handling hybrid topographic data in a topographic information system. Automatic building extraction combining the mentioned data sources consists of three steps. First, candidate regions for buildings have to be detected. After that, initial polyhedral building models have to be created in these candidate regions in a bottom-up procedure. Third, these initial polyhedral building models have to be verified in the images to improve the accuracy of their geometric parameters. This paper describes the current state of development, the overall work flow, and the algorithms used in its individual stages. Intermediate results are presented.

1. INTRODUCTION

1.1 Motivation and goals

The great demand for 3D building models for various applications collides with the enormous costs of acquisition of these data at an appropriate level of detail. Quite some efforts have been spent in the past to automatically extract buildings from digital aerial images (Lang, 1999) or from digital surface models (DSM) derived from laser scanner data (Weidner, 1997). The most recent achievements in the field of automated acquisition of 3D building models are based on the integration of data from two or more sources in order to overcome the drawbacks of specific sensor types. For instance, a DSM is combined with existing 2D GIS data (Brenner, 2000) or with digital aerial images (Ameri, 2000). The trend towards combining data from multiple sensors for the automatic reconstruction of topographic objects is triggered by the fact that new sensor types such as LIDAR, SAR, and high-resolution space borne scanners, have become available. In the future, the task of data acquisition for topographic information systems (TIS) might be performed by "multi-sensor-grammetry" rather than by traditional photogrammetry. It is the goal of this paper to give a contribution to the development of this "multi-sensor-grammetry" by presenting a new method for the automatic creation of polyhedral building models in densely built-up areas by combining high-resolution LIDAR DSM and aerial images. Along with that method, an object-relational technique for storing large sets of building models in a TIS is described.

Automatic building extraction by combining DSM and aerial images consists of three steps (Ameri, 2000). First, building candidate regions have to be detected in the DSM. Second, initial polyhedral building models have to be created in the candidate regions by a bottom-up procedure using both the DSM and the aerial images. Third, the initial polyhedral building models have to be verified in the aerial images to improve the accuracy of their geometric parameters.

Our technique follows this three-step procedure. It is still work in progress and, thus, not all of its modules have already been finished. This paper describes the current state of development, the overall work flow, and the individual modules. Intermediate results will be presented for the modules already having been implemented.

1.2 Related work

1.2.1 Building detection from DSM: For that purpose, the points of the DSM have to be classified according to whether they belong to the terrain, to buildings or to other object classes, e.g., vegetation. Morphological opening filters or rank filters are commonly used to determine a digital terrain model (DTM) which is subtracted from the DSM. By applying height thresholds to the normalized DSM (nDSM) thus created, an initial building mask is obtained (Weidner, 1997; Ameri, 2000). The initial classification has to be improved in order to remove vegetation areas. In (Brunn and Weidner, 1997), this is accomplished by a framework for combining various shape cues in a Bayesian network. Our algorithm for building detection from DSM is based on the method of linear prediction presented in (Kraus and Pfeifer, 1998).

1.2.2 Generic procedures for creating polyhedral building models: The data driven generation of a polyhedral model starts by finding initial hypotheses for planar patches in object space. This has been performed by first reconstructing 3D line segments from aerial images and then finding tilted "half planes" delimited by these line segments (Baillard et al., 1999) or by grouping co-planar line segments supported by the results of a color-based segmentation of aerial images (Moons et al., 1998). Other possibilities are given by propagating the results of gray level segmentation of an aerial image to object space using a DSM (Ameri, 2000) or by a curvature based segmentation of the DSM, e.g. (Brenner, 2000). The initial planar patches having been found, neighboring patches are grouped (Baillard et al., 1999, Ameri, 2000). After that, the polygons delineating the borders of planar patches have to be derived, which involves determining consistent intersections at the building vertices (Moons et al., 1998). Finally, the 3D polygons have to be combined in order to obtain consistent building models. At the building outlines, vertical walls have to be added to the model, and the floor has to be modeled. Parametric models which are typically used in model driven techniques for building extraction (Veldhuis, 1998; Rottensteiner, 2001) are used in a data driven context by (Brenner, 2000): the polygons delineating a building in a 2D map are split into rectangular regions. In each rectangle, the parameters of parametric models are determined using a DSM, and the model achieving the best fit is accepted.

1.2.3 Model verification by wire frame fitting: If an approximate object model exists, its (roof) edges can be back-projected to the images, where the model edges are matched with image edges. Thus, the accuracy of the model is increased considerably, especially with respect to the building outlines. Such model driven techniques are used for measurement of parametric primitives (Veldhuis, 1998; Rottensteiner, 2001). The method described in (Rottensteiner, 2001) is general enough to be applicable to any polyhedral model, and it is used in our work. Another technique for wire frame fitting of polyhedral models based on other observation equations is described in (Ameri, 2000).

1.2.4 Management of building data: In (Koehl and Grussenmeyer, 1998) and in (Grün and Wang, 1998), relational data base concepts for managing building models and DTMs are presented. By these concepts, the full power of standard relational data base systems can be exploited, especially with respect to semantic attributes. This advantage is contrasted by the fact that the data representing a single building are distributed over several tables. In order to perform an operation, an application has to construct the objects involved from these data. (Yang et al., 2000) use a relational data base system for managing building models, too, but in their system, the buildings are represented by binary large objects (BLOBs), and a tree-like structure is used for structuring the data by geometric criteria in order to increase access rates. The advantage of fast access is contrasted by the programming effort required to interpret the BLOBs and by the problems in exchanging data. The method described in our paper is also based on relational data bases and BLOBs. It is an expansion of the technique for the management of DTMs by (Hochstöger, 1996).

2. WORK FLOW FOR BUILDING EXTRACTION

The work flow for our method for automatic building extraction is presented in Figure 1. First, building candidate regions have to be detected in the DSM by a hierarchical classification procedure of the LIDAR points. As a result, regions of interest for the geometric reconstruction of the buildings are obtained. In the subsequent processes, these regions of interest are treated individually. Building detection is described in section 3.

In the regions of interest, initial planar patches are reconstructed in object space by the results of a segmentation of the DSM, classifying grid points as being "planar". This segmentation can be improved by a segmentation of the digital images giving regions of homogeneous gray level or color distribution. Neighboring patches being co-planar have to be merged, and a further analysis of neighborhood relations gives hypotheses for lines of intersection and/or step edges. As soon as no more hypotheses for grouping of planar patches can be found in the data, the resulting polyhedral model consisting of a conglomerate of intersecting (mostly: roof) planes is completed by adding vertical walls at the bordering edges of the planar patches not yet being neighbored by another plane and by adding a horizontal plane for modeling the floor. The creation of the initial polyhedral models is described in section 4.

The initial polyhedral building models have to be verified in the images by 3D-model fitting to improve the accuracy of their geometric parameters. The parameters having been determined, the models can be regularized by introducing hypotheses about planes being orthogonal to each other. The 3D-model fitting procedure is explained in section 5.

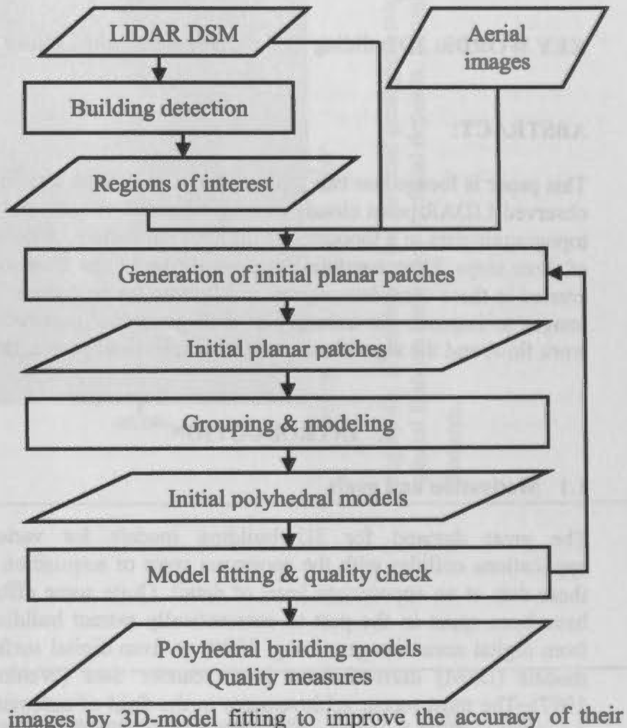

LIDAR DSM
Aerial images
Building detection
Regions of interest
Generation of initial planar patches
Initial planar patches
Grouping & modeling
Initial polyhedral models
Model fitting & quality check
Polyhedral building models
Quality measures

Figure 1. Work flow for building extraction.

The feedback loop shown in Figure 1 indicates a coarse-to-fine strategy by which initially a model is created from the most salient structures found in the data. This initial model then guides further processing as additional data become considered and need to be "explained" by the model.

In all phases of building extraction, a close integration of 3D object and 2D image spaces is needed to improve the degree of automation, the reliability and accuracy of the results. Multi-view optical images are needed to cope with occlusions, ambiguities, and to get the 3rd dimension. On-line self-checking has to be performed in all stages of the extraction process.

3. BUILDING DETECTION FROM LIDAR DSM

3.1 DTM generation from LIDAR data in urban regions

In (Kraus and Pfeifer, 1998), a method for linear prediction is proposed to separate vegetation from terrain points for the interpolation of DTMs from LIDAR data. The error distribution of the LIDAR heights is assumed to be skew with a strong bias towards off-terrain elevations. This assumption is used for robust estimation in order to eliminate these points. After that, a classification of the LIDAR points is performed by thresholding the residuals. For LIDAR DSMs having high point densities, the algorithm has to be modified to work in a hierarchical framework because in densely built-up areas, it has to cope with relatively large blocks of houses which share the statistical behavior of steep hills rather than that of single outliers. The algorithm has to be applied to thinned-out data first to get an initial classification of terrain points. After that, iteration is continued starting from pre-classified data. In (Briese, 2001), this strategy has been evaluated for the generation of a high-quality DTM of a test site in the City of Vienna.

3.2 Detection of buildings by comparing DSM and DTM

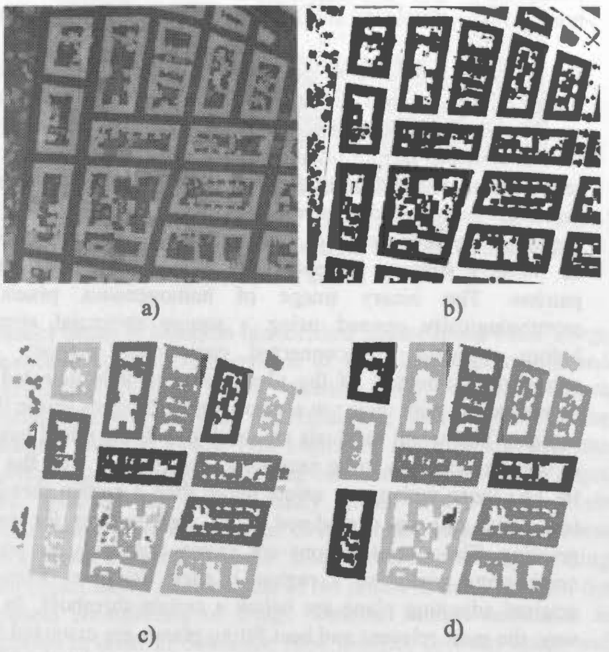
Two digital elevation models of identical grid width are derived by linear prediction: a DTM is computed from the points classified as "terrain points" with a high degree of smoothing, whereas a DSM is computed from all points without smoothing (Figure 2a). An initial building mask is created by thresholding the height differences between the DSM and the DTM. This initial building mask still contains areas covered by vegetation, and some individual building blocks are not correctly separated (Figure 2b). A morphological opening filter using a small (e.g., 5 x 5) square structural element is applied to the initial building mask in order to get rid of small elongated objects and to separate regions just bridged by a thin line of pixels. A connected component analysis of the resulting image is applied to obtain the initial building regions. At this instance, regions smaller than a minimum area (e.g., 40 m^2) and regions at the border of the DSM are discarded (Figure 2c).

Some of the remaining regions correspond to groups of trees. They can be eliminated by evaluating a "terrain roughness" criterion derived by an analysis of the second derivatives of the DSM. In (Fuchs, 1998), a method for polymorphic feature extraction is described which aims at a classification of texture as being homogeneous, linear, or point-like, by an analysis of the first derivatives of a digital image. This method is applied to the first derivatives of the DSM using a large (e.g., 9 x 9) integration kernel. For each initial building region, the number of "point-like" pixels is counted. Regions containing more than 50% of pixels classified as being "point-like" (thus, pixels being in a neighborhood of great, but anisotropic variations of the surface normals) are very likely to contain vegetation rather than buildings, and they are eliminated.

The terrain roughness criterion is very efficient in classifying isolated vegetation regions, but it cannot find vegetation areas which are still connected to buildings. In a final stage of analysis, we try to eliminate such areas. By morphological opening using a square structural element, regions just connected by small bridges are separated. The resulting binary image is analyzed by a connected component analysis which results in a greater number of regions, and the terrain roughness criterion is evaluated again. Pixels being in regions now classified as containing vegetation are erased in the initial building label image. Thus, in vegetation areas originally connected to buildings, only the border pixels remain classified as "building pixels". Again, morphological opening helps to erase these border pixels. The resulting building label image only contains a small percentage of erroneously classified pixels in a few backyards (Figure 2d).

At a very coarse level of detail, a 3D city model can be derived by creating prismatic models from the boundary polygons of the building regions using the average building heights computed from the DSM. An example for such a city model with a height accuracy of about ±5 m is shown in Figure 3.

Figure 2. Building detection in a test site in the City of Vienna. Original resolution: 0.1 m (in-flight) by 1 m (cross-flight).



a) DSM; grid width: 0.5 x 0.5 m^2; extent: 410 x 435 m^2. b) Initial building mask (height threshold Δh_{min}=3.5m). c) initial building label image before evaluating terrain roughness. d) Final building label image.

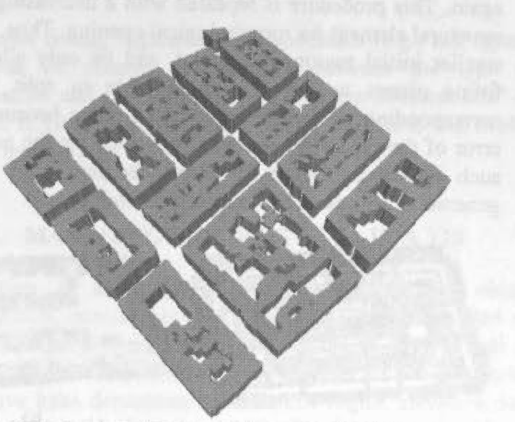


Figure 3. VRML visualization of prismatic models created from the boundary polygons of the building regions from Figure 2.

4. MODEL GENERATION

4.1 Generation of initial 3D planar segments

To start with model generation, initial 3D planar segments, their geometrical parameters, and their initial border polygons have to be found in the regions of interest. This can be achieved by generating a "segment label image" defined in object space with an appropriate grid width. Each pixel of that image is assigned the label of the planar segment it belongs to.

The framework for polymorphic feature extraction (Fuchs, 1998) is applied for the generation of planar segments, too. Just as described in section 3.2, the framework is applied to the first derivatives of the DSM, this time using a small integration kernel of 3 x 3 pixels. Pixels classified as being homogeneous are surrounded by pixels having similar components of the normal vector, i.e., they are in a region containing co-planar points (Brunn and Weidner, 1997). The binary image of the homogeneous pixels is used for further processing (Figure 4a).

By applying a connected component analysis to this binary image, planar patches should be detectable. However, due to errors in the classification of homogeneous pixels, especially at the intersections of roof planes which are almost horizontal, the regions thus detected often turn out to be too large. Typically, this leads to L-shaped segments such as the one at the upper left corner of Figure 4a. In order to avoid these segmentation errors, an iterative strategy is applied for the generation of planar patches. The binary image of homogeneous pixels is morphologically opened using a square structural element before applying the connected component analysis. The geometric parameters of the planar patches thus detected are derived along with their r.m.s. errors from all points inside these patches. The height residuals are used to split the initial patches in case this appears to be necessary. The patches with the best fit, i.e., those with r.m.s. errors better than a certain threshold (e.g., ±10 cm) are considered to be seed regions for region growing. These seed regions are grown iteratively by adding neighboring pixels to a region if their distances from the original adjusting plane are below a certain threshold. In this way, the most relevant and best fitting planes are extracted from the DSM. After that, the plane parameters are updated, and the pixels already being assigned to a planar patch are erased in the binary images. The connected component analysis is repeated, and the parameters of the new planar patches are evaluated again. This procedure is repeated with a decreasing size of the structural element for morphological opening. Thus, smaller and smaller initial regions are found, and by only allowing well-fitting planes to grow, it is possible to split the regions corresponding to more than one roof plane because the r.m.s. error of the planar fit is a good indicator for the occurrence of such situations. Figure 4b shows the results of seed region generation for one of the buildings from Figure 2.

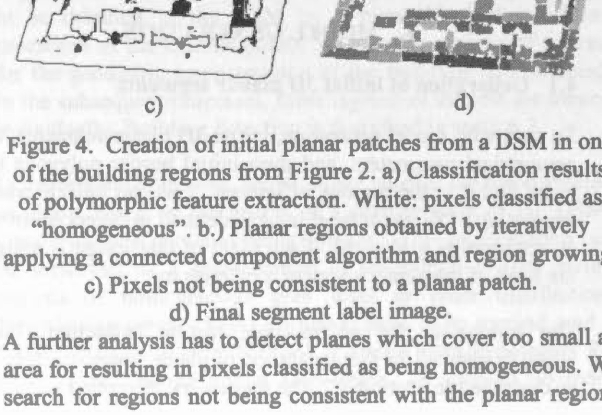


Figure 4. Creation of initial planar patches from a DSM in one of the building regions from Figure 2. a) Classification results of polymorphic feature extraction. White: pixels classified as "homogeneous". b.) Planar regions obtained by iteratively applying a connected component algorithm and region growing. c) Pixels not being consistent to a planar patch. d) Final segment label image.

A further analysis has to detect planes which cover too small an area for resulting in pixels classified as being homogeneous. We search for regions not being consistent with the planar regions detected so far (Figure 4c). The borders of the buildings are typically found in that process, which is caused by laser points on the walls. Again, we get rid of these points by a morphological opening operation using a 3 x 3 square structural element, and a connected component analysis is applied to the resulting image in order to create additional planar patches. Figure 4d shows the final segment label image created for one of the building regions from Figure 2. The r.m.s. errors of planar adjustment varies between ±5 cm and ±15 cm for the segments corresponding to the "homogeneous" points. The segments having a r.m.s. error larger than ±10 cm possibly still correspond to more than one roof plane. In the planar regions created by the analysis of the originally inconsistent points, the r.m.s. errors vary between ±25 cm and ±5 m, which is caused by the fact that some of these regions still correspond to trees, whereas other regions also correspond to more than one roof plane. In the future, a further analysis will split these planar patches into smaller ones corresponding to even smaller planes in object space. This can be accomplished, e.g., by a height segmentation of the DSM in these regions. Another possibility for doing so is the integration of the information from the digital aerial images into the process of finding the initial planar patches. The digital images are also segmented based on the algorithm for polymorphic feature extraction, which results in a set of edge and point features and a set of patches of homogeneous gray levels for each image. The image segments could be considered in region growing, for instance by applying additional conditions with respect to homogeneity of gray levels at projected positions. In addition, the geometrical resolution of the segment label image can be increased considerably because the resolution of digital aerial images is still superior to the resolution of a laser scanner DSM.

4.2 Grouping planar segments to create polyhedral models

To derive the neighborhood relations of the planar segments, a Voronoi diagram based on a distance transformation of the segment label image has to be created (Ameri, 2000): each pixel inside the region of interest not yet assigned to a planar segment is assigned to the nearest segment. The distances of pixels from the nearest segments are computed by using a 3-4 chamfer mask. Figure 5 shows a Voronoi diagram of the segment label image from Figure 4d. From the Voronoi diagram, the neighborhood relations of the planar segments are derived, and the borders of the Voronoi regions can be extracted as the first estimates for the border polygons of the planar segments.

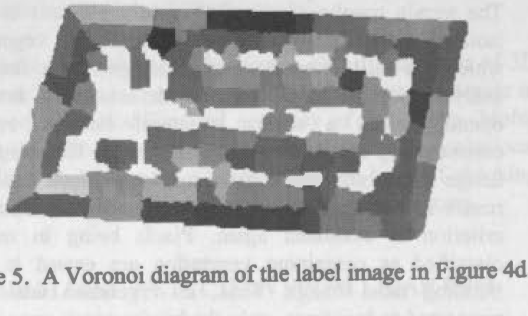


Figure 5. A Voronoi diagram of the label image in Figure 4d.

After deriving the neighborhood relations, neighboring planar segments have to be grouped. There are three possibilities for the relations of two neighboring planes (Baillard et al., 1999). First, they might be co-planar, which is found out by a statistical test applied to the plane parameters. In this case, they have to be merged. Second, two neighboring planes might intersect consistently, which is the case if the intersection line is close to the initial boundary. In this case, the intersection line has to be computed, and both region boundaries have to be updated to contain the intersection line. Third, if the planes do not intersect in a consistent way, there is a step edge, and a vertical wall has to be inserted at the border of these segments.

After grouping neighboring planes, the bounding polygons of all enhanced planar regions have to be completed. (Moons et al., 1998) give a method for doing so and for regularizing the shape of these polygons at building corners. Finally, the completed planar polygons have to be combined to form a polyhedral model. Polygon sides not being intersections or step edges are supposed to be situated at the building outlines. Vertical walls and a floor have to be added to the model.

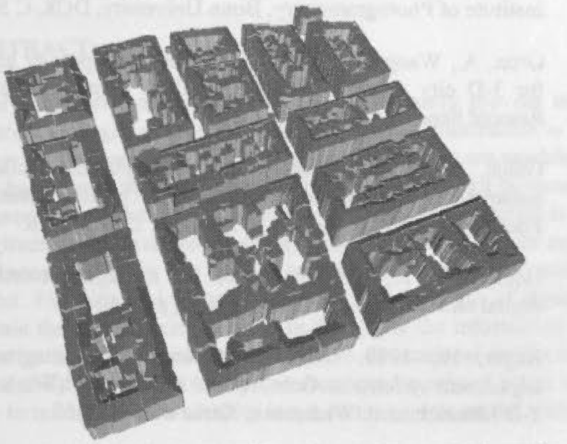


Figure 6. A VRML visualization of a model created from the boundary polygons of the Voronoi diagram in the test area from Figure 3.

The tools for grouping planes and for computing intersections and the positions of step edges have not yet been implemented. Figure 6 shows a VRML visualization of a 3D model created from intersecting vertical prisms bounded by the borders of the Voronoi regions with the respective 3D roof planes. The structure of the roofs is correctly resembled, but the intersection lines of neighboring roof planes are not yet computed correctly. However, the visualization shows the high potential of the method for generating roof planes from LIDAR data.

5. WIRE FRAME FITTING

The initial polyhedral model is back-projected to the images (Figure 7), where its edges can be matched with image edges. This is necessary in order to improve the accuracy of the building models, especially with respect to the building outlines. Our model fitting algorithm has first been used for automatic fine measurement of parametric and prismatic building models in a system for semi-automatic building extraction (Rottensteiner, 2001). It is based on the integration of object modeling by boundary representation (B-rep) and hybrid adjustment in the way it is realized in the adjustment package ORIENT (Kager, 1989). Each face of the B-rep corresponds to a set of "shape" ("GESTALT") observations, i.e., a set of points declared to be situated on the same surface (Kager, 1989), and the parameters of these surfaces have to be determined. In parameter estimation, the topology of the B-rep is represented by the fact that a building vertex gives support to at least three surfaces and by implicit geometric conditions imposed by a specific way of formulating the surface equations (Rottensteiner, 2001). Before matching is performed, the parameterization of the planes of the initial polyhedral models has to be decided upon, and all "shape observations" and plane parameters which represent the building model in the reconstruction process have to be prepared.

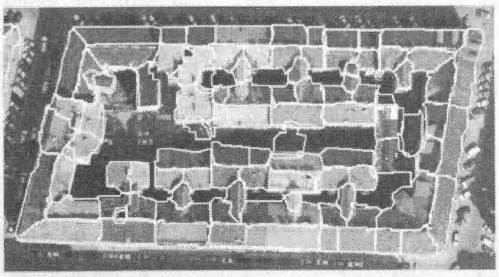


Figure 7. The roof polygons of the building in Figures 4 and 5 back-projected to one of the images (white). Approximate ground pixel size: 15 cm.

Model fitting is based on hierarchical feature based multi-image matching applying the scheme of hypotheses generation and verification (Gülch, 1994). Starting from the approximate values, the wire frame of the building is back-projected to the images. Straight image line segments are matched with the object edges. Whereas matching candidates for all object edges are searched for independently in all images, hypothesis verification is performed in an overall robust estimation process in a hybrid adjustment of the "shape observations" representing the object models, the heights of the points of the DSM, and the camera co-ordinates of image features. Robust estimation is applied in order to determine false matches between image and object edges. In a test project in the context of semi-automatic building extraction (scale 1:4500, six-fold overlap) the method has been shown to achieve results with an accuracy in the range of a few centimeters (Rottensteiner, 2001).

After model fitting, regularization hypotheses (e.g., orthogonal walls) can be created. A mapping between the type of regularization and the formulation of "shape observations" as described in (Rottensteiner, 2001) has to be found, and model fitting is repeated.

6. MANAGING BUILDING DATA IN A TIS

In addition to investigating the work flow for object reconstruction, considerable efforts have been spent to find an object-relational technique for the management of large sets of building data in a TIS. This method is based on the principles which have been demonstrated to handle digital elevation data of whole states (Hochstöger, 1996). The program TopDM, a development of our institute, is used as a TIS. It is based on a relational data base with additional topologic and geometric elements (Hochstöger, 1996). An object-relational principle (Kraus, 2000) is used for the management of building data in order to use the existing relational data base and to offer an object oriented view at the data base for application programs.

TopDM offers an area called "derived products market" (DPM), where DTMs are managed in a table of the relational data base. Each line of that table corresponds to a DTM. However, only the meta data of the DTM are actually contained in that table, whereas the actual DTM is stored on the disk. The meta data comprise attributes such as object type, data format, file name and extension in object space, the latter one being used for queries according to geometric criteria. The principle of managing the meta data in a relational data base while treating the actual data as BLOBs stored separately on disk was shown to be well-suited for managing digital elevation data of whole states in (Hochstöger, 1996). That principle could also be applied to building data by introducing new possible values for the attributes "object type" and "data format". The B-reps of the buildings are stored on disk in a specific binary data format.

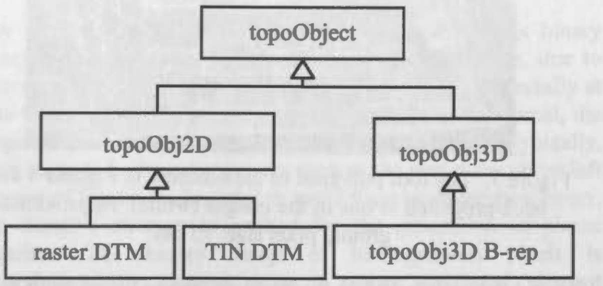


Figure 8. Inheritance tree for topographic objects.

An object oriented interface for the DPM was realized by creating an inheritance tree for topographic objects (Figure 8). The base class *topoObject* corresponds to a topographic object containing the meta data described above. The actual method of geometric modeling is not yet implemented in that class. In a first level, two classes corresponding to objects being modeled by 2.5D and 3D modeling techniques, respectively, are derived from class *topoObject*, but these classes do not yet correspond to real objects. The classes derived from them, however, correspond to actual data modeled by means of specific (2.5D or 3D) modeling techniques: DTMs in 2.5D grid-based representation or as TIN, and 3D objects in B-rep. In order to realize the object oriented view at the topographic data, an interface sending queries to the DPM had to be implemented. This interface has to interpret the attribute "object type" in the results of the query, instantiate an object of the class corresponding to the actual value of that attribute and use the class specific methods for reconstructing the object in memory from the data found on the file corresponding to the value of the attribute "file name". The results of the query are presented to an application as a list of instances of class *topoObject*, and the application can use the methods of that class, e.g., for visualizations such as the one presented in Figure 7.

7. CONCLUDING REMARKS

In this paper, we have presented a method for automatic extraction of buildings from multiple data sources. Parts of that method are still work in progress, but we were also able to present preliminary results achieved by the modules already implemented in a test project in the City of Vienna. These preliminary results show the high potential of the method.

REFERENCES

Ameri, B., 2000: Automatic Recognition and 3D Reconstruction of Buildings from Digital Imagery. PhD Thesis, Institute of Photogrammetry, Stuttgart University, DGK-C 526.

Baillard, C., Schmid, C., Zisserman, A., Fitzgibbon, A, 1999: Automatic line matching and 3D reconstruction of buildings from multiple views. IAPRS 32 (3-2W5), pp. 69-80.

Brenner, C., 2000: Towards fully automatic generation of city models. IAPRS 33 (B3), 85-92.

Briese, Ch., 2001: Digitale Modelle aus Laser-Scanner-Daten in städtischen Gebieten. Diploma Thesis, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology.

Brunn, A., Weidner, U., 1997: Extracting Buildings from Digital Surface Models. IAPRS 32 (3-4W2), pp. 27-34.

Fuchs, C., 1998: Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung. PhD Thesis, Institute of Photogrammetry, Bonn University, DGK-C 502.

Grün, A., Wang, X., 1998: CC-Modeler: a topology generator for 3-D city models. ISPRS Journal of Photogrammetry and Remote Sensing 53 (5), 286-295.

Gülch, E., 1994: Erzeugung digitaler Geländemodelle durch automatische Bildzuordnung. PhD Thesis, Institute of Photogrammetry, University of Stuttgart, DGK-C 418.

Hochstöger, F., 1996: Software for managing country-wide digital elevation data. IAPRS 31 (B2), pp. 160-163.

Kager, H., 1989: ORIENT: A universal photogrammetric adjustment system. In: Grün, A. and Kahmen, H. (Eds.): Optical 3-D Measurement. Wichmann, Karlsruhe, 447-455.

Koehl, M., Grussenmeyer, P., 1998: 3D data acquisition and modelling in a Topographic Information System, IAPRS 32(4), pp. 314-319.

Kraus, K., 2000: Photogrammetrie Band 3. Topographische Informationssysteme. 1st ed., Dümmler Verlag, Köln.

Kraus, K., Pfeifer, N., 1998: Determination of terrain models in wooded areas with aerial laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing 53 (4), pp. 193-203.

Lang, F., 1999: Geometrische und semantische Rekonstruktion von Gebäuden durch Ableitung von Gebäudeecken. PhD Thesis, Institute of Photogrammetry, Bonn University, Shaker, Aachen.

Moons, T., Frere, D., Vandekerckhove, J., Van Gool, L., 1998: Automatic modeling and 3D reconstruction of urban house roofs from high resolution aerial imagery. ECCV'98, Lecture notes in computer vision 1406(I), pp. 410-425.

Rottensteiner, F., 2001: Semi-automatic extraction of buildings based on hybrid adjustment using 3D surface models and management of building data in a TIS. PhD Thesis. Geowissenschaftliche Mitteilungen 56, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology.

Veldhuis, H., 1998: Performance analysis of two fitting algorithms for the measurement of parametrised objects. IAPRS 32 (3/1), pp. 400-408.

Weidner, U., 1997: Gebäudeerfassung aus digitalen Oberflächenmodellen. PhD Thesis, Institute of Photogrammetry, Bonn University, DGK-C 474.

Yang, B., Li, Q., Li, D., 2000: Building model creating and storing in 3D urban GIS. IAPRS 33 (B4), pp. 1192-1198.

ASSIGNING THE 3RD DIMENSION TO ROADS BY USE OF LASER SCANNING DATA

Carsten Hatger

Institute for Cartography and Geoinformatics, University of Hanover, Appelstraße 9a, 30167 Hanover, Germany - Carsten.Hatger@ikg.uni-hannover.de

Commission IV, WG IV/7

KEYWORDS: Feature Extraction, Geometric Modelling, High resolution DEM/DTM, Integration

ABSTRACT:

Most geographic data sets available today mainly provide two-dimensional geometric information about topographic objects in different aggregation levels. Three dimensional information is distributed separately by means of digital terrain models. In contrast, many applications presume that integrated data sets are available. The extraction of man-made objects from images and laser scanner data has been in the focus of image processing research in recent years. In this paper an approach is presented to identify road objects in laser scanner data sets. The information that is of interest is the outline of the road, its height and further tributary properties like longitudinal and transversal slope, width and curvature. In order to derive the width and further properties of roads, features and attributes of 2D data sets have to be evaluated against their contribution to the integration of height information into a common data model. Planning and construction facilities make use of standardised longitudinal and cross sections. Topographic data sets only contain the middle axis of roads. In this paper the information will be detected, analysed and measured by a model based approach, investigating the neighbourhood of two dimensional road elements in terms of cross sections. After the optimal position of the section is detected, the road parameters can be refined by an estimation procedure. It is subject to further investigations where and how to introduce cross sections into a common data model with respect to handling, storage and quality aspects.

1. INTRODUCTION

Today, most geographic data sets available mainly provide two-dimensional geometric information about topographic objects. Popular examples are the European geographic data File (GDF) data set used by intelligent transportation systems for vehicle navigation or the digital topographic data set of the German national mapping agencies ATKIS (Authoritative Topographic Cartographic Information System). The third dimension is typically stored separately in terms of a digital terrain model (DTM) describing landscape's surface by regular, irregular or hybrid grids.

Recently, higher sophisticated solutions like digital city models became available to the market. They not only describe the terrain surface as a DTM but also contain additional information about buildings and roof structures. Such options are needed by carriers of mobile phone service networks like GSM allowing for more precise network cell planning. A high quality data source for the extraction of such information are high resolution laser scanner data. Also in car navigations systems there is a growing awareness for the necessities of enriched data sets in terms of 3D object descriptions including the detailed shape of the features (eg. roads, buildings, bridges).

Many applications like those mentioned in the previous section before presume that integrated data sets are available. In this paper an approach is presented to identify road objects in laser scanner data sets. The aim is to extract on the one hand the detailed height information and derived measures like slope, and on the other hand the exact road geometry. The approach relies on the integration of prior information from GIS, and has to take the different quality levels of the integrated information into account. Examples for the extraction of the road geometry for two different road types will be given.

2. MOTIVATION

The integration of planimetric features with height information to is a demanding task, offering potential for the development of new techniques with respect to the third dimension. In comparison with conventional measurement techniques like tachymetry and photogrammetry, laser scanning has become an effective alternative for capturing digital elevation models (Ackermann, 1999; Baltsavias, 1999a). Because of the inherent measurement technique, laser scanner data based DTMs can provide dense height information. Although that makes it difficult for processing and demands new methods for data reduction, it offers new opportunities to disciplines being in need of high resolution data sets.

With respect to traffic systems, the third dimension can be applied to facilities related to transportation like roads. Such enriched topographic objects can help to improve the process of planning, construction, operation and use of roadway networks by several aspects. Applying height information to traffic related objects gives us the opportunity to record additional object attributes which are relevant for roads. Besides height, longitudinal and transversal slope form new attributes enabling for more precise description of the roadway network. Higher sophisticated attributes like the exposition or curvature of objects can be derived as well.

Applications could lie in the analysis of visibility ranges in the landscape. That information then to be used as hints or warnings by driver information systems for an increased safety. Changes in road curvature and gradient might be used to generate driver information for the readjustment of vehicles velocity. Automobiles carrying heavy or dangerous load can be automatically decelerated in front of hazardous gradients.

Due to the high density of the data, not only height information can be derived from the 3D-surface, but also the detailed shape of the objects. This allows to automatically extract the exact the road geometry, as long as it is bordered by typical height changes. In the context of car navigation, this allows for a lane-precise vehicle positioning and more precise route guidance.

Additional possibilities lie in the more precise prediction of emissions rates of harmful substances and noise depending on varying road gradients. Last but not least the exposition of lanes might be an indicator for glazed frost sections during cold seasons. This information can either be used as driver information or as an advice for gritters to spread salt on the affected roads.

3. PREVOUS WORK

The extraction of man-made objects from images and laser scanner data has been in the focus of image processing research in recent years (Grün, 1995). The interpretation of laser scanner data often relies on geometric properties that is analyzed and manipulated by use of filtering techniques.

Image based filter techniques closely related to mathematical grayscale morphology have been used by Vosselman (2000) as well as Kilian et al. (1996) and Lindenberger (1993). Results on road tracing by profile matching are given in Vosselman et al. (1995). Brügelmann (2000) and Wild et al. (1996) make use of slopes and even further derivatives. Procedures with respect to the data stochastics have been developed by Kraus & Pfeifer (1998) and Lohmann et al. (2000). Methods for separating man- made and natural objects from the terrain have been developed by Weidner (1997) and Pfeiffer et al. (2001)
Besides that several approaches make use of additional information in terms of GIS-information like Brenner & Haala (2000), Vosselman (2001) and Simonse (2000), or image data (Walter 1998, Schiewe 2000). Earlier efforts to extract buildings from laser data can be found in the researches of Haala et al. (1997a, 1997b).

4. SUGGESTED APPROACH

Existing geographic data sets provide detailed planimetric information on topographic features. The idea is to use both sources - 2D and 3D dataset - to describe roads more precise. The information that is of interest is the outline of the road, its height and further tributary properties like longitudinal and transversal slope, width and curvature. The information will be detected, analysed and measured by a model based approach, investigating the neighbourhood of two dimensional road elements in terms of cross sections.

4.1 Data Sources

4.1.1 3D Data Sources Three dimensional spatial data often is acquired by airborne laser scanners. The data is collected by sending out laser beams to the ground using pulsed or continuous wave techniques. Therefore the distance is derived by measuring travel time or phase shift of the emitted signal. To improve coverage of the sampled area system manufacturers use varying techniques to deflect the laser beam during flight time (Baltsavias 1999b). The result is called a Digital Surface Model (DSM) because it contains both points from the ground surface and points from objects on top of the surface, like buildings and trees. Several methods have been developed to calculate the difference between DSM and DTM. Many of these approaches use filtering techniques. Both DSM and DTM data sets are available by commercial companies.

The planimetric accuracy of the laser points is approximately 0.5 m (Balsavias, 1999c; Lohr, 1999) where the point density is up to 4 points per square meter. The accuracy in height is 0.01 up to 0.15 m (Briese et al 2001, Wever & Lindenberger, 1999).

4.1.2 2D-data sources. There are different data sets available that contain information about roads. First of all, there are cadastral and topographic data, but also data sets used for road operation and maintenance and car navigation.

The most accurate information is available at the road operation and maintenance divisions. They provide precise information about the geometry of road objects and also the road furniture like traffic signs, street markings, etc. However, this information is typically only available in analogue form.

Another potential data source is the cadastral data set in Germany (ALK), that also contains road objects with high geometric accuracy (approximately 0.05 up to 1 m). Unfortunately, only the legal boundaries of the roads are given, not the real borders. Furthermore, this data set is not yet available in digital form for the whole country.

Two data sets that are available in digital form for the whole country are the digital topographic data set ATKIS Digital Landscape Model (ATKIS-DLM) (scale approximately 1:25.000) and Geographic Data File (GDF) with a similar resolution. In both data sets, roads are modelled as linear features, typically by their centre line. The accuracy of GDF and ATKIS is similar and lies in the order of 3 to10 m..

4.2 Analysing Feature Attributes

In order to derive the width and other properties of road features, the attributes of 2D data sets have to be evaluated against their contribution to the integration of height information into a common data model. As described above, in the data sets available nationwide, road objects are modelled as linear features. Both ATKIS-DLM and GDF provide different road categories on which conclusions related to the current feature extent can be drawn. Some of them provide this information even in an explicit way by storing its width as an attribute value. Those features and attributes have to be identified, which are suitable for storing the newly derived information like slope and curvature.

ATKIS-DLM, for instance, provides the following attribute information being suitable for describing and deriving the features spatial extent:

- Width of traffic route, which represents the features total width including side strips, classified in 3 m intervals,
- width of carriage way, referring to the features width including lanes and side strips, measured in decimeters (accuracy approx. 0.5 m), which depends on the feature type,
- number of lanes, denoting the number of strips per direction and
- category of traffic route, distinguishing the features administrative, maintenance and/or operation responsibility or construction principles according to design standards.

Three dimensional information is captured and distributed separately from the ATKIS-DLM by means of a digital terrain model. By its nature, the ATKIS Digital Terrain Model (ATKIS-DTM) usually describes the landscapes surface using regular grids. The resolution is about 12.5 meters including additional morphological information like breaklines. The vertical accuracy is about 0.5 m.

Geographic Data File provide explicit information on roads including three dimensional properties, too, in detail (ISO, 1996):

- Functional Road Class, which denotes a classification based on the importance of the role that the road plays in the connectivity of the total road network and
- the maximum number of lanes, describing the total number of lanes associated with one particular driving direction.
- The width of a road,
- the road gradient along the features axis,
- the road inclination perpendicular to the feature axis and
- the structures for storing three dimensional coordinates.

Although the data model itself provides several facilities for storing and retrieving three dimensional feature attributes, nowadays commercial products available to the European market do not provide such information (Marwitz, 2001). In case of feature width is not available, it has to be derived from more general information describing the objects function, then assuming a certain width depending on the objects functional road class.

4.3 Creating Cross Sections

One essential property of any operational road is its continuous surface along and perpendicular to its axis (Vosselman, 1995). Besides that, vehicles require this surface to be typically plane up to a certain degree. Additionally the surface probably will be curved or at least inclined to get the water running off in case of rainfalls. Planning and construction facilities make use of standardised longitudinal and cross sections for the encoding of these properties. The approach in this paper uses a very similar technique for the description of the three dimensional road geometry because of two reasons. Firstly, sections are an efficient tool to describe the road geometry because of their two dimensional nature. Therefore, their properties can be described by use of existing geometric primitives supported by common GIS-tools that usually are capable of 2D analysis. Secondly, cross sections play a significant role during the merging process of two and three dimensional data sets, as will be shown later on. The length of the sections is obtained from the length of road elements and by feature width, respectively.

4.4 Merging 2D with 3D features

As 2D and 3D data are recorded separately with different measurement techniques and also different accuracies the objects typically do not match. A mere overlay of 2D and 3D data set will result in discrepancies that have to be solved by use of buffering and filtering techniques.

One common approach of merging height information and two dimensional data sets is by simply draping the planimetric geometry onto the DTM. The height at any position of features is assigned by use of suitable interpolation methods. The most naïve solution is given by linear interpolation between existing vertices. Higher sophisticated methods are based on surface approximation by use of spline functions (Fritsch, 1990).

In this approach height information has been derived by using the interpolation method of a common GIS-tool (here ArcGIS from ESRI) assigning heights to vertices depending on the DEMs resolution.

4.5 Cross Section Model

We are starting with an overlay of the 2D-geometry of ATKIS and the 3D surface data. At discrete distances, cross sections are established. The width of the cross sections is derived from the road class attribute, as well as from the width attribute given in the data set.

In addition to this, the accuracy of the centre line of the road is assumed to be 3-10 m, the positional accuracy of the laser scanner data is given by 0.5-1 m. This leads to a initial length of the cross section of up to 50 m.

The distance between successive cross sections depends on the expectance of how exact the shape of the road can be extracted: in dense city areas, there will be "obstacles" at the border of the road like parking cars or trees, which typically will not be the case on highways. In both cases, however, it can be expected that cars are on the road, that lead to peaks in the interpolated 3D cross section.

Often, slopes or ditches are located parallel to linear features like traffic routes. They form discontinuities of the landscape surface. This assumption can help to recognize the roads axis position by comparison with typical cross sections as given in figure 1. Based on this model template we are able to create more detailed road model and to perform exact merging of roads and DEMs.

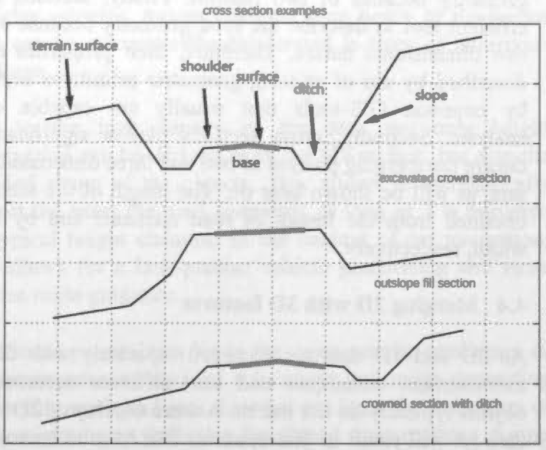


Figure 1. Cross section examples

A model of the expected cross section will be set up, that depends on the prior information. A typical highway cross section is given in Figure 2a, whereas a cross section in a city will look like in Figure 2b. The two peaks that frame the road section can be interpreted as vegetation, which is growing on the road shoulder (The vertical green line shows the position of the road axis and the blue dotted red line represents the surface heights. Note the different scale for the two axes.)

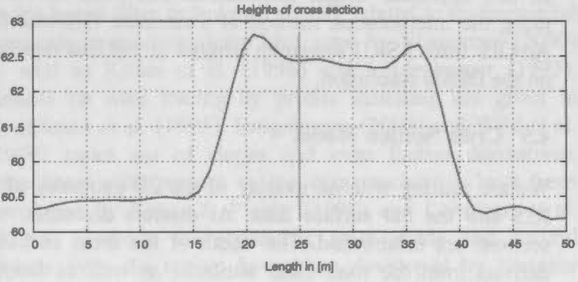


Figure 2a. Typical cross section of a highway

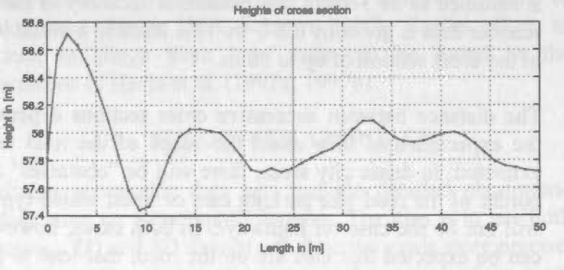


Figure 2b. Typical Cross section in a city

Surface discontinuities can be interpreted as local minima or maxima of surface curvature. Besides that, curvature values usually will reach their maximum absolute value perpendicular to the linear features axis. Relying on that effect several basic types of geometries can be identified for use in sections (cf. fig. 1). They indicate surface discontinuities along transportation related linear features like roads. Within three dimensional data sets the appearance of this geometric shape may differ slightly as it can appear distorted and mirrored.

In the most simple case the morphology of a road cross section can be described as a straight line. Using the morphology of cross sections as a model it becomes feasible to achieve exact matching between two dimensional features like roads and three dimensional data sets. The assumption of a planar surface – which is sensible in order to make sure that vehicles can pass over – leads to rather constant differential height changes along the section.

These differential height changes can also be substituted by computing the values of the first derivative describing the slope along the section. Normally these values will vary within the sections length around a certain value characterising the transversal road inclination. A graphical example is given in figure 4.

Local maxima and minima of surface curvature are also indicators for the existence of a possible topographical breakline. Therefore the analysis of several successively arranged cross sections along the feature can also be used for breakline detection related to slopes or ditches.

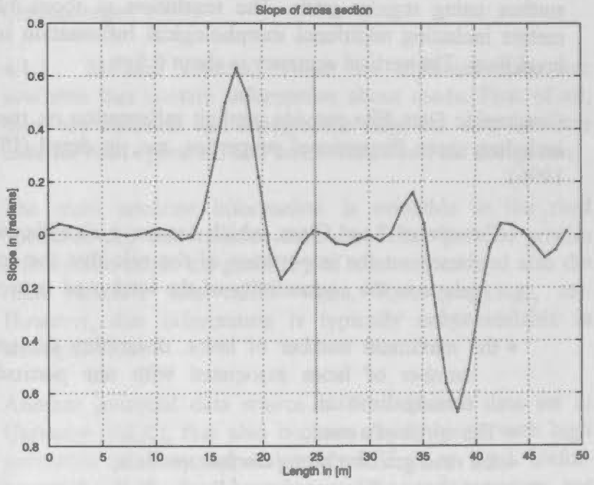


Figure 4. Slope of cross section of a major road

By computing the second derivative of the given heights one gets curvature along the cross section. As already mentioned before local minima and maxima of curvature indicate surface discontinuities while the value itself helps to figure out if it is a concave or convex curvature from above. The aim is now to search for these extreme values within the computed curvature (cf. fig 5). An exact road location and measurement is possible by matching the extracted profile with one of the given road section models.

Given approximate values for the road width, as well as the assumption of a constant inclination and a low curvature of road surfaces along their axes gives us the opportunity to estimate an optimal position for the centre of the road, its axis.

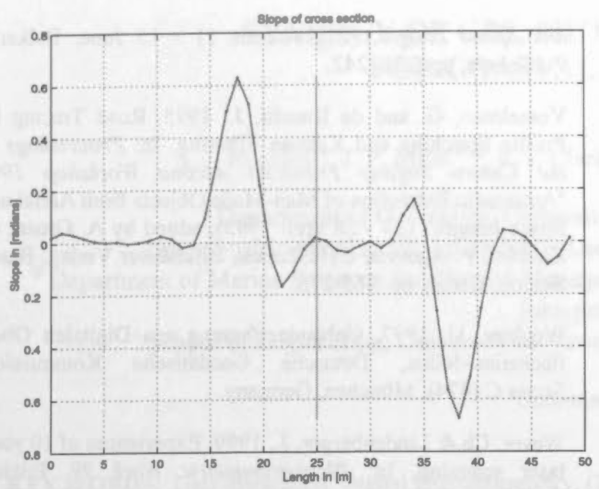


Figure 5: Curvature of a major road

After the extraction of zero crossings of curvature – indicating zero curvature – their position along the axis of the cross section has to be analysed and intervals bigger than the corresponding feature width and accuracy have to be built. Inlying local extreme values of these intervals denote the beginning of the road surface. If any one of the road section models is applicable, the distance between the extreme values should be in the range of the features width. Even if there is only one extreme value or zero crossing to be found, its neighbourhood along the section should distinguish itself by reasonable low inclination and curvature values in the range of the features width. The search ranges applied in this matching procedure take the accuracy of the original data into account. Only if there is no significant extreme value or zero crossing available, the original position of the features axis has to be maintained and the model cannot be applied. In this case the road feature cannot be improved with the given 3D data.

4.6 Estimation of Optimal Cross Section and Assignment of Attributes

After the optimal position of the section is detected, the road parameters can be refined by an estimation procedure. On the one hand, the exact width of the road can be determined. Furthermore, the shape of the section can be refined by estimating an optimally fitting curve – either a straight line, or a parabola opened downwards. Also, optimal position and heights for 2D linear features can be obtained by interpolation at any position between given heights. In this way sections can be built in the direction of and perpendicular to the features axis.

Within the ATKIS-DLM data set changes in road feature width only can occur after 500m as predefined by capturing rules. Therefore, it makes sense to refine this attribute, too. Comparing the distance between the zero crossings of slope to the initial feature width, updated values can be obtained. The differences between estimation and original values are in the range of two up to three meters. Assuming only smooth changes in the geometry of the road the estimated values of consecutive sections is expected to be similar. Thus the integration of results from neighbouring sections will lead to a more robust and reliable result.

4.7 Data Model

As a result of this integration of 2D and 3D-data, it is expected that the accuracy of the 2D-planimetry will be enhanced. Furthermore, additional information will be available, that was – up to now – not available or foreseen in the given data models. An open question therefore is how the refined, more accurate, road information on the one hand, and the enriched data on the other hand can be integrated into the given 2D-data sets.

None of the investigated data models, neither ATKIS nor Geographic Data File are able to completely integrate the derived information. Besides that only few attributes seem to be suitable for partly integration of height information to makes users benefit from the results.

It is subject to further investigations where and how to introduce cross sections into a common data model with respect to handling, storage and quality aspects. One simple approach is to store evenly spaced cross sections and one longitudinal section as poly lines together with the features identity into a common data model. Another possibility is given by storing the information using segmented attributes (Marwitz, 2001; ISO, 1996).

5. FUTURE WORK

It is obvious that by varying the spacing between the sections similar results for the estimation of section models will be achieved. Further investigations will be undertaken in the analysis of this effect for aims of data reduction with respect to quality and topographic morphology.

The concepts presented in this paper will be implemented and thoroughly tested with respect to the reliability and accuracy of the derived information. In addition to this, the results will be classified about their function within the roadway network.

REFERENCES

Ackermann, F., (1999). Airborne Laser scanning – present status and future expectations. In: *ISPRS Journal of Photogrammetry and Remote Sensing*, (54), pp. 64-67.

AdV, 2002. ATKIS – Objektartenkatalog. http://www.atkis.de:8080/dstinfo/plsql/dstinfo2.dst_gliederung2?dst_ver=dst (accessed 27. Feb. 2002).

Baltsavias, E. P., 1999a. A comparision between photogrammetry and laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, (54), pp.83-94.

Baltsavias, E. P., 1999b. Airborne Laser Scanning: existing systems and firms and other resources. *ISPRS Journal of Photogrammetry and Remote Sensing*, (54), pp. 164-198.

Baltsavias, E. P., 1999c. Airborne Laser Scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, (54), pp. 199-214.

Brenner, C. & Haala, N., 2000. Erfassung von 3D Stadtmodellen. In: *PFG - Photogrammetrie, Fernerkundung, Geoinformation*, Issue 2/2000, pp.109 – 118.

Briese, Ch., K. Kraus, G. Mandlburger and N. Pfeifer, 2001. Einsatzmöglichkeiten der flugzeuggetragenen Laser-Scanner. In: *Tagungsband der 11. Internationalen Geodätischen Woche in Obergurgl*, Innsbruck, Austria, pp. 17-26.

Brügelmann, R., 2000. Automatic breakline detection from airborne laser scanner data. In: *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXII, Part B3. Amsterdam, pp. 109-116.

Fritsch, D., 1990. Raumbezogene Informationssysteme und digitale Geländemodelle. Technische Universität München,

Haala, N., Brenner, C., 1997a.: Interpretation of Urban Surface Models using 2D building information. In: *`Automatic Extraction of Man-Made Objects from Aerial and Space Images'*, A. Gruen, O. Kuebler M. Baltsavias, (eds.), Birkhäuser Verlag, Basel, Boston, Berlin.

Haala, N. & Anders, K.-H., 1997b. Acquisition of 3D urban models by analysis of aerial images, digital surface models and existing 2D building information. In: *SPIE Conference on Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision*, (III), Orlando, Florida, pp. 212-222.

ISO, 1996. Geographic Data File ENV ISO 14825:1996

Kilian, J., Haala, N. and Englich, M., 1996. Capture and elevation of laser scanner data. In: *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXI, Part B3. Vienna, pp. 109-116.

Lohmann, P., Koch, A. and Schaeffer, M., 2000. Approaches to the filtering of laser scanner data. In: *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXII, Part B3. Amsterdam pp.540-547.

Lindenberger, J., 1993. Laserprofilmessung zur topographischen Geländeaufnahme, Deutsche Geodätische Kommission, Series C (400), München, Germany.

Lohr, U., 1999. High Resolution Laserscanning, not only for 3D-City Models, In: *Photogrammetric Week 99*, Fritsch/Spiller (eds.), Wichmann, Stuttgart Germany, pp. 133-138.

Marwitz, S., 2001. Integration von Höheninformation in eine vorhandene Datenbank für die Kfz-Navigation, (unpublished), Berlin, Germany.

Pfeifer, N., Stadler, P. and Briese, Ch., 2001. Derivation of digital terrain models in the SCOP++ environment. In: *OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for digital elevation models*, Stockholm, Sweden.

Schiewe, J., 2000. Combining geometrical and semantical image information for the improvement of Digital Terrain Models. In: *A Decade of Trans-European Remote Sensing cooperation*, Buchroithner, M.F. (Ed.) Proceedings of the 20th EARSEL-Symposium, A. A. Balkema Publishers, pp. 175-180.

Vosselman, G. and Suveg, I., 2001. Map Based Building Reconstruction from Laser Data and Images. In: *Proceedings of Automatic Extraction of Man- Made Objects from Aerial and Space Images*, (III), Ascona 11 - 15 June, Balkema Publishers, pp. 231- 242.

Vosselman, G. and de Knecht, J., 1995. Road Tracing by Profile Matching and Kalman Filtering. In: *Proceedings of the Centro Stefano Franscini Ascona Workshop 1995* "Automatic Extraction of Man-Made Objects from Aerial and Space Images" (24 - 28 april 1995), edited by A. Gruen, O. Kuebler, P. Agouris, ETH Zurich, Birkhäuser Verlag, Basel, Boston, Berlin. pp. 265-275.

Weidner, U. 1997. Gebäudeerfassung aus Digitalen Oberflächenmodellen, Deutsche Geodätische Kommission, Series C (474), München, Germany.

Wever, Ch.& Lindenberger, J., 1999, Experiences of 10 years laser scanning, In: *Photogrammetric Week 99*, Fritsch/ Spiller (eds.), Wichmann, Stuttgart, Germany, pp. 125-132.

Wild, D. & Krzyustek, P., 1996. Automatic breakline detection from airborne laser scanner data. In: *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXI, Part B3. Vienna, pp. 946-952.

ACKNOWLEDGEMENTS

The laser scanner data have been provided by courtesy of TopScan GmbH, Steinfurt, Germany and Emschergenossenschaft, Essen, Germany. ATKIS-DLM: © Geobasisdaten: Landesvermessungsamt NRW, Bonn, Germany, 1116/2002.

SYNERGY OF GPS, PHOTOGRAMMETRY AND INSAR FOR COASTAL ZONE MONITORING

S. J. Buckley [a], J. P. Mills [a] *, P. J. Clarke [a], S. J. Edwards [a], J. Pethick [b], H. L. Mitchell [c]

[a] Department of Geomatics, University of Newcastle upon Tyne, United Kingdom –
simon.buckley@ncl.ac.uk, j.p.mills@ncl.ac.uk, peter.clarke@ncl.ac.uk, s.j.edwards@ncl.ac.uk
[b] Department of Marine Sciences and Coastal Management, University of Newcastle upon Tyne, United Kingdom –
john.pethick@ncl.ac.uk
[c] School of Engineering, University of Newcastle, Australia – harvey.mitchell@newcastle.edu.au

Commission IV, WG IV/7

KEY WORDS: Data Integration, Digital Photogrammetry, GPS, InSAR, Monitoring

ABSTRACT:

This paper reports on research that aims to develop a technique for rapidly monitoring coastal erosion over wide areas, by deriving synergy from three integrated measurement technologies. These technologies are the global positioning system (GPS), automated digital photogrammetry using imagery acquired with a small format digital camera, and synthetic aperture radar interferometry (InSAR). The technology for all three of these techniques has been available for the last few years but only recently has it become potentially accessible for multiple use by coastal managers. Through integration, it is anticipated that the optimum monitoring solution in terms of spatio-temporal coverage, precision, accuracy, time and expense will be achieved.

1. INTRODUCTION

1.1 Background

The issue of coastal zone management is increasingly important to government authorities that are confronted with problems of erosion, marine flooding, landslides and other phenomena that affect the existence of coastal assets within their administrative boundaries. The displacement of property and population can be a costly business for local government, especially in built-up areas. Faced with legal consequences of not predicting land movements, it is imperative that monitoring systems are deployed to regularly record changes to the coastline, so that prediction can be attempted. Areas of Britain's coastline are very susceptible to coastal erosion, high profile examples being Gunner Point at the western tip of Hayling Island in Hampshire, Beachy Head in East Sussex and the Hythe coast in Kent.

In areas of erosion or deformation, it is not uncommon to make use of field survey techniques to monitor discrete points in the areas of interest. The standard survey tool for this has traditionally been the total station, but more recent research has concentrated on the use of the global positioning system (GPS). For wider area surveys, photogrammetry, with imagery taken using large format metric film cameras and measurements performed manually on analytical instrumentation, is an accepted method of data production for coastal management purposes. The efficiency of photogrammetric processing in coastal management has been improved in recent years following the introduction of digital photogrammetric workstations that have provided the ability to automate measurements when using scanned photography (Smith and Waldram, 1996). Nevertheless, the inherent problems of slow image acquisition due to film processing and the need to perform extensive ground survey remain (Mills and Newton, 1996). Indeed the processing chain is further complicated by the need for analogue to digital conversion of the imagery by scanning the film. In addition to these factors, deteriorating atmospheric conditions in Europe mean that the number of days suitable for acquiring imagery using large format film cameras is in decline (Reed, 1999). For organisations involved in standard topographic map production this is a problem but it has a far greater impact when monitoring environments that can be subject to rapid change.

1.2 Motivation

The motivation behind this research is to develop a survey technique that is based on three mapping technologies for rapidly monitoring coastal zone erosion. The three technologies are GPS, digital photogrammetry (using imagery from a digital sensor rather than a film camera) and synthetic aperture radar interferometry (InSAR). The technology for all three of these techniques has been available for the last few years but only now has it become accessible for multiple use by the type of user targeted by this research. This research is not only developing each technique individually, but also integrating them to form the optimum monitoring solution in terms of spatio-temporal coverage, precision, accuracy, time and expense.

2. METHODOLOGY

2.1 GPS

The first stage of the monitoring technique involves a comprehensive ground survey of the area of interest by GPS. Early GPS research into coastal monitoring saw the technology used primarily to observe and monitor single points. However, with the development of on-the-fly kinematic post-processing algorithms, GPS can equally be used to record strings of points using a roving receiver and a static base station (Edwards et al., 1999). Dual frequency GPS receivers are used to acquire a coarse resolution, but cm accurate, wireframe digital elevation model (DEM) of the beach, cliff and cliff top areas, by following the breaks in slope, thus defining the shape of the terrain surface with the least amount of data. Where environmental concerns allow, the roving GPS receiver is mounted on an all-terrain vehicle (Figure 1) to speed up data collection and improve efficiency of the survey.

Figure 1. GPS roving receiver mounted on an all-terrain vehicle

2.2 Photogrammetry

The second component technology is digital photogrammetry, utilising a small format digital camera mounted on a microlight platform. Small format cameras have been used previously for aerial survey (Graham, 1988), providing a cost-effective solution for small-area surveys. Advances in the field of digital camera technology have resulted in high resolution sensors, comparable in usage with small and medium format film and with the added practicality of instantly available and inherently stable imagery. Use of a microlight camera platform (Figure 2) reduces costs and enables air survey to take place rapidly, from nearer to the coastal strip than a conventional aircraft. Because a microlight is fast to scramble and can fly below the cloud base, it is less dependent on weather conditions, giving a larger flying window (Warner et al., 1996). Although the ground coverage of a small format camera is limited when compared with conventional large format photography, for narrow coastal strip surveys where only one single strip is needed, this problem is minimised.

Figure 2. Thruster microlight aircraft

The use of small format digital imagery results in a large number of images covering the coastal strip, having implications on the number of photo control points required to transform the stereomodels into the ground coordinate system. Because of automated aerial triangulation methods employed in digital photogrammetric workstations, the amount of control, formerly a minimum of two plan and three height points per stereopair (Rosenholm and Torlegård, 1988), has been greatly reduced in recent years. However, a difficulty associated with surveying in the coastal zone is that few natural or man-made features exist that are easily identifiable (Warner et al., 1996). A common solution to this problem is the use of pre-fabricated ground markers, positioned and coordinated before a flight takes place, but the added time and expense, combined with tidal patterns and unpredictable weather conditions still makes the absolute orientation process the least efficient and most costly in the photogrammetric processing chain.

Conventional photogrammetric processing follows a set workflow: interior orientation, where image space and camera parameters are defined; relative orientation, where stereomodels are formed by identifying conjugate points in the Von Gruber positions of the overlapping images; and absolute orientation, where the stereomodel is scaled and orientated into an object-space reference coordinate system (Wolf and Dewitt, 2000). Using conventional ground control, the absolute orientation phase comprises a least squares adjustment transforming the coordinates of a set of points in the model space to known control in the object space using a three dimensional conformal transformation (comprising a scale factor, three translations and three rotations).

Instead of using ground control points to give absolute orientation to the strips of digital imagery, this research uses independently collected DEMs to scale, translate and orientate the elevation model produced from the relative orientation stage of photogrammetric processing. A least squares three dimensional surface matching algorithm is used, based on the above transformation, in a similar manner to image matching methods employed elsewhere in digital photogrammetry. Rosenholm and Torlegård (1988) introduced the theory of least squares surface matching, and applied it to the absolute orientation of blocks of aerial photography using coarse national-level DEMs. Karras and Petsa (1993) and Mitchell and Chadwick (1999) used surface matching for ultra small-scale medical and dental applications, where the comparison of digital surfaces at different epochs was required, but the use of control markers was both undesirable and unethical. As no control points are used in surface matching, the procedure is instead to register two DEMs, which may have differences due to data collection methods or differences caused by deformation, by a set of transformation parameters, so that the vertical differences between the surfaces are minimised.

To remove the need for costly ground control, the floating photogrammetric surface produced from the relative orientation stage of processing is fused with the DEM collected using GPS. The merged GPS and photogrammetric DEM forms the base model of the coastal zone. A product of the least squares surface matching approach is the ability to detect differences between surfaces (Mitchell and Chadwick, 1999). These show up as residual points that differ greatly in the match. By examining the residual plot it is possible to identify areas of surface difference. Because GPS measures to the ground height and photogrammetry measures to the highest part of the terrain, some of the residuals can be identified as differences due to the data collection, and may be recognised as vegetation, buildings and vehicles. The photogrammetric survey is repeated episodically to develop a temporal model of the coastline. By matching the new DEM on to the old DEM, true surface differences can be identified, indicating where change may have occurred.

2.3 InSAR

The integrated GPS/photogrammetric surface provides a single epoch of data, in essence a 'snapshot' of the coastline at the time of collection. However, to increase the temporal resolution of the model, SAR Interferometry is used to detect changes occurring between aerial photographic surveys. Although the spatial resolution of radar imagery is poor, interferometric processing allows a change detection sensitivity of mm between coherent image pairs. During development of the InSAR technique, applications in earthquake displacement studies (Wright et al., 1999), subsidence monitoring (Strozzi et al., 2001) and landslide detection (Kimura and Yamaguchi, 2000) have been researched. The all-weather nature of radar allows frequent images to be acquired, based on the orbit rate of the satellite, without the need for field survey.

3. CASE STUDY

3.1 Test Site

The North Yorkshire coast is comprised mainly of unstable materials that are highly susceptible to erosion, as the Holbeck Hall landslide in Scarborough in June 1993 showed (Ordnance Survey, 1994). This much-publicised landslide caused considerable damage to property and involved Scarborough Borough Council, the local government authority, in a drawn-out legal battle. Scarborough Borough Council has identified many other potential problem areas, one such area being Filey Bay, the test site for this research study.

Filey is a small town of around 7000 inhabitants on the North Yorkshire coast of England, situated in a 12 km bay (Figure 3). The northern 8.6 km of the bay is mainly comprised of soft glacial till, a material susceptible to erosion by runoff caused by rain, as well as erosion by the sea. A fault line splits the bay, with stable vertical chalk cliffs at the southern end; because of this the larger glacial till stretch of the bay was chosen as the test site. A report commissioned in 1991 estimated an average erosion rate of 0.25 m per year in the bay (Elliott et al., 1991). However, this information was based on only seven discrete erosion posts distributed along the bay, and conceals a more complex periodicity of cliff failures related to storm events and beach dynamics. The local shoreline management plan (one of a series of plans that have been set in place by the Department of Environment, Food and Rural Affairs, a government body, to manage each individual coastal system in Britain) identified the monitoring of cliff processes and beach levels as key issues for the management of Filey Bay.

During the development of this new methodology three sets of fieldwork have been carried out at Filey Bay, in August 2000, August 2001 and March 2002, allowing a temporal change detection model to be initiated. For each of these three fieldwork periods, GPS and photogrammetric data were captured, along with total station cliff profiles for verification of both the surface matching and change detection results.

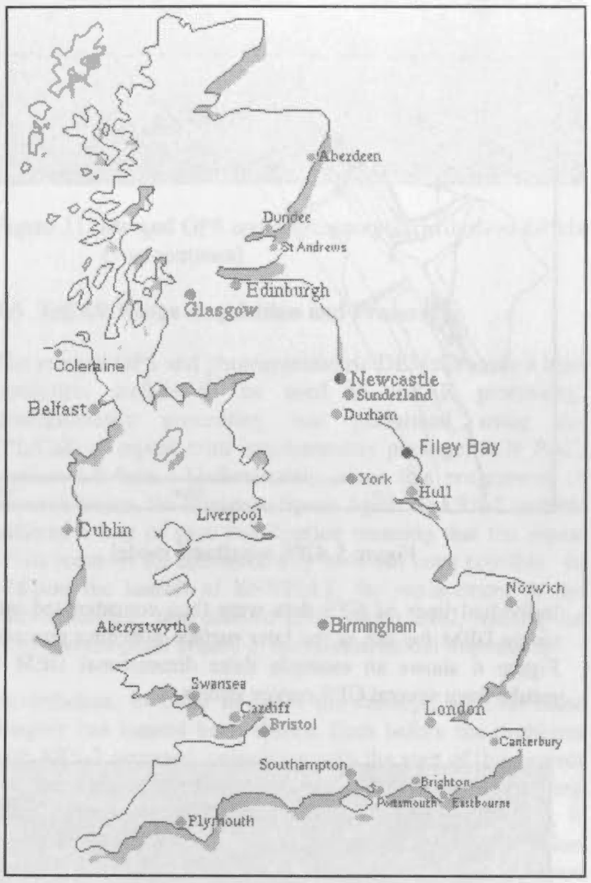


Figure 3. Location of Filey Bay, North Yorkshire, UK

3.2 GPS Survey

Leica System 500 dual frequency GPS receivers were used to traverse the breaks in slope, creating a wireframe DEM of the beaches and cliffs at each epoch of data collection. On the smooth and unobstructed beach an all-terrain vehicle was used to reduce data acquisition times. In other areas the GPSycle (Buckley and Mills, 2001), a standard surveyor's detail pole with a mountain bike wheel attached, was used (Figure 4).

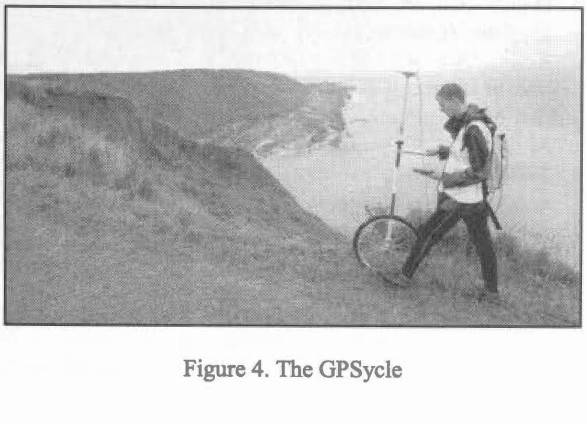


Figure 4. The GPSycle

Multiple height repeatability studies of this GPS methodology were carried out along a known baseline, resulting in a standard deviation of 0.014 m. This figure is better than the precision expected of the photogrammetric measurements made at the second stage of the survey. A wireframe model from a GPS survey is shown in Figure 5.

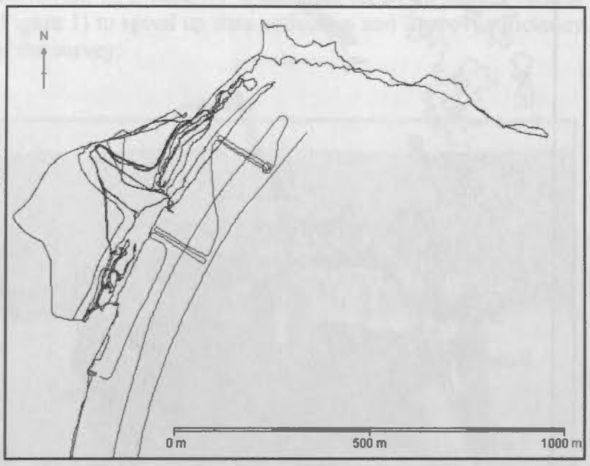


Figure 5. GPS wireframe model

Individual lines of GPS data were then concatenated into a single DEM for use in the later surface matching procedure. Figure 6 shows an example three dimensional DEM that results from several GPS survey strings.

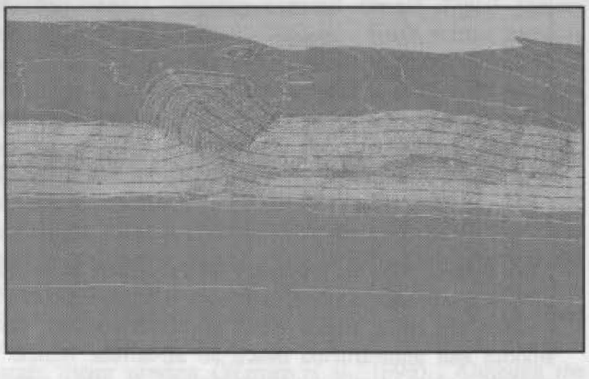


Figure 6. DEM created by GPS strings of points (1 m contours)

3.3 Aerial Survey and Photogrammetric Processing

Digital imagery was acquired from a Thruster Microlight platform (Graham, 1988) as shown in Figure 2, using a Kodak DCS 660 camera (Figure 7). The Kodak DCS 660 is one of a successful series of high resolution SLR cameras that have been employed for mapping (Mason et al., 1997; Mills et al., 1996) and surface modelling (Chandler et al., 2001). With a six megapixel charge-coupled device (CCD) and 9µm element size, the DCS 660 provides a cost effective approach to aerial survey, especially when used in conjunction with the low-cost microlight aircraft.

At a flying height of 2000 feet (600 m), the camera was set at ISO200 and f/4 with an exposure time of 1/400 s. A 28 mm Nikkor lens was used, giving an approximate photoscale of 1:22 000. A fore/aft overlap of 60% was achieved, this configuration providing a ground pixel size of 0.20 m and an expected heighting precision of 0.35 m (Light, 2001).

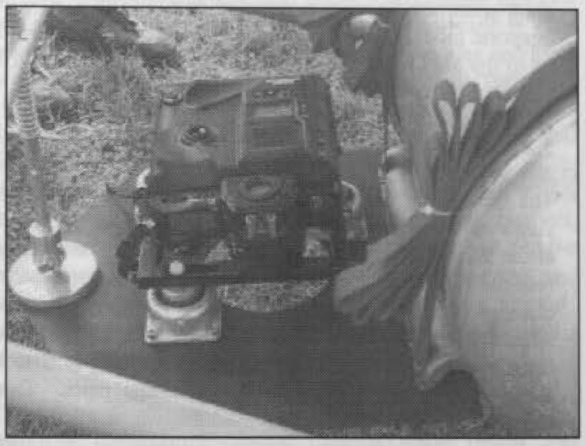


Figure 7. Kodak DCS 660 digital camera mounted in the Thruster microlight

The non-metric nature of the DCS 660 meant that calibration was needed to give accurate internal orientation; this was performed using 54 ground markers laid and coordinated before the flight took place. Because of the small format of the camera, around 50 images were required to give complete stereo coverage of Filey Bay (a typical image can be seen in Figure 8). From this it can be seen that even with sophisticated aerial triangulation algorithms, following conventional photogrammetric workflow would mean a correspondingly large amount of ground control would be needed, increasing the time and expense of the solution.

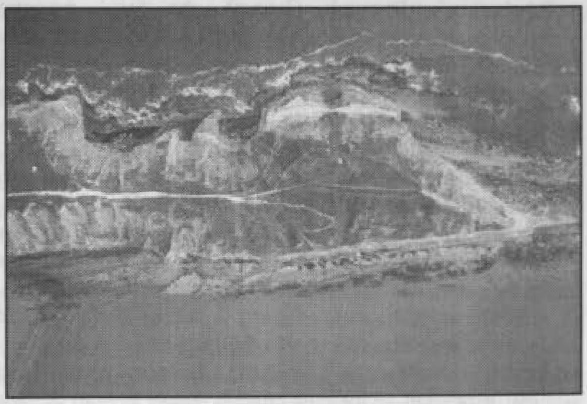


Figure 8. Near-vertical aerial image taken with the Kodak DCS 660 digital camera

Imagery was processed in LH Systems' SOCET Set version 4.3.1, a digital photogrammetric workstation capable of performing model orientation and automatic DEM extraction. Because of the nature of the triangulation algorithms in SOCET Set, some ground control is required to give initial approximations of the block, as both the relative and absolute orientation stages are performed simultaneously. To bypass this, three approximate points were scaled from existing mapping and measured in one stereopair. The remaining images in the Filey Bay strip were then added to this first pair using tie points, until the end of the strip was reached. With the strip orientated, DEMs were then automatically extracted and edited within SOCET Set (Figure 9). The resolution of the DEMs created varied according to the nature of the monitored features. Coarser DEMs were measured over large stretches of coastline, and then smaller, finer models added in areas of more detailed processes, such as landslides or embayments.

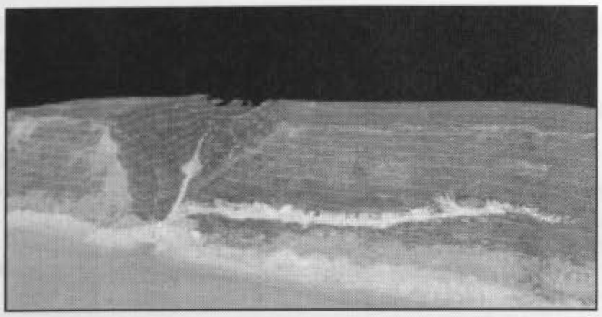


Figure 9. DEM generated in LH Systems SOCET Set from the Kodak DCS 660 aerial imagery (3 m contours)

3.4 GPS and Photogrammetric Data Fusion

As only three ground control points were used in the transformation of the model into the absolute reference coordinate system, the accuracy of this transformation decreased along the strip as the distance to the control increased. As a result of this, a comparison with the GPS surface towards the strip end showed that the photogrammetric DEM was rotated slightly and was around 60 m below the 'true' surface level (Figure 10); hence surface matching was crucial in removing this transformation error and merging the two data collection techniques.

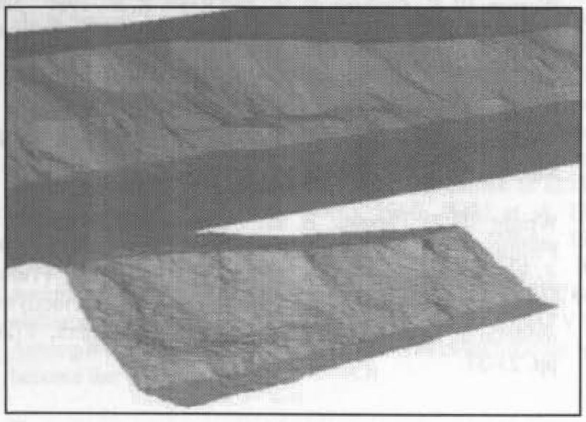


Figure 10. Photogrammetric DEM (bottom) in relation to the GPS derived DEM (top) at the end of the strip

Photogrammetric DEMs were matched to the corresponding GPS areas using the least squares surface matching algorithm, resulting in integrated surfaces (Figure 11), more accurate than when using a single technique alone. Total station cliff profiles were measured at six 'control areas' along the 8.6 km coastal strip; these were used in match verification, showing correlation values of up to 0.98 between the merged GPS and photogrammetric surfaces.

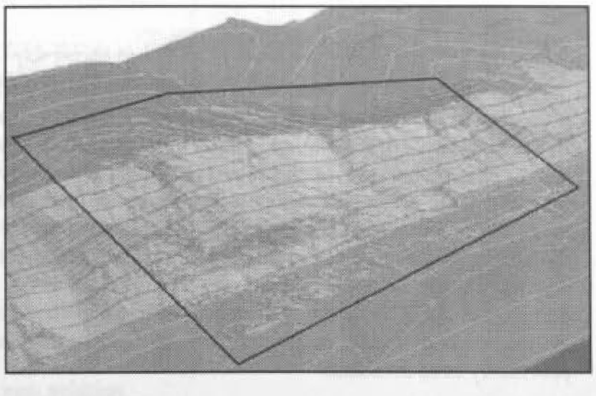


Figure 11. Merged GPS and photogrammetric derived DEMs (1 m contours)

3.5 InSAR Image Acquisition and Processing

The merged GPS and photogrammetric DEMs provide a high resolution surface to be used in InSAR processing. Interferometric processing was performed using the JPL/Caltech repeat orbit interferometry package, ROI_PAC, version 1.0 beta. Unfortunately, since this programme of research began, the European Space Agency's ERS-2 satellite suffered a loss of gyro stabilisation meaning that the repeat orbits required for interferometry have not been possible. In addition the launch of ENVISAT, the replacement to the ERS-2 satellite, was delayed until March 2002, making the proposed temporal update to the coastal model impossible.

Nevertheless, in order to prove the concept, archival radar imagery has instead been chosen from before the problems with ERS-2 occurred, coinciding with the start of this project and the time of the first fieldwork. Current work is being carried out to process these images (Figure 12) which, if successful, will give an historical insight into the erosion processes active in Filey Bay. Areas of coherence between image pairs will highlight small coastal changes, whereas it is expected that incoherent areas may be indicative of larger changes having occurred, requiring field inspection or further survey to be carried out.

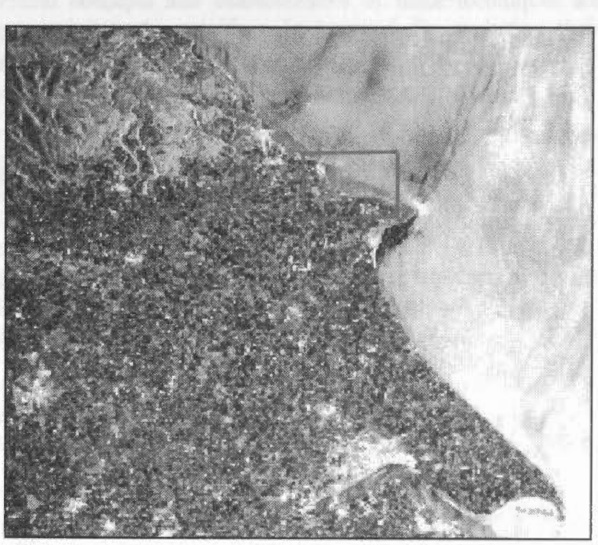


Figure 12. 100 km^2 ERS-2 SAR Scene of North Yorkshire, UK. Image acquired on 15th May, 2000. Filey Bay is Boxed. © ESA

4. CONCLUSIONS

An ongoing project to develop an optimum solution to the problems of monitoring coastal erosion has been described. In the past surveying techniques have often proved difficult and expensive because of the inherent problems associated with the wide areas and dynamic processes in the coastal zone. By incorporating the high positional accuracy of GPS, the wide area coverage of digital photogrammetry and the all-weather change detection capabilities of InSAR, synergy is derived to form a faster and more efficient monitoring system, with a higher spatio-temporal resolution than has previously been available.

Although GPS and photogrammetry form the core of this new monitoring solution, the technique is generic so that different surface measuring technologies could be integrated to form a coastal model. The use of least squares surface matching is an elegant method of fusing data, and could be applied to Lidar data (both terrestrial or airborne) in preference over, or in addition to, GPS or photogrammetry. Further, the application of this methodology is not restricted to coastal zone monitoring and modelling. Environmental and engineering applications that could adopt this type of monitoring methodology are plentiful, with the technique now being utilised, for example, in an investigation into the reactivation of post-mining ground subsidence in the UK.

REFERENCES

Buckley, S. and Mills, J., 2001. Synergy of new geomatics technologies for coastal zone studies. *Engineering Surveying Showcase*, 2001(2), pp. 29-31.

Chandler, J. H., Shiono, K., Rameshwaren, P. and Lane, S. N., 2001. Measuring flume surfaces for hydraulics research using a Kodak DCS460. *Photogrammetric Record*, 17(97), pp. 39-62.

Edwards, S.J., Cross, P.A., Barnes, J.B. and Betaille, D., 1999. A methodology for benchmarking real time kinematic GPS. *Survey Review*, 35(273): 163-174.

Elliott, M., Jones, N. V., Lewis, S., Pethick, J. S. and Symes, D. G., 1991. *Filey Bay Environmental Statement*. Institute of Estuarine and Coastal Studies, University of Hull, Hull, pp. 131.

Graham, R. W, 1988. Small format aerial surveys from light and microlight aircraft. *Photogrammetric Record*, 12(71), pp. 561-573.

Karras, G. E. and Petsa, E., 1993. DEM matching and detection of deformation in close-range photogrammetry without control. *Photogrammetric Engineering and Remote Sensing*, 59(9), pp. 1419-1424.

Kimura, H. and Yamaguchi, Y., 2000. Detection of landslide areas using satellite radar interferometry. *Photogrammetric Engineering and Remote Sensing*, 66(3), pp. 337-344.

Light, D., 2001. An airborne direct digital imaging system. *Photogrammetric Engineering and Remote Sensing*, 67(11), pp. 1299-1305.

Mason, S., Ruther, H. and Smit, J., 1997. Investigation of the Kodak DCS460 digital camera for small-area mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(5), pp. 202-214.

Mills, J. P. and Newton, I., 1996. A new approach to the verification and revision of large scale mapping. *ISPRS Journal of Photogrammetry & Remote Sensing*, 51(1), pp. 17-27.

Mills, J. P., Newton, I. and Graham, R. W., 1996. Aerial photography for survey purposes with a high resolution, small format, digital camera. *Photogrammetric Record*, 15(88), pp. 575-587.

Mitchell, H. L. and Chadwick, R. G., 1999. Digital photogrammetric concepts applied to surface deformation studies. *Geomatica*, 53(4), pp. 405-414.

Ordnance Survey, 1994. Frontispiece. *Photogrammetric Record*, 14(83), pp. 712.

Reed, R. E., 1999. Aspects of change in air survey primary data acquisition. *Photogrammetric Record*, 16(93), pp. 417-422.

Rosenholm, D. and Torlegård, K., 1988. Three-dimensional absolute orientation of stereo models using digital elevation models. *Photogrammetric Engineering and Remote Sensing*, 54(10), pp. 1385-1389.

Smith, M. J. and Waldram, D. A. 1996. Automatic digital terrain modelling of coastal zones. *International Archives of Photogrammetry and Remote Sensing*, 31(B4): 919-924.

Strozzi, T., Wegmuller, U., Tosi, L., Bitelli, G. and Spreckels, V., 2001. Land subsidence monitoring with differential SAR interferometry. *Photogrammetric Engineering and Remote Sensing*, 67(11), pp. 1261-1270.

Warner, W. S., Graham, R. W. and Read, R. E., 1996. *Small Format Aerial Photography*. Whittles Publishing, Caithness, Scotland, pp. 348.

Wolf, P. R. and Dewitt, B. A., 2000. *Elements of Photogrammetry (with Applications in GIS), Third Edition*. McGraw-Hill, Inc., New York, pp. 624.

Wright, T. J., Parsons, B. E., Jackson, J. A., Haynes, M., Fielding, E. J., England, P. C. and Clarke, P. J., 1999. Source parameters of the 1 October 1995 Dinar (Turkey) earthquake from SAR interferometry and seismic bodywave modelling. *Earth and Planetary Science Letters*, 172(1), pp. 23-37.

ACKNOWLEDGEMENTS

This research is funded by the Engineering and Physical Sciences Research Council (EPSRC), grant number GR/N23721. The authors would also like to thank Scarborough Borough Council, Aerial Imaging Systems, Geotechnologies, and the Natural Environment Research Council (NERC) sponsored CHASM project for their assistance in undertaking the fieldwork and loan of equipment.

PROBLEMS IN THE FUSION OF COMMERCIAL HIGH-RESOLUTION SATELLITE AS WELL AS LANDSAT 7 IMAGES AND INITIAL SOLUTIONS

Yun Zhang
Department of Geodesy and Geomatics Engineering
University of New Brunswick
Fredericton, New Brunswick, Canada E3B 5A3
Email: yunzhang@unb.ca

Commission IV, WG IV/7

KEY WORDS: Image fusion, IKONOS, Landsat 7, colour distortion, new solution

ABSTRACT:
Currently available image fusion techniques are not efficient for fusing the new satellite images such as IKONOS and Landsat 7. Significant colour distortion is one of the major problems. Reasons for this distortion are analyzed in this paper. A new initial automatic solution for fusing the new images is introduced. The new solution has reached an optimum fusion result – with minimized colour distortion and maximized spatial detail.

1. INTRODUCTION

The fusion of low-resolution multispectral and high-resolution panchromatic satellite images is a very important issue for many remote sensing and mapping applications. Usually, a panchromatic band covers a broad wavelength range of the visible and near infrared spectrum, while a multispectral band just covers a narrow spectral range. For receiving a same amount of incoming energy, the size of a panchromatic detector can be smaller than that of a multispectral detector. Therefore, most sensors of earth recourse satellites, such as SPOT, IRS, Landsat 7, IKONOS and QuickBird, provide panchromatic and multispectral images at different spatial resolutions. To effectively utilize such images, techniques that can effectively combine high-resolution panchromatic and low-resolution multispectral images into one colour image are demanded. Such techniques can largely extend the application potential of the raw remote sensing images.
Many image fusion algorithms and software tools have been developed (Pohl and Van Genderen 1998). The well-known techniques are, for example, the IHS (Intensity, Hue, Saturation), PCA (Principal Components Analysis), arithmetic combination based fusion, and wavelet based fusion (Cliche et al.1985, Welch and Ehlers 1987, Albertz *et al.* 1988, Chavez *et al.* 1991, Ehlers 1991, Shettigara 1992, Munechika et al.1993, Garguet-Duport et al. 1996, Yocky 1996, Wald et al. 1997, Zhou et al. 1998, Zhang and Albertz 1998, Zhang 1999, Ranchin and Wald 2000). Most techniques have been developed based on the fusion of the SPOT pan with other multispectral images, such as Landsat TM and SPOT HRV XS. Among the fusion techniques the IHS and PCA techniques have become the very popular.

However, a common problem associated with the available techniques is the colour distortion of the fused images. To date, well-fused satellite images with SPOT or IRS panchromatic band as a high-resolution input have been presented in some publications; however, publications on well-fused IKONOS and Landsat 7 images have been rarely seen. Another common problem associated with the existing algorithms is that the fusion quality is operator and data dependent. Different operators or data sets may lead to different fusion qualities.

In this study, the principles of different fusion techniques are reviewed. The spectral characteristics of the available earth recourse satellite images are analyzed. Available software tools (IHS, PCA etc.) are applied to fusing different images. Based on the analyses of the spectral relationships and different fusion results, a new technique has been developed. This technique can effectively fuse the new satellite images, e.g., IKONOS and Landsat 7, and reduce the manual interactions during the process. The spectral distortion of the fused images is significantly reduced and the fusion quality becomes operator and data set independent.

2. CHARACTERISTICS OF AVAILABLE TECHNIQUES

Pohl and Van Genderen (1998) provided a comprehensive review of image fusion techniques. Among the numerous techniques the following four types of fusion techniques have demonstrated most promising results for effectively fusing SPOT pan with other images, e.g., IHS (Intensity, Hue, Saturation), PCA (Principal Components Analysis), arithmetic combination based fusion, and wavelet based fusion. The general concepts and characteristics of these techniques are reviewed here to provide a background for analyzing their limitation for IKONOS as well as Landsat 7 fusions.

2.1. HIS technique

The IHS colour transform can effectively separate a standard RGB (Red, Green, Blue) image into spatial (I) and spectral (H, S) information. The basic concept of IHS fusion is shown in Figure 1. The most important steps are: (1) transform a colour image composite from the RGB space into the IHS space, (2) replace the I (intensity) component by a panchromatic image with a higher resolution, (3) reversely transform the replaced components from IHS space back to the original RGB space to obtain a fused image. The IHS technique has become a standard procedure in image fusion (Albertz *et al.* 1988, Chavez *et al.* 1991, Ehlers 1991, Shettigara 1992, Zhang and Albertz 1998, Zhang 1999).

2.2. PCA technique

The PCA is a statistical technique that transforms a multivariate inter-correlated data set into a new un-correlated data set. The basic concept of PCA fusion is shown in Figure 2.

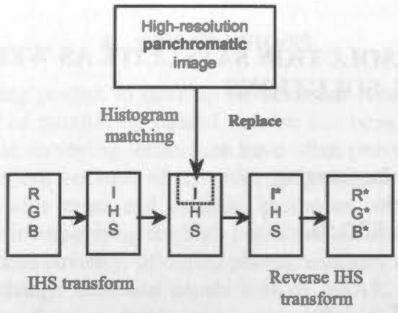

High-resolution panchromatic image
Histogram matching
Replace
R G B
I H S
I H S
I' H S
R* G* B*
IHS transform
Reverse IHS transform

Figure 1. Schematic flowchart of IHS image fusion

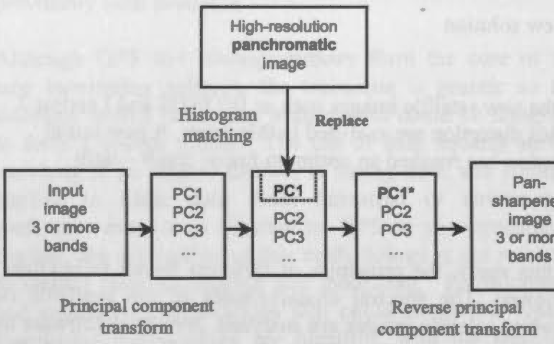

High-resolution panchromatic image
Histogram matching
Replace
Input image 3 or more bands
PC1 PC2 PC3 ...
PC1 PC2 PC3 ...
PC1* PC2 PC3 ...
Pan-sharpened image 3 or more bands
Principal component transform
Reverse principal component transform

Figure 2. Schematic flowchart of PCA image fusion

Its most important steps are: (1) perform a principal component transformation to convert a set of multispectral bands (three or more bands) into a set of principal components, (2) replace one principal component, usually the first component, by a high resolution panchromatic image, (3) perform a reverse principal component transformation to convert the replaced components back to the original image space. A set of fused multispectral bands is produced after the reverse transform (Chavez *et al.* 1991, Shettigara 1992, Zhang and Albertz 1998, Zhang 1999).

2.3. Arithmetic Combination technique

Different arithmetic combinations have been employed for fusing multispectral and panchromatic images. The arithmetic operations of multiplication, division, addition and subtraction have been combined in different ways to achieve a better fusion effect. Brovey Transform, SVR (Synthetic Variable Ratio), and RE (Ratio Enhancement) techniques are some successful examples for SPOT pan fusion. Brovey Transform uses addition, division and multiplication for the fusion of three multispectral bands (ERDAS 1999). Its concept can be described with equation 1. Its basic processing steps are: (1) add three multispectral bands together for a sum image, (2) divide each multispectral band by the sum image, (3) multiply each quotient by a high resolution pan. The SVR and RE techniques are similar, but involved more sophisticated calculations for the sum image (Cliche et al.1985, Welch and Ehlers 1987, Chavez *et al.* 1991, Munechika et al.1993, Zhang and Albertz 1998, Zhang 1999).

$$Fusion_i = \frac{Multi_i}{Multi_{Sum}} \times Pan \qquad (1)$$

with $i = 1, 2$ and 3, and $Multi_{Sum} = Multi1 + Multi\,2 + Multi\,3$.

2.4. Wavelet technique

Wavelet transform is a mathematical tool developed in the field of signal processing. It can decompose a digital image into a set of multi-resolution images accompanied with wavelet coefficients for each resolution level. The wavelet coefficients for each level contain the spatial (detail) differences between two successive resolution levels. The wavelet based fusion is performed in the following way (Figure 3): (1) decompose a high resolution panchromatic image into a set of low resolution panchromatic images with wavelet coefficients for each level, (2) replace a low resolution panchromatic with a multispectral band at the same resolution level, (3) perform a reverse wavelet transform to convert the decomposed and replaced panchromatic set back to the original panchromatic resolution level. The replacement and reverse transform is done three times, each for one multispectral band (Garguet-Duport et al. 1996, Yocky 1996, Wald et al. 1997, Zhou et al. 1998, Ranchin and Wald 2000).

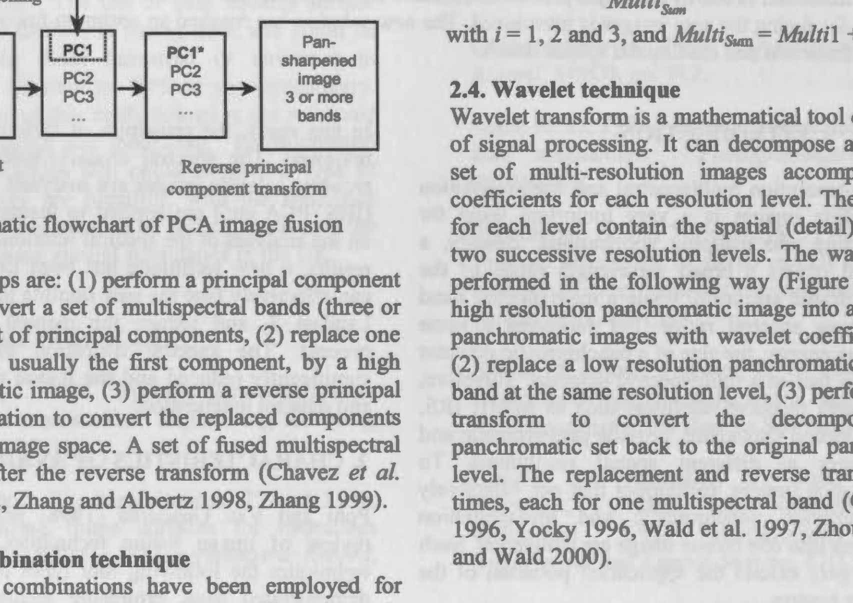

Multi 1 (20 m)
Multi 2 (20 m)
Multi 3 (20 m)
2.
Pan (10 m)
1.
Pan (20 m)
Vertical Details (10m)
Horizontal Details (10m)
Diagonal Details (10m)
2.
Multi 1 (20 m)
Vertical Details (10m)
Horizontal Details (10m)
Diagonal Details (10m)
3.
Pan-sharpened 1 (10 m)
Multi 2 (20 m)
Vertical Details (10m)
Horizontal Details (10m)
Diagonal Details (10m)
3.
Pan-sharpened 2 (10 m)
Multi 3 (20 m)
Vertical Details (10m)
Horizontal Details (10m)
Diagonal Details (10m)
3.
Pan-sharpened 3 (10 m)
1. Wavelet transform
2. Replace Pan (20m) with Multi (20m) respectively
3. Inverse wavelet transform

Figure 3. Schematic flowchart of wavelet based image fusion

3. PROBLEMS OF AVAILABLE TECHNIQUES

Problems and limitations associated with the available fusion techniques have been reported by many studies (Chavez et al. 1991, Pellemans et al. 1993, Wald et al. 1997, Van Der Meer 1997). The most significant problem is the colour distortion of fused images. Another most common problem is the operator and data dependency of the fusion quality. To reduce the colour distortion, different strategies have been developed for different techniques; nevertheless operator's experience still spays an important role for a good fusion result.

- The colour distortion of IHS technique is often significant. To reduce the distortion, the suggested strategies are, for example, matching the panchromatic to the intensity before the replacement, stretching the hue and saturation components before the reverse transform, or stretching individual I, H or S components with respect to individual data sets.
- In PCA image fusion, dominant spatial information and weak colour information is an often problem. The reason is that the first principal component being replaced usually contains a maximum variance. Such replacement maximizes the effect of panchromatic image in the fused image. Suggested solutions have been, for example, stretching the principal components to give a spherical distribution, or discarding the first principal component. The PCA approach is sensitive to the choice of area to be fused.
- With Brovey Transform, it is necessary to select only three bands. In general, the pixel grey values are smaller than those of other fusion techniques. The colour distortion is obvious and varies depending on the band combinations being fused. With the SVR and RE techniques, the colour effect of the fused image is generally better than Brovey Transform, but a suitable preprocessing is necessary before the fusion. Operator's experience plays an important role for all the three techniques.
- Wavelet based fusion is a computationally intensive process. It extracts spatial details from a high-resolution panchromatic image, and then adds them into the multispectral bands. In this manner, the colour distortion can be reduced to a certain extent, but the fused image appears like a result of a high-pass filtering fusion, e.g., the colour seems not being smoothly integrated into the spatial features. Other disadvantages have also been reported, for example, the loss of spectral content of small objects.

To date, successful results of fusing Landsat TM, SPOT HRV, or IRS LISS with SPOT panchromatic or IRS panchromatic images have been shown in many publications. Different strategies for reducing colour distortions and operator's fusion experiences have played an important role for such successes. However, colour distortions still remain. Figure 4 shows an example of the fused Landsat TM and SPOT pan images using HIS, PCA and SVR techniques. Different colour distortions can still be recgonized.

When the same fusion techniques are applied to the fusion of IKONOS or Landsat 7 images, the colour distortions are much more significant, even if great operator efforts have been employed. Figure 5 shows IKONOS multispectral, panchromatic and IHS fused images in two different areas – Fredericton, Canada and San Diego, USA. The same IHS algorithm and fusion parameters were applied to the two image sets. It can be clearly seen in the fused results that (1) the colour is totally distorted from their original colour, and (2) the colour distortions tend in different ways for different sets. Other techniques, such as PCA and SVR, produce similar results with significant colour distortions.

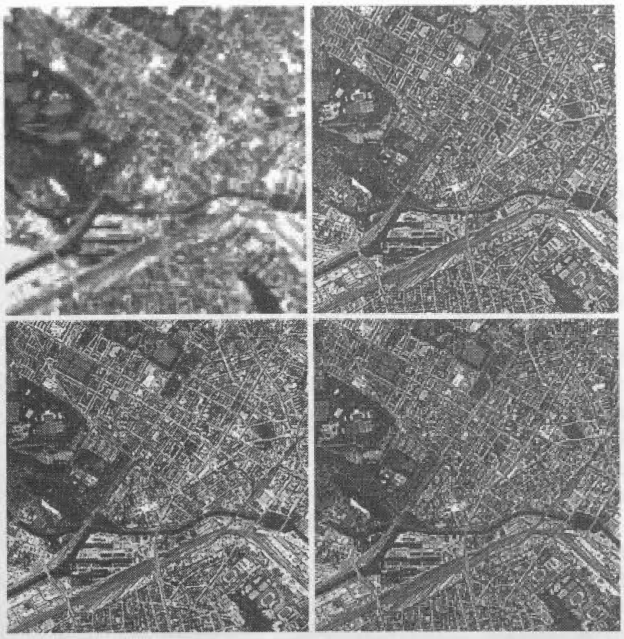


Figure 4: Landsat TM and SPOT pan fusion. (a) Original TM colour image, (b) PCA fusion result, (c) SVR fusion result, (d) IHS fusion result. (with a: upper left, b: upper right, c: lower left, d: lower right)

With wavelet base fusion technique, the colour distortion may be reduced to some extent. But, the integration of colour and spatial feature usually does not appear smooth. The spectral content of small objects is lost for many cases (Yocky 1996, Ranchin and Wald 2000). Therefore, this technique has not been tested in this study.

4. REASONS FOR THE COLOUR DISTORTION

A major reason for the insufficiency of available techniques for IKONOS fusion as well as Landsat 7 fusion is the change of the panchromatic spectral range. Different from SPOT panchromatic and IRS panchromatic images, the wavelength of IKONOS and Landsat 7 panchromatic images (same to QuickBird pan) was extended from visible to near infrared range (Table 1). This change makes the grey value relationship of an IKONOS or Landsat 7 panchromatic image significant different from that of a SPOT panchromatic or IRS panchromatic image (Figure 6a and 6b). For example, vegetation areas in the SPOT panchromatic image (Figure 6a) appear darker than pavement areas. However, they appear brighter than pavement areas in the IKONOS pan image because of the influence of near infrared content. Therefore, it is not a wonder that conventional fusion algorithms, which have been successful for the fusion of SPOT panchromatic and other multispectral images, cannot effectively fulfill the fusion of the new images.

Figure 6c shows the intensity (I) image transformed from IKONOS multispectral bands 1, 2 and 3. The grey value difference between the IKONOS pan and the I image is much more significant than that between the SPOT pan and the I image. When the I image is replaced by a significantly different image – IKONOS pan, the IHS fusion will, of course, lead to a significant colour distortion. In the opposite, if the I image was replaced by a pan image like the SPOT pan, the colour distortion of the IHS fusion would be much smaller. This comparison clearly illustrated the reason of the colour distortion of the IHS technique. The first component of the PCA technique (PC1) is also significantly different from the original IKONOS pan image. If the grey values of an IKONOS pan could be adjusted to the grey values similar to an I or PC1 image before the replacement, the colour distortion would be significantly reduced.

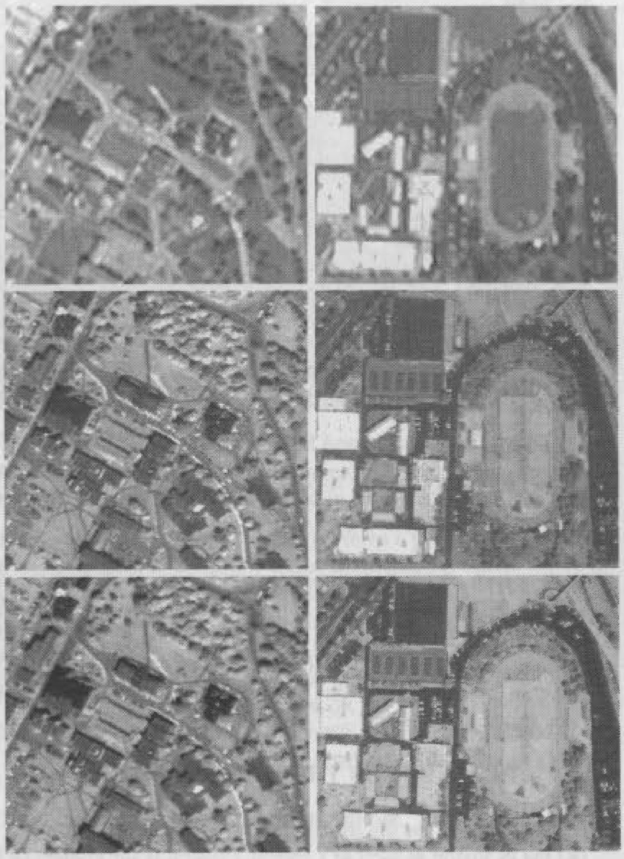


Figure 5. Significant colour distortions of the IHS technique for IKONOS fusion. From top to bottom: original IKONOS natural colour images with 4 m resolution, original IKONOS panchromatic images with 1 m resolution, and IHS fusion results. Left: Fredericton, Canada. Right: San Diego, USA.

Table 1. Spectral ranges of different panchromatic images

Satellite sensor	Spectral range (µm)
SPOT 1, 2, 3 (HRV)	0.51 – 0.73
SPOT 5 (HRG)	0.51 – 0.73
IRS 1C, 1D (PAN)	0.50 – 0.75
Landsat 7 (ETM+)	0.52 – 0.90
IKONOS	0.45 – 0.90
QuickBird	0.45 – 0.90

5. INITIAL SOLUTIONS AND RESULTS

Based on thorough studies and analyses of existing fusion algorithms and their fusion effects, an initial new automatic fusion approach has been developed by the author. This new technique solved the two major problems in image fusion – colour distortion and operator dependency. A technique based on least squares was employed for a best approximation of the grey value relationship between the original multi, panchromatic and the fused images for a best colour representation. Statistic approaches were applied to the fusion for standardizing and automating the fusion process.

Figure 7 shows two results of the new automatic fusion technique. No colour adjustment and manual interaction were employed during the fusion. In other words, the operator just needs to input the registered original multispectral and panchromatic images, and then receive the fusion result. Compared to the original colour image, the colour of the fused image of the new technique stays almost the same as the original one. The spatial details of the panchromatic image are also perfectly integrated into the fused image. An enormous improvement of the new fusion technique over the IHS fusion can clearly seen by comparing Figure 7 and Figure 5.

Another fusion result of the new technique is shown in Figure 8. The image covers a part of the main campus of the University of New Brunswick, Canada. The original 4m multispectral and 1m panchromatic images were taken in October 2001 when the maple trees began turning red. Compared to the original colour image, the colour of the fused image stays almost the same as the original image for all objects. In comparison with the original panchromatic image, the detail of spatial features in the fused image is as clear as in the pan image. Further, the colour and spatial detail of the original images have been perfectly integrated. No artificial effects between colour and spatial features can be recognized. Even spatial details under shadows can be seen in the fused image.

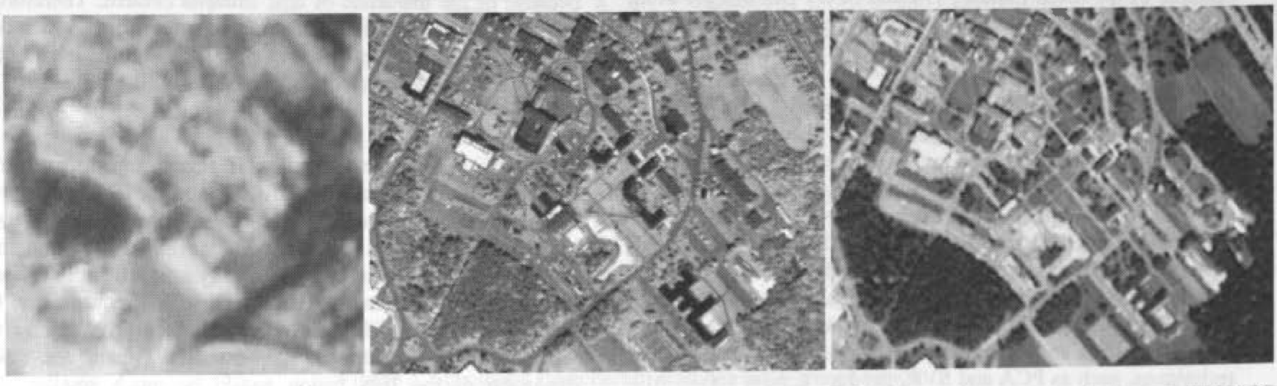


Figure 6. Comparison of grey value differences between SPOT panchromatic (left), IKONOS panchromatic (middle), and IKONOS Intensity (right) images (with a: left, b: middle, c: right)

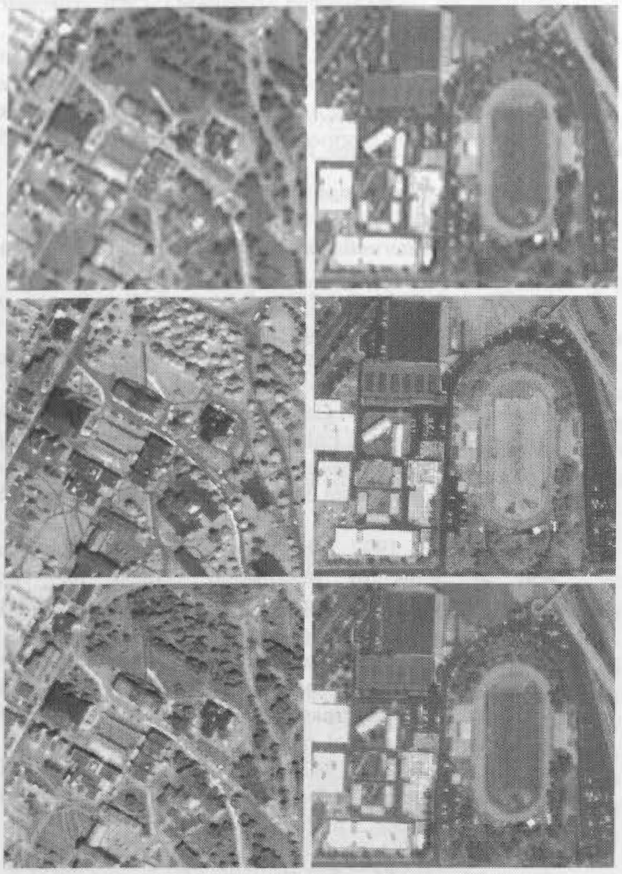


Figure 7. Minimized colour distortions of the new fusion technique. From top to bottom: original IKONOS natural colour images with 4 m resolution, original IKONOS panchromatic images with 1 m resolution, and IHS fusion results. Left: Fredericton, Canada. Right: San Diego, USA.

For different application purposes, three different fusion models – spectral information preserving model, spatial information preserving model, and colour enhancing model – have also been developed.

- The spectral information preserving model is developed mostly for digital classification purposes (Figure 9b). In this model, the colour of the fused image is kept as same to the original image as possible. The grey value variances within individual objects are kept smooth. These properties are important to conventional spectral based classifications for higher classification accuracy.
- The spatial information preserving model is developed for visual interpretation, image mapping, and photogrammetric purposes (Figure 9c). Spatial details and colour fidelity are both important in this model. The colour fidelity of the spatial model is not as high as the spectral model for some parts of objects; however, details of an object are significantly increased and the clarity of the image is improved.
- The colour enhancing model is developed for visualization and GIS integration purposes. In many cases, the colour of an IKONOS natural colour composite does not appear fresh. For example, vegetation areas are often darker and with less green intensity than we usually see on the ground. With this model, the colour of vegetation areas can be enhanced appearing fresher than in the original image, but the colour of other objects are keeping unchanged (Figure 9d).

6. CONCLUSIONS

The initial fusion technique has been tested with many IKONOS and Landsat 7 data sets covering different areas. The sizes of some data sets were more than 10,000 by 10,000 pixels. All the fusions have shown a perfect result – maximum detail increasing and minimum colour distortion.

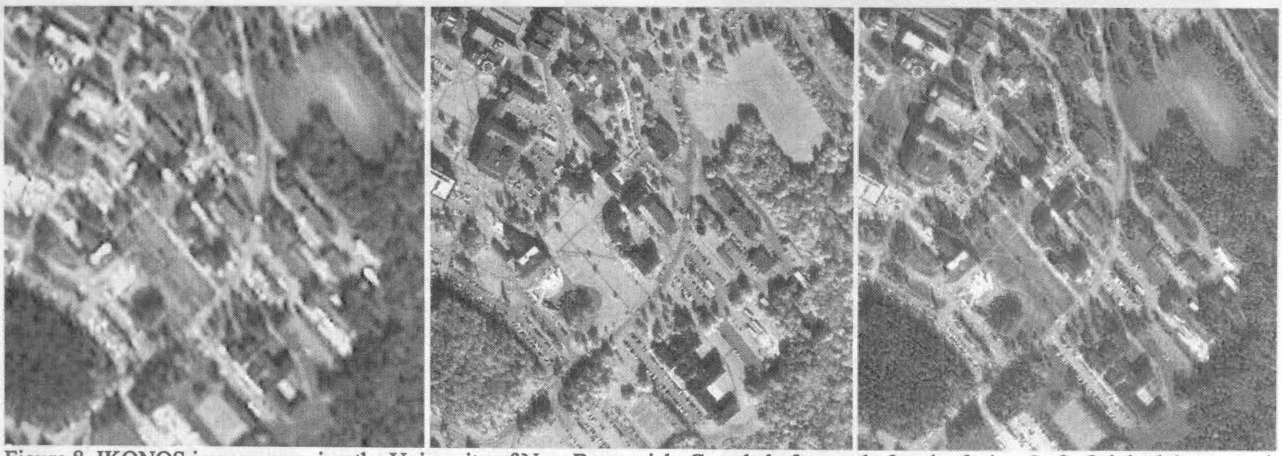


Figure 8. IKONOS images covering the University of New Brunswick, Canada before and after the fusion. Left: Original 4m natural colour image, Middle: Original 1m panchromatic image, Right: Automatically fused natural colour image of the new technique.

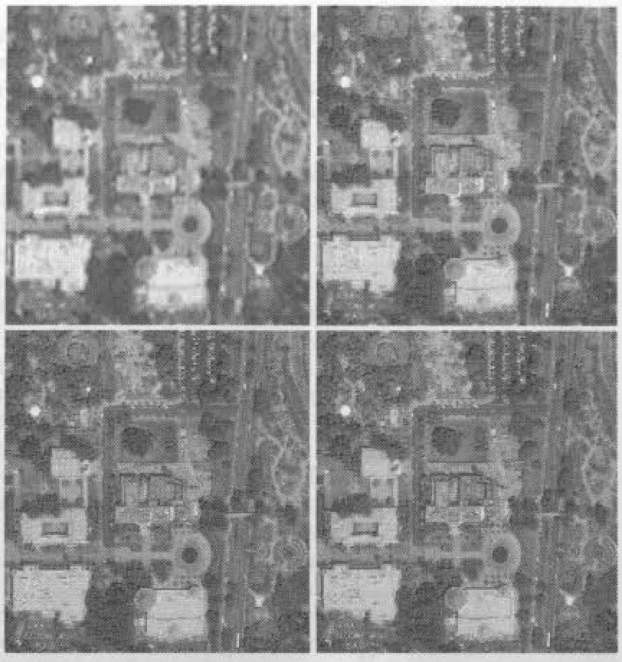


Figure 9. Effects of three fusion models of the new fusion technique. (a) Original IKONOS 4 m natural colour image, (b) Fusion result of the spectra preserving model. (c) Fusion result of the spatial detail preserving model. (d) Fusion result of the colour enhancing model. (with a: upper left, b: upper right, c: lower left, d: lower right)

ACKNOWLEDGEMENTS

The author thanks Mr. Rob Lunn, GIS supervisor of the City of Fredericton, Canada and Space Imaging for providing raw IKONOS images for this research.

References

Albertz, J., Lehmann, H., Mehlbreuer, A., Scholten, F., Tauch, R., 1988. Herstellung hochauflösender Satelliten-Bildkarten durch Kombination Multisenoraler Datensätze. In: Internationales Jahrbuch für Kartographie, band 28, Ulm: Universitätsverlag, pp. 11-27.

Chavez, P. S., Sides, S. C., Anderson, J. A., 1991. Comparison of Three Different Methods to Merge Multiresolution and Multispectral Data: Landsat TM and SPOT Panchromatic. *Photogrammetric Engineering & Remote Sensing*, Vol. 57, No. 3, pp. 295-303.

Cliche, G., Bonn, F., Teillert, P., 1985. Integration of the SPOT Panchromatic Channel into Its Multispectral Mode for Image Sharpness Enhancement. *Photogrammetric Engineering & Remote Sensing*. Vol. 51, pp. 311-316.

ERDAS, 1999. On-Line Manuals version 8.4, Atlanta, Georgia.

Ehlers, M., 1991, Multisensor image fusion techniques in remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 46, pp. 19-30.

Garguet-Duport, B., Girel, J., Chassery, J.-M., and Pautou, G., 1996. The Use of Multiresolution Analysis and Wavelets Transform for Merging SPOT Panchromatic and Multispectral Image Data. *Photogrammetric Engineering & Remote Sensing*, Vol. 62, No. 9, pp. 1057-1066.

Munechika, C. K., Warnick, J. S., Salvaggio, C., Schott, J. R., 1993. Resolution Enhancement of Multispectral Image Data to Improve Classification Accuracy. *Photogrammetric Engineering & Remote Sensing*, Vol. 59, No. 1, pp. 67-72.

Pellemans, A., Jordans, R., Allewijn, R., 1993. Merging Multispectral and Panchromatic SPOT Images with Respect to the Radiometric Properties of the Sensor. *Photogrammetric Engineering & Remote Sensing*, Vol. 59, No. 1, pp. 81-87.

Pohl, C. and Van Genderen, J. L., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*, Vol. 19, pp. 823-854.

Ranchin, T. and Wald, L., 2000. Fusion of High Spatial and Spectral Resolution images: The ARSIS Concept and Its Implementation. *Photogrammetric Engineering and Remote Sensing*, vol. 66, pp. 49-61.

Shettigara, V. K., 1992. A Generalized Component Substitution Technique for Spatial Enhancement of Multispectral Images Using a Higher Resolution Data Set. *Photogrammetric Engineering & Remote Sensing*, Vol. 58, No. 5, pp. 561-567.

Van Der Meer, F., 1997. What does multisensor image fusion add in terms of information content for visual interpretation? *INT. J. Remote Sensing*, Vol. 18, No. 2, pp. 445-452.

Wald, L., Ranchin, T., Mangolini, M., 1997. Fusion of Satellite Images of Different Spatial Resolutions: Assessing the Quality of Resulting Images. *Photogrammetric Engineering & Remote Sensing*. Vol. 63, No. 6, pp. 691-699.

Welch, R., Ehlers, M., 1987. Merging Multiresolution SPOT HRV and Landsat TM Data. *Photogrammetric Engineering & Remote Sensing*. Vol. 53, No. 3, pp. 301-302.

Yocky, D.A., 1996. Multiresolution Wavelet Decomposition Image Merger of Landsat Thematic Mapper and SPOT Panchromatic Data. *Photogrammetric Engineering and Remote Sensing*, vol. 62, pp. 1067-1074.

Zhang, Y., and Albertz, J., 1998. Vergleich verschiedener Verfahren zur Kombination multisensoraler Satelliten-Bilddaten. Photogrammetrie-Fernerkundung-Geoinformation, Vol. 5, pp. 261-274.

Zhang, Y., 1999. A new merging method and its spectral and spatial effects. *International Journal of Remote Sensing*, vol. 20, pp. 2003-2014.

Zhou, J., Civco, D.L., and Silander, J.A., 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. *International Journal of Remote Sensing*, Vol. 19, No. 4, pp. 743-757.

UBIQUITOUS GIS, DATA ACQUISITION AND SPEECH RECOGNITION

Andrew Hunter[1,*] and Dr. C. Vincent Tao[2]

[1] Department of Geomatics Engineering, University of Calgary,
[2] Department of Earth and Atmospheric Science, York University
Commission IV, WG IV/7 - Data Integration and Digital Mapping

KEY WORDS: Mobile GIS, Wireless Computing, Wearable Computers, Speech Recognition, Data Acquisition

ABSTRACT:

The research asked whether Mobile GIS that incorporated speech recognition was a viable tool for locating defects in pavement, curbs and footpaths. A Xybernaut MA IV wearable computer and Dragon NaturallySpeaking's speech recognition software were tested, and the Geography Markup Language, GML 2.0, was used to implement an application schema for street condition surveys. Both technical and overall accuracy exceeded 95% over six speech recognition tests in environments that were quiet or constantly loud. However, for three tests while walking along a busy road during which the noise level varied, the accuracy of the speech recognition plummeted to 58%. A "standing" test for capturing the position of defects (N=30) gave an error of 0.41m at the 95% confidence interval. Finally, a web-based questionnaire was answered by 80 GIS project managers, who indicated that they are unhappy with the quality of their data, although they do not require the data in real-time.

1. INTRODUCTION

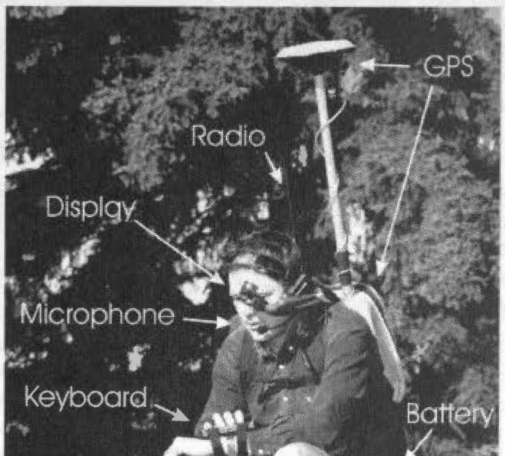
This paper investigates the use of a data acquisition tool that has been designed to minimise both positional and attribute errors. A key objective of this work has been to provide the field operator with a tool that allows them to take ownership of the whole data acquisition process, thus improving workflow by removing duplicated effort and the need to communicate difficulties to office based colleagues. The fundamental research question is whether a Mobile GIS, that includes speech recognition and wireless connectivity for real time access to spatial data, is a viable tool for data acquisition?

The mobile GIS tool has been designed around a wearable computer, which utilizes a multi-modal interface. Speech recognition (SR) and Text to Speech (TTS) technologies have been combined with the traditional keyboard, mouse, and visual display unit interface in an effort to create a hands-free computing environment. Wireless communication has been incorporated to provide increased mobility by allowing the GIS to be free from location constraints such as physical network connections (see Figure 1).

Traditionally, due to the high cost of field acquisition, the capture of geographic information (GI) for inclusion in a Geographic Information System (GIS) has been undertaken using techniques such as digitization or scanning of paper maps. However, the scale of the map, the data collection and manipulation techniques utilized, the symbolization employed by the cartographer, and the medium upon which the map is produced all determine the accuracy of any measurement, and therefore the accuracy of any information derived from the measurements. Data accuracy is further compounded by the need to produce clear maps. This often requires that the cartographer rectify, enlarge or move elements on a map in order to clarify a situation. This practice leads to local differences in scale, rotation and translation (de Knecht *et al*, 2001). The result is that the positional quality of data digitized from maps can be significantly more than map accuracy requirements promoted by organizations such as the US Geological Survey.

Personal experience also indicates that GI users are inclined to believe that the accuracy of their spatial data is better than it really is. For example, while working for Truebridge Callender Beach Ltd. capturing water facilities for Wellington City, New Zealand, it was found that the digitized positions of features with respect to positions obtained from the City Mapping Database had a positional accuracy of 2.38m at a 99% Confidence Interval (CI) for Johnsonville, a suburb of Wellington City (Truebridge Callender Beach Ltd., 1999). The Council required an accuracy of 0.15m (the accuracy of the City Mapping Database) at the 99% CI, and refused to acknowledge that the plans provided for conversion could not achieve their accuracy requirements.

This paper commences with a synopsis of the mobile computing paradigm followed by an examination of SR. Development of a Geography Markup Language (GML) Schema is discussed as a vehicle for the storage and transport of GI between field and office, followed by a review of current GI requirements assembled from a Data Accuracy Requirements Survey undertaken for this research. The paper concludes with initial test results of the mobile GIS under development and a summary of future research activities.



2. MOBILE COMPUTING AND GIS

Mobile computing is the use of computing devices at some distance from the normal, fixed workplace that enables the mobile worker to: create, access, process, store, and communicate information without being constrained to a single location (Zimmerman, 1999). It can be viewed as a combination of three important and related properties: computation, communication and mobility. Computation includes the computing devices at either end of the network; communication systems include the different wireless and wired networks that link the computing devices; and mobility is an aspect of user behaviour (Liu *et al*, 1995).

At the heart of mobile computing lies a need to deliver intelligence to the field to improve productivity and provide a competitive edge in the marketplace. In order for a mobile GIS to be successful it must emulate existing field practices and eliminate repetitive time-consuming tasks.

The mobile workforce often has little computing experience (Wilson, 2000), as such a mobile application should help a mobile user to automate their entire workflow and improve their efficiency. However, a mobile GIS application should also support primary functions such as mapping and navigation (zoom and pan), data collection, query, update and transmission, remote data and component access (wireless), location determination (GPS), coordinate transformations. But the most important requirement is that the application is able to work in the same environment as that of a field user.

2.1 Wearable Computers

Conventional mobile devices include Personal Digital Assistants (PDA's). While cost and convenience make devices such as a PDA attractive, limited disk space, memory and battery capacity can impose considerable restrictions on mobile applications. An alternative is the wearable computer, whose most distinguishing feature is that it can be used while moving around (Rhodes, 1997). They emphasize hands-free computing, and concentrate on speech input and heads-up displays. Wearable computers are designed to be usable at any time with the minimum amount of cost or distraction from the wearer's primary task, i.e., work or recreation.

3. SPEECH RECOGNITION

Speech recognition has long been recognized as an appealing computer input mode (Licklider, 1960; Murray, Jones, and Frankish, 1996; Tyfa, Howes, 2000). It is generally accepted that speech recognition can play a useful role in hands-busy, eyes-busy environments such as field acquisition of GI.

Speech recognition transforms human voice input into commands and characters using one of two modes; command-and-control (discrete SR) and dictation. Typically, discrete SR is used to navigate an application. For example, a user speaks the command "Launch [Application]" to launch the application. The user is required to speak discrete words or short phrases that imitate an application menu. Dictation is typically used for word processing based activities and requires a considerably larger vocabulary.

To perform SR, an application requires an audio source object (a microphone) to accept speech input, a SR engine that resolves speech input into text or commands, and a list of speech commands, a grammar, supplied to the SR engine.

3.1 Vocabularies and Templates

The words that a discrete SR engine resolves are its vocabulary. A vocabulary consists of one or more templates for each word in the vocabulary. In SR, a template is a pattern that can be used to identify or be matched to a speaker's pronunciation of a word. When a SR engine tries to resolve a word, it compares the audio input stream to its templates until it finds a match, or determines no match is available. If a command is recognized the engine then notifies the associated application.

Templates typically fall into two categories, Speaker-independent and Speaker-dependent. Speaker independent templates combine multiple pronunciations of a word and are intended to require no user training. Speaker dependant templates contain single words or phrases and require the user to train the speech engine so that their particular pronunciation can be recognized.

Many speaker-independent applications are field-driven. That is the application will "anticipate" certain vocabulary only when a particular form or field has been activated. In a street condition survey application, for example, saying the word "defect" could activate the Defect vocabulary, for which there may only be a limited number of acceptable inputs; "distortion", "ravelling" and "cracking", for instance. By limiting the number of acceptable values for this field, accuracy and computing efficiency are maximized. By utilizing unique commands, these inputs are not acceptable in any other field, minimizing the risk that they will be inserted in the wrong place by accident.

3.2 Grammar

A grammar is a subset of the available vocabulary that contains only the words used by the application. A grammar is normally created to avoid comparing a spoken command to all the words in a speech engines vocabulary. A grammar should contain only those words that the application requires. A single application may utilize more than one grammar. As a user moves from one application context to another, the active grammar changes to the commands required for the new context. Managing grammars efficiently makes speech-enabled applications easier to use and speech recognition more accurate.

For speech recognition to work effectively, there are a few guidelines that should be adhered to when developing a grammar. Firstly, choose grammars that are consistent with other applications. Use intuitive words for commands that identify the tasks that the application performs, or the objects that are being captured. Speech menus should be short and should use words that are easy to remember. Avoid using commands that sound alike, and use vocabulary included with speech applications when possible.

3.3 Speech Algorithms: The Basis of Speech Technology

There are many words that sound alike but have different spellings and meanings. This presents challenges for speech recognition software vendors and speech-enabled application developers.

There are two basic approaches to deciphering spoken commands. One approach uses matching algorithms that compare bit patterns of the spoken commands to standard bit patterns stored in the SR software's library of patterns. The commands and related bit patterns are matched for associated actions. The library of bit patterns is created by averaging the patterns of a large multicultural sampling of pronunciations of a finite vocabulary.

The second approach relies on speech patterns that are created by users "training" the speech software. The SR software creates patterns based on the user's pronunciation of the phonemes*. This is the approach employed for this research. Spoken words are then matched against patterns in the libraries and those created during the user's training of the software.

3.3.1 How Speech Recognition Works

Speech recognition consists of three broad processes, being the transformation of digital audio from a sound card into a better acoustic representation; determination of the phonemes that have been spoken; and matching the phonemes with a grammar that has been supplied to the SR engine so as to establish the spoken words.

Firstly the Pulse Code Modulation (PCM) digital audio produced by the computer's sound card is transformed into the frequency domain using a band limited Fast-Fourier Transform (FFT). This allows the frequency components of a sound to be identified. A sound card typically samples an audio stream 16,000 times per second. In order to reduce processing time the FFT samples the PCM audio stream every 1/100th of a second. From this sampling process a graph of the amplitudes of each frequency component is created to describe the sound heard for that 1/100th of a second.

The SR engine contains a database of graphs that identify different types of sounds the human voice can make. The graph produced by the FFT is then "identified" by matching it to the closest entry in the database.

Background noise and speaker variation problems are minimized via the use of statistical modelling to determine which phoneme was spoken. Statistical analysis is possible because a phoneme often lasts for 50 to 100 frequency components, and it is likely that one or more sounds dominate during that time that can be used to predict the phoneme spoken. Determination of a phoneme by a speech engine works by hypothesizing a number of different states at once. Each state contains a phoneme with a history of previous phonemes. The state with the highest score (statistically most likely state) is used as the final recognized phoneme.

Once all phonemes have been identified they are compared against a dictionary of pronunciations to determine the word that was spoken. However, if a word is not pronounced as described in the dictionary no match will be found, or the recognizer may select an incorrect word.

4. A DATA MODEL FOR THE TRANSFER OF GEOGRAPHIC INFORMATION

A data model is used to describe and represent some aspect of the real world in a computer. The Open GIS Consortium's (OGC) Interoperability program was been used in the development of a data model for testing the mobile GIS under development. This research has used Street Condition surveys as the vehicle for testing the application. The purpose of a Street Condition survey is to locate defects within the street pavement, along curbs and on sidewalks.

4.1 Geography Markup Language

The Geography Markup Language is an eXtensible Markup Language (XML) encoding for the transport and storage of geographic information, including both the spatial and non-spatial properties of geographic features. GML defines an XML Schema syntax that provide an open, vendor-neutral framework for the definition of GI and objects. It supports the description of GI application schemas for specialized domains and enables the creation and maintenance of linked geographic application schemas and data sets, such as a national road network that passes through a city. The municipal organization may develop a schema that describes the type of road features within its domain and import a schema that describes the types of national road features that pass through the city, rather than duplicating road features that have already been described by the national organisation.

GML also supports the storage and transport of application schemas and data sets (Open GIS Consortium, 2001b). GML allows organisations to either store geographic application schemas and information in GML, as the schema can be directly mapped to a database, or to convert from some other storage format on demand and use GML only for schema and data transport. For this research, data was transported using GML and is stored on the server in a database conforming to the GML schema.

GML supports interoperability through the provision of basic geometry tags, a common data model (features/properties), and a mechanism for creating and sharing application schemas. In general terms a schema defines the characteristics of a class of objects: how the data should be marked up. By publishing a schema an organization can enable or enhance the interoperability of their applications.

GML 2.0 defines three base schemas for encoding spatial information. The Geometry schema (geometry.xsd) includes the detailed geometry components. The Feature schema (feature.xsd) defines the general feature-property model (as feature types) and includes common feature properties such as `fid` (a feature identifier), `name` and `description`. The XLink schema (xlink.xsd) provides the attributes that can link different data sets or schemas. The schema documents alone do not provide a schema suitable for constraining data instances; rather, they provide base types and structures, which may be used by an application schema. An application schema declares the actual feature types and property types of interest for a particular domain using components of GML.

4.2 Street Defect Schema Implementation

Figure 2 is a UML diagram for the Road Defect data model. As shown, allowable defect members must be `Road` or `Footpath` instances. The `<<restriction>>` stereotype applied to the generalization relationship indicates that subtypes defined in the schema are derived by restriction from their super type. For instance, a `RoadDefect` can only contain class members that are members of the class `RoadDefectMember`, in this case, `Road` and `Footpath`.

---

* Phonemes are a set of similar speech sounds, which are perceived to be a single distinctive sound in a language.

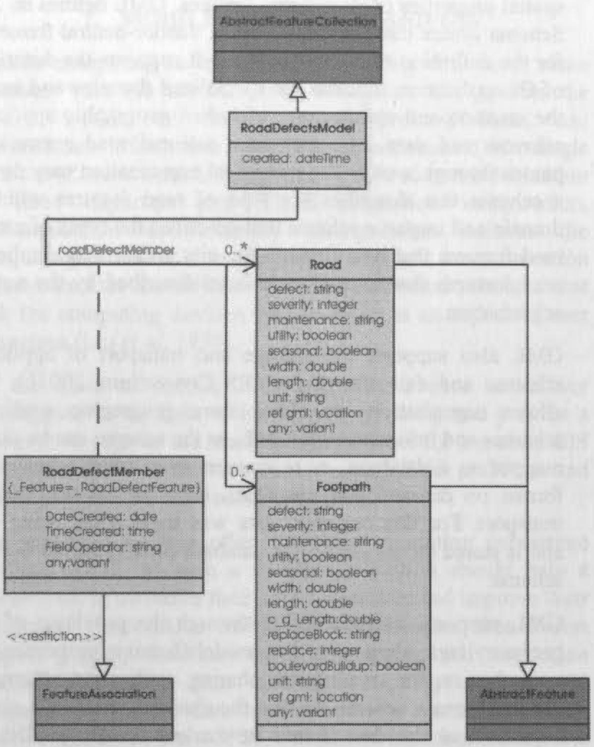

AbstractFeatureCollection
RoadDefectsModel
created: dateTime
roadDefectMember
0..*
Road
defect: string
severity: integer
maintenance: string
utility: boolean
seasonal: boolean
width: double
length: double
unit: string
ref gml: location
any: variant
RoadDefectMember
(_Feature=_RoadDefectFeature)
DateCreated: date
TimeCreated: time
FieldOperator: string
any:variant
0..*
Footpath
defect: string
severity: integer
maintenance: string
utility: boolean
seasonal: boolean
width: double
length: double
c_g_Length:double
replaceBlock: string
replace: integer
boulevardBuildup: boolean
unit: string
ref gml: location
any: variant
<<restriction>>
FeatureAssociation
AbstractFeature

Figure 2: UML Data Model of Defects Schema

Each member class then defines its attributes by extending the base `AbstractFeatureType` class in the GML feature schema.Road Defects Instance

An instance of the data model for this research (see Figure 3) has a single feature collection of type `RoadDefectsModel` and contains one feature using a containment relationship called `DefectMember`. The feature collection has a string property called `<Created>` which holds the date and time that the GML file was created and a geometric property called `<boundedBy>` with a `<box>` value. The box geometry (the 'bounding box' of the feature collection) is expressed in terms of the spatial reference system (SRS) identified by the value of the `srsName` attribute. The `srsName` attribute of the geometry types are those described by the European Petroleum Survey Group (EPSG) as proposed by the OGC (Open GIS Consortium, 2001c).

The feature member is an instance of `RoadType` representing a "Distortion" `<Defect>` of `<Severity>` "2", etc. The member has a `<defectID>` of "1395" and contains other properties describing the date and time the feature was captured and by whom. It has a geometric property called `<location>` with a `<point>` value, which are derived from the base GML schemas.

5. DATA ACCURACY REQUIREMENTS SURVEY

A Data Accuracy Requirements Survey was created and sent out to a number of GIS and governmental bulletin boards such as those managed by GITA, AURISA, URISA, the NZ ESRI User Group and the GISList maintained by the GeoCommunity at the beginning of June, 2001. The purpose of Survey was to determine: the spatial accuracy needs of end users; present data capture methods; "time to use" requirements and the level of validation of captured data. The survey was intended for Utility, Local, Provincial and Federal/National Government GIS project managers.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- File: Defect.xml -->
<RoadDefectsModel
xmlns="http://www.ucalgary.ca/~ahunter/gml"
xmlns:gml="http://www.opengis.net/gml"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation=
"http://www.ucalgary.ca/~ahunter/gml/defects.xsd">
  <gml:boundedBy>
   <gml:Box
srsName="http://www.opengis.net/gml/srs/epsg.xml#26911"
>
    <gml:coord>
     <gml:X>750921.789</gml:X>
     <gml:Y>5322419.521</gml:Y>
    </gml:coord>
    <gml:coord>
     <gml:X>750921.789</gml:X>
     <gml:Y>5322419.521</gml:Y>
    </gml:coord>
   </gml:Box>
  </gml:boundedBy>
  <DefectMember>
   <Road>
    <Defect>Distortion</Defect>
    <Severity>2</Severity>
    <Maintenance>Hot box</Maintenance>
    <Utility>True</Utility>
    <Seasonal>False</Seasonal>
    <Width Units="meters">4.3</Width>
    <Length Units="meters">5.8</Length>
    <gml:location>
     <gml:Point
srsName="http://www.opengis.net/gml/srs/epsg.xml#26911"
>
      <gml:coord>
       <gml:X>750921.789</gml:X>
       <gml:Y>5322419.521</gml:Y>
      </gml:coord>
     </gml:Point>
    </gml:location>
   </Road>
   <DefectID>1395</DefectID>
   <DateCreated>2001-12-13</DateCreated>
   <TimeCreated>08:42:53</TimeCreated>
   <FieldOperator>Andrew Hunter</FieldOperator>
  </DefectMember>
  <Created>2000-12-13T08:45:22</Created>
</RoadDefectsModel>
```

Figure 3: Defect Schema Instance

By mid-October 2001, 80 responses had been received. The respondents were grouped into functional groups: AM/FM, Electrical and Gas (12 respondents), Business GIS (5 respondents), Environment (10 respondents), Local Government (36 respondents), Water Resources (4 respondents), and others (13 respondents).

	Accuracy > 1m	Accuracy < 1m	Significance
Actual Accuracy	34	46	
Practical Accuracy†	45	35	$p = 0.11$
Desired Accuracy	63	17	$p < 0.01$

Table 1: Comparison of Accuracy Requirements in Relationship to Cost of Acquisition

There is no significant difference between the accuracy of data currently being used and end-users' accuracy requirements. They make do with what they have. However, comparing the respondents who require data at an accuracy of less than one metre (with cost as a constraint), to the respondents who would like data with an accuracy of less than one metre if cost was not an issue it is evident that respondents desire more accurate data (refer to Table 1). In other words, the cost of data acquisition significantly affects the accuracy of the data.

---

† Practical accuracy was defined as the accuracy specification considered ideal for an application when cost of data acquisition is a consideration.

Improved data accuracy was most desirable for Local Government organisations, however all domains would like more accurate data if they could afford it. The significance values indicated in the table have been obtained by performing a Chi-squared test between actual and practical accuracy, and actual and desired accuracy.

Respondents were provided a number of options ranging from 1 day or less through to 6 months or more, and were asked to indicate actual time-to-use and preferred time-to-use for their organization. Seventeen respondents are currently using data within one day; the majority (61) are using their data within one month. In terms of preferred time-to-use, 32 respondents would like their data within one day and the majority (68) would like to use their data within 2 weeks. However, while the survey indicates that virtually all users would like data to be obtained rapidly, the most desirable timeframe is somewhere between one week and two weeks (see Table 2).

	Actual Time-to-use	Desired Time-to-use	Significance (p)
$t \leq 1$ day	17	32	0.017
1 day < $t \leq 3$ days	10	9	0.037
3 days < $t \leq 1$ wk	12	23	<0.001
1 wk < $t \leq 2$ wks	5	4	<0.001
2 wks < $t \leq 1$ mth	17	8	0.016
1 mth < $t \leq 2$ mths	9	1	0.042
2 mths < $t \leq 6$ mths	5	1	0.246
$t \geq 6$ months	5	2	

Table 2: Comparison of Time to Use (t) Requirements

If the respondents are grouped into Local Government and "Others" we see that "Others" have a greater preference (p = 0.005) for 1 week or less and Local Government has a greater preference (p = 0.002) for two weeks or less.

Half of the respondents use conflation, of which 53% consider the results acceptable. Only 60% of the respondents do any form of validation, 53% digitize maps in order to acquire data, and 54% are not happy with the quality of their data. To conclude, GI users would like more accurate data, although they do not require the data in real-time. The majority of GI users (79%) continue to use scanning and digitizing for the capture of data, both of which are methodologies that add to the existing spatial error in source data.

6. APPLICATION TESTING

6.1 Speech Recognition Testing

Speech recognition was undertaken by recording spoken commands with a Dictaphone and then comparing them with a log of commands recognised by the speech engine. Agreement is binary, in that the computer recognised command either matches the recorded command or not. In order to obtain a reasonable picture of speech recognition performance, testing was been carried out in three different environments with respect to background noise. The first test was performed in a quiet environment (office) where background noise was minimised. The second test was in an environment where the background noise was relatively loud but constant (tested while driving the car along a motorway). The background noise in the final environment was variable in that there were moments of very low background noise and very high background noise, while walking along a busy street around the University. Quiet periods were observed when there was no traffic; noisy periods occurred when traffic lights turned green or when public transit buses passed. In the quiet environment, the system was also tested on two different computers to determine if computing power affected SR performance.

Table 3 sumarises the results in each of the three environments described above. The results have been characterized into Technical and Overall accuracy. Technical accuracy looks only at the recognition errors made by the speech engine, which where either the misrecognition of a command, or the inability to recognise a command. Overall accuracy includes user errors, which consist of poor enunciation of commands or giving an unacceptable command.

Environment	Numbers of Commands	Technical Accuracy	Overall Accuracy
Quite	507	99.3%	98.2%
Constantly Loud	463	96.8%	95.5%
Variable	234	58.8%	57.3%

Table 3: Speech Recognition Results

The results for the first two categories are acceptable (correct better than 95% of the time); however the third category is not, given that this is the environment in which most data acquisition will be performed. Having said that, when analysing the results more broadly it was often found that the speech engine did recognise a command correctly when in this environment, but it took an intolerable amount of time to do so. With Internet based applications a time of 5 to 8 seconds is typically used as the maximum criteria that a user must wait for a response. This is the same criteria used for testing in this research. While waiting for the recognition of a command, additional commands were often issued, hence the very poor result. What is actually happening is that the speech engine has to process all the sounds that it hears; if traffic is busy the computer hears this and tries to make sense of it. Speech recognition is extremely processor-intensive, so in times of high background noise it was found that it could take several minutes before the speech engine actually caught up and recognised a valid command.

The reason that speech recognition worked in a constantly noisy environment is that the speech engine can sample the background noise at the beginning of an exercise and then attempt to remove this from everything that it hears. As is evidenced by these results the major source of error is the microphone. The wearable computer comes with its own specialized directional microphone, but it is clearly not adequate for working in variable noise environments.

There is also an apparent reduction in ability to remember pieces of information, numbers for example, that are otherwise possible in a normal environment minus the computer. This does point to the question of user safety. Does talking to a computer interfere with your ability to recognise what is happening around you? Studies such as those performed by Strayer *et al*, 2001, have provided evidence which shows that those engaged in phone conversations miss twice as many simulated traffic signals as when they were not talking on the phone, and that they took longer to react to the signals that they did detect. Is talking on the phone any different than talking to a computer in terms of user safety?

6.2 Positional Accuracy

Various modes of transportation will be used while testing positional accuracy in order to match the application to different data acquisition requirements. Four methods are to be analysed: standing on a feature to be captured, and capturing features while walking, riding a bike or driving a car over them. To date, only the initial phase has been completed. The "standing" test involves walking up to a local survey control monument, standing on it, observing a differential GPS position, and then moving away. This process was repeated sufficiently (30 times) to ensure a statistically reliable result. The mean error for the standing test is 0.25m with a standard deviation of 0.08m. This equates to a positional error of ±0.41m at a Confidence Interval of 95%.

7. CONCLUDING REMARKS

One of the primary tasks of this research has been to look at an alternative method of data acquisition, the objectives being to improve spatial accuracy, improve attribute accuracy, minimised acquisition timeframes, and removal of intermediate processes typically required to get GI from the field into an application.

Speech recognition plays a major role in this application in terms of system architecture and user interaction. Results indicate that the current system does not respond as instructed, nor in a timeframe that is acceptable when working in a typical environment. However, if the background noise issue can be resolved it is apparent that this mode of interaction with a computer shows immense promise. When in a quiet environment, talking to the computer is fast, and accuracy is acceptable. Speech recognition's ability to interact with the user via text to speech capabilities (play back to the user what it has heard) provides an opportunity to verify captured data on the fly, rather than having somebody else verify it. If head mounted displays are to be utilised, speech recognition is a necessity. Without it, interaction with the computer via mouse and miniature keyboard is difficult.

Positional accuracy tests completed to date would indicate that the system developed can meet or better the accuracy requirements of most GI users' who desire accurate (sub-metre) spatial data. Further testing will provide an indication of system accuracy while on the move.

Given the difficulties encountered with speech recognition and in light of the survey undertaken as part of this research, we ask whether it really is necessary to deliver data in real-time to a local government? The study would indicate that, as yet, there is not a great need for instant geographic information.

Further research is required to investigate alternative methods of capturing spoken commands issued by the user. Alternatives include throat and skull microphones, which record sound directly from vocal cord vibrations, rather than through the air. Significant cognitive based testing is also required to ensure that the system is simple and intuitive to use. Lastly, with regards to the wireless/Internet component further investigation with respect to reliability, data transfer capabilities and latency are warranted.

8. BIBLIOGRAPHY

de Knecht, J. P., Schavemaker, J. G. M., Reinders, M. J. T., Vossepoel, A. M., (2001) Utility Map Reconstruction, International *Journal of Geographical Information Science*, Vol. 15(1), pp. 7-26

Licklider, J.C.R., (1960) Man-Computer Symbiosis, *IRE Transactions of Human Factors in Electronics*, Vol. 1, pp. 4-11

Liu, G., Marlevi, A., Maguire, G. Q. Jr., (1995) A Mobile Virtual Distributed System Architecture for Supporting Wireless Mobile Computing and Communications *Proceedings of the First Annual International Conference on Mobile Computing and Networking (MOBICOM '95)*, Berkeley, California, USA, pp. 111-118

MobileInfo.com, Mobile Computing Hardware & Software Components, *MobileInfo Hardware & Software Components*, http://www.mobileinfo.com/hwsw_components.htm, (accessed May 15, 2001)

Murray, A. C., Jones, D. M., Frankish, C. R., (1996) Dialog Design in Speech-mediated Data-entry: The Role of Syntactic Constraints and Feedback, *International Journal of Human-Computer Studies*, Vol. 45, pp. 263-286

United States National Map Accuracy Standards - Version 6/1947 http://rockyweb.cr.usgs.gov/nmpstds/nmas647.html, (accessed 6/2/02)

Open GIS Consortium (2001a) OGC Document Number: 01-029, Geography Markup Language (GML) 2.0, *OpenGIS® Implementation Specification, 20 February 2001*, http://www.opengis.net/gml/01-029/GML2.html, (accessed 8 July, 2001)

Open GIS Consortium (2001b) *The Open GIS™ Abstract Specification Topic 1: Feature Geometry (ISO 19107 Spatial Schema) Version 5* http://www.opengis.org/techno/abstract/01-101.pdf, (accessed 8 July, 2001)

Open GIS Consortium (2001c) *The Open GIS™ Abstract Specification Topic 2: Spatial Reference Systems Version 4* http://www.opengis.org/techno/abstract/99-102r1.pdf, (accessed 15 July, 2001)

Rhodes, B. J., (1997) The Wearable Remembrance Agent: A System for Augmented Memory, *Personal Technologies*, Issue 1, pp. 218-224

Strayer, D., Drews, F., Albert, R. and Johnston, W. (2001) *Does Cell Phone Conversation Impair Driving Performance?* National Safety Council | Cellphone Article, http://www.nsc.org/library/shelf/inincell.htm (accessed 15 July, 2001)

Trubridge Callender Beach Ltd. (1999) *Data Characteristics: Data Conversion of Water Network Record* – Johnsonville Portion Report to Wellington City Council, New Zealand

Tyfa, D. A., Howes, M., (2000) Speech Recognition for Command Entry in Multimodal Interaction, *International Journal of Human-Computer Studies*, Vol. 52, pp. 637-667

Wilson, J. D., (2000). Mobile Technology takes GIS to the Field, *GEOWorld*, Vol. 13(6), pp. 32-36

Zimmerman, J. B., (April 1999) Mobile Computing: Characteristics, Business Benefits, and the Mobile Framework, *Mobile Computing: Characteristics, Business Benefits, and the Mobile Framework*, http://faculty.ed.umuc.edu/~meinkej/inss690/zimmerman/INSS 690 CC - Mobile Computing.htm (accessed May 24, 2001)

THE DESIGN AND IMPLEMENTATION OF A NOVELL OBJECT RECOGNITION METHOD BASED ON INFORMATION FUSION AND ARTIFICIAL INTELLIGENT I CONCEPT

[a] Farhad Samadzadegan, [a] Ali Azizi, [b] Micheal Hahn

[a] Department of Geomatics, Faculty of Engineering, University of Tehran, Tehran, IRAN
[b] Applied University of Stuttgart, Stuttgart, Germany
samadz@chamran.ut.ac.ir, aazizi@ut.ac.ir

Commission IV, WG IV/7

KEY WORDS: Object Extraction and Recognition, Fuzzy Logic, Neuro-Fuzzy, DSM, Morphological Operators, Finite Elements, Matching

ABSTRACT:

The idea of having a fully automatic three dimensional object recognition (OR) system to replace the human operator has been one of the main aspirations and the final goal of many photogrammetric and computer vision investigators. This would reduce the need for the human experts for object recognition and extraction and hence decreases the production time and leads to a more economical knowledge-based information extraction. Nevertheless, due to the tremendous complications and complexities associated with the natural objects appearing in digital images, fully automatic object recognition have faced serious obstacles and, thus, only in a relatively simple imaging environment or controlled laboratory conditions a reliable result is normally expected.

Regarding the human potentials in object perception and recognition, it seems that a comprehensive OR system should be able to integrate the following capabilities: (a) all available descriptive information components of an object (such as: structure, textural information and spectral responses) must be simultaneously taken into account, (b) a fuzzy formulation for the object definition, regarding its complexities, should be devised and (c) learning capabilities to modify the defects accompanied by the objects definition needs to be considered. This will enhance the recognition potentials when encountering new and undefined objects.

In this paper, for the first time, an attempt has been made to design a system that integrates all above features in a total and comprehensive solution. The approach presented here takes advantage of the concept of fusion in two levels of descriptive information and logical planes. That is, information fusion to exploit the multi-level characteristics of the objects and logic fusion for enhancing the learning and hence recognition abilities of the system. The designed OR process comprises the following procedures:

Pre-processing

To facilitate the OR operations, a set of preprocessing modules are initially applied to the input data. These are basically fundamental radiometric and geometric corrections such as the gray scale filtering, the determination of the sensor attitude and altitude parameters, epipolar re-sampling and the generation of the image pyramids.

Object definitions

This stage starts with the definitions of the features (objects) and their corresponding classes according to the required map scale. It has been assumed that the objects can be described in an integral manner by three attributes of structural, textural and spectral (STS) components. Structural information express the 3D geometric characteristics of the objects surface. This parameter is indeed an efficient mechanism by which a quick recognition of many 3D objects can be conducted without further involvement with the textural complications. Nevertheless, the structural information alone, by no means, gives an exhaustive description of the objects. The textural information and spectral response attributes of the objects also play a decisive rule.

Rules generating engine

Based on the object and class definitions in the previous stage, linguistic variables, labels and membership functions are built for STS attributes of each object. These object descriptive parameters are made closer to the subjective perceptions of a human operator by an expert with a long experience in photogrammetric map compilation. Based on the previously defined descriptive concepts, fuzzy rules are generated and the logical structures of the system are constructed.

Training engine

Due to the inevitable complications and ambiguities of the objects, in some conditions it is not possible to define the membership functions in a perfect manner. These functions will, thus, can be modified by the learning capabilities of the neuro-fuzzy system which adapts new descriptive information and re-adjusts the parameters of reasoning system.

Recognition engine

Based on the STS components extracted in stage "b", a preliminary analysis is carried out to generate the following discriminative information: for the 3D structural information, dimension; height; and shapes are analyzed and quantified. The shape analysis is performed according to the dimensional compatibility of the three elements of the length, width and height. Textures are also quantified by a texture analysis approach by means of a mean curvature computation. If the extracted regions in image space do not demonstrate a homogenous texture (e.g. a planar facet having a rugged texture and/or with a totally unexpected spectral response), the 3D structure will be divided into two or more objects according to the case. The full discriminative information accompanied by their corresponding spectral responses are, then, simultaneously fed into the fuzzy recognition module which utilizes the predefined rules to identify and recognize the objects. If the object recognition process for a feature is failed or if an object is flagged as unrecognizable, it may be transferred into the training engine by which either a new object is introduced or a better membership function is determined.

The proposed OR approach is implemented as a software package entitled 'StereoMetrica'. This paper presents a detailed analysis of the individual components that govern the entire system. To verify the implemented approach a pair of large scale aerial photographs of an urban area of Stuttgart, Germany was selected. A detailed report of the tests conducted in this area is presented and discussed in the final section of this paper.

EVALUATION OF THE POTENTIAL OF THE HIGH RESOLUTION SATELLITE IMAGERIES (IKONOS) FOR LARGE SCALE MAP REVISION

[a]Farhad Samadzadegan, [a]Ali Azizi, [b]Mohammad Sarpoulaki, [c]Ahmad Talebzadeh

[a] Departmen of Geomatics Engineering, Faculty of engineering, Tehran University,
[b] National Cartographic Center
[c] Iranian remote Sensing Center
e-mail addresses: [1]samadz@chamran.ut.ac.ir , [1]aazizi@ut.ac.ir, [2]sarpulki@ncc.neda.net.ir

Commission IV, WG IV/7

KEY WORDS: High Resolution Satellite Images, Map Revision, Mathematical Model, Accuracy, IKONOS

ABSTRACT:

High and accelerating rates of the urban changes and the urban area growth especially in developing countries like Iran with more than 1000 cities calls for efficient and fast techniques for map production and revision with the required accuracy. The availability of the new generation of the commercial one-meter resolution satellite imageries have opened a new era heralding a promising future for producing and updating large scale digital maps. Although several investigations have already began in this area, still the subject demands more comprehensive tests for the determination of the suitability of such high resolution images for large scale map revisions. In this study the mapping potentials of the IKONOS images have been thoroughly examined in a test site in Iran. The test is conducted based on the following three-stage experimentation.

The first stage is concerned with the 2D and 3D geometric accuracy assessment of the IKONOS images over the test site. A wide range of mathematical models for geo-referencing is adapted and implemented. For the 2D geometric evaluation these were multi quadratic functions, projective and polynomial transformations. 3D geometric accuracy assessment is also investigated using Direct Linear Transformation, as well as rational and rigorous functional models. Since the stereo images for the IKONOS data and the corresponding orbital parameters were not available, a large scale DTM of the test area was used to compensate for the relief displacements. In order to make the geometric evaluation as reliable as possible, a test site (city of Rasht, in Iran) was selected for which 200 ground control points was available. The coordinates of these points were measured with differential GPS with an accuracy in the order of 10 to 15 cm. These control points were well distributed over the test area with an overall distribution density of 2 to 3 ground control points in each square km. Moreover, photogrammetricaly generated 1:1000 and 1:10000 scales digital maps (acquired from 1:5000 and 1:40000 scale aerial photographs respectively) and the 3D city model and the corresponding DTM covering nearly 80 squares km was also available. The geometric accuracy assessments were made based on the residual errors calculated for ground control points as well as uniformly spaced checkpoints. The test results indicate that the multi quadratic function offers a better image distortion modeling for flat areas provided that large number and well-distributed ground control points are available. One of the drawbacks of the multi quadratic functions was the fact that accuracy assessment could not be studied on ground control points and hence existence of checkpoints was inevitable. The rational functions on the other hand indicated a better performance for non- flat areas. The rational functions also demonstrated more flexibility in terms of the number and distribution of the ground control points.

In the second stage of the project the information content and the sharpness of the features of the image in comparison with the standard aerial photographs were evaluated by visual inspection. This visual inspection was quantified using frequency domain analysis approach. In this context the possibility of improving the quality of radiometric information using image restoration techniques such as de-blurring operations were also investigated. To perform this stage, it was necessary to estimate the imaging system's point spread function (psf). This was evaluated using the smallest detectable symmetric feature with a uniform grey scale background that could be found in the image. The 2-sigma of the psf was evaluated assuming a Gaussian model for it. The pictorial restoration operations were then performed by modified inverse and Wiener filters respectively. The test results indicated a limited marginal radiometric improvement.

In the third stage the feature tracing and pointing capabilities of IKONOS image was compared with the feature tracing and pointing potentials of standard aerial photographs for the same test area. For further assessment of the feature tracing potentials of IKONOS images some linear features such as non-occluded road and building edges were traced and digitized on geo-referenced images and the result was compared with the same features extracted from 1:1000 and 1:10000 scale digital maps and the corresponding city model. In the final stage of this study, based on the preceding metric and pictorial evaluations, the suitability of the images for large scale map revision were assessed. Based on the overall evaluation of this study it may be concluded that map revision for the scale of 1:10000 is feasible with IKONOS images.

APPLICATION OF HIGH-RESOLUTION SATELLITE IMAGES IN FORESTRY AND HABITAT MAPPING - EVALUATION OF IKONOS IMAGES THROUGH A HUNGARIAN CASE STUDY

Kristóf, D.[a], Csató, É.[b], Ritter, D.[a]

[a]Institute of Environmental Management – Szent István University
Páter Károly u. 1., Gödöllő, Hungary, H-2103 Tel: +36 28 522 000/ext.1057 Fax: +36 28 415 383
e-mail: kdan@nt.ktg.gau.hu

[b]Remote Sensing Centre – Institute of Geodesy, Cartography and Remote Sensing
Bosnyák tér 5., Budapest, Hungary, H-1149 Tel: +36 1 222 5101/ext.110 Fax: +36 1 222 5112
e-mail: e.csato@rsc.fomi.hu

KEY WORDS: forestry, mapping, thematic data extraction, geometric correction

ABSTRACT:

The launch of Ikonos satellite at the end of September 1999 marked the beginning of a new era for the application of satellite images. Being able to provide up to one-metre resolution panchromatic and four-metre resolution multispectral images, thus available in the past for military purposes or from aerial surveys only, this satellite offers considerable improvements and broadens the application possibilities of satellite imagery. In our study, two of the lower-price "Ikonos Carterra Geo" images were processed. Geometric accuracy enhancement and thematic data extraction possibilities were examined in order to evaluate this sensor for forestry applications. The whole study focuses on a selected Hungarian research area, but the results can be useful in a more general context.

1 INTRODUCTION

1.1 Challenge

In Hungary, large-scale forest management is primarily based on special geographical information systems, with "traditional" airborne orthophotos as a base map. The characteristics of Ikonos make it an interesting sensor for large-scale forestry and environmental applications. Temporal resolution of image databases can be largely increased, due to the known revisit time and pointing capabilities of the satellite platform, which facilitates large-scale change-detection and monitoring of selected areas in order to keep GIS databases up-to-date. However, the practical application of such images for the above-mentioned purposes requires a geometrical accuracy comparable to airborne orthophotos. As the satellite is owned by a profit-oriented private company, image prices are relatively high, and the images are priced in function of the geometric accuracy. Being extremely price-sensitive, sectors such as forestry won't be able to purchase the geometrically most precise images, although their work requires this precision.

Furthermore, the above-mentioned applications require efficient thematic data extraction methods. Compared to airborne photography, high-resolution satellite imagery has the advance of being meaningful from a radiometric point of view, therefore not only qualitative but quantitative analysis can be carried out. As the image resolution increases, large-scale mapping becomes possible, but digital image interpretation such as thematic data extraction becomes harder and harder, requiring new and more efficient methods.

1.2 Aims

As a consequence of the above-mentioned facts, the first aim is to *improve the geometrical accuracy* of a lower-price image. This goal doesn't seem to be easily acquired, since the sensor and orbit parameters are kept in strict confidence by Space Imaging and therefore aren't available to the public. Arising from the above-mentioned facts, multiple methods, including polynomial, *rational polynomial* and *satellite orbital modeling* calculations were tested and compared in order to find the method that provides the highest geometrical accuracy.

The *extraction of thematic information* is always crucial for forestry and environmental applications. In the second part of our paper, the efficiency of *visual interpretation*, *spectral* and *texture-based* classification and *image segmentation* methods are compared. *Data fusion* with other raster- and vector-based databases was also examined.

2 DATA

2.1 Study Area

As this study deals with forestry applications of Ikonos images, a study area characterized by important forestry activities had to be selected. The area of the town Zirc and surroundings was chosen, where more than 40% of the surface is covered by woodland, which makes it one of the most forested regions of Hungary. The study area lies in the western part of Hungary, and covers about 192 square kilometres, between the latitudes going from 47°19' to 47°13' North and longitudes going from 17°45' to 17°58' East. The area is a rolling, hilly terrain, characterized by hills and valleys with elevations going from 200 to 680 meters above sea level, so the maximal variations reach about 500 meters.

2.2 Ikonos images

Two images, covering together the study area, were purchased to carry out the research. These separate, partially overlapping images can be identified by the product order number: the western, larger image has the number 69566, and the number of the smaller eastern image is 69567. *Table 1* shows the most important characteristics of the two images.

Product order number	69566	69567
Product level	Ikonos Carterra Geo 1 m (Pan-sharpened)	Ikonos Carterra Geo 1 m (Pan-sharpened)
Sensor	IKONOS-2	IKONOS-2
Date and time of acquisition	02 may 2001. 10:00 am	02 may 2001. 9:59 am
Raw image pixel size • Cross scan • Along scan	 0.92 m 0.87 m	 0.94 m 1.01 m
Nominal Collection Azimuth	262.46°	335.81°
Nominal Collection Elevation	70.66°	62.58°
Sun Angle Azimuth	159.48°	159.38°
Sun Angle Elevation	56.99°	56.98°
Image product pixel size • Cross scan • Along scan	 1.00 m 1.00 m	 1.00 m 1.00 m
Projection	UTM north zone 33	UTM north zone 33
Ellipsoid	WGS 84	WGS 84
Upper left corner (UTM, meters)	X: 707899.86 Y: 5244719.30	X: 718426.72 Y: 5244867.12
Columns (pixels)	12704	6104
Rows (pixels)	12672	12436
Bit/pixel/band	11 (stored on 16)	11 (stored on 16)
Number of bands	4 (R,G,B,NIR)	4 (R,G,B,NIR)
File format	GeoTIFF	GeoTIFF

Table 1. *Most important characteristics of the two Ikonos images (from the product metadata file).*

2.3 Topographic maps

To carry out geometric accuracy examinations and enhancements, topographic maps of scale 1 : 10,000 covering the totality of the study area were purchased. The maps have been scanned and georeferenced in order to obtain a highly accurate reference database. *Table 2* shows relevant data about the topographic map-based reference database.

Number of map sheets	12
Scanned pixel size on ground	0.8 m
Nominal horizontal accuracy of paper maps	3 m
Nominal vertical accuracy of contour lines	0.7 – 0.8 m
Projection	EOV (Hungarian Uniform National Projection System)

Table 2. *Characteristics of the digital 1 : 10,000 topographic map database.*

Some features, e.g. important and well-visible roads, railways were digitized using the topographic map raster files. These vectorized line features were used for visual geometric accuracy examination of Ikonos images.

2.4 Digital Elevation Model (DEM)

Digital elevation model of the entire study area is required to examine geometric accuracy tendencies depending on surface relief and also for geometric accuracy enhancement (Davis and Wang 2001).

Digitized contour lines of the above-mentioned 1 : 10,000 topographic map database have been used to calculate DEM of the study area.

The topographic maps contain one contour line on every 2.5 meters, with the values of the current elevation above sea level. After digitizing all of the contour lines of the entire geocoded topographic map database, line coverage has been created containing every contour line and the appropriate elevation value. Linear rubber sheeting interpolation method was used to produce a raster DEM. The resulting DEM has a pixel size of 2.5 m, and contains interpolated elevation data for each pixels (or grid cells) of the entire research area.

2.5 Digital Forestry Map Database

To test, evaluate and develop data extraction techniques, digital forestry map database was purchased from the National Forest Service.

The forestry map database was shipped in Arc/Info Export format. This vector database, covering most of the woodland area of the research site, is composed of 2534 polygons altogether, with each polygon containing one forest subcompartment. The database linked to each polygon contains over 70 attributes describing climatic, water and soil properties, tree species, mixture rate, closure and many other descriptive and administrative properties.

2.6 Other data sources

Although this study focuses on the application of Ikonos images, it is also important to make a comparison between this new high-resolution satellite image product and other available remote sensing data. Therefore, other images, including Landsat 5 TM, SPOT-4 HRVIR satellite images, true-colour and false-colour infrared aerial photographs have also been investigated during our work.

2.7 Software used

Geometric accuracy investigations, polynomial geometric corrections and classifications were carried out using ERDAS Imagine 8.4 Professional. PCI OrthoEngine 8.2 Satellite Edition with Ikonos Model was used to carry out geometric corrections based on rational polynomials and satellite orbital modeling. For image segmentation, Definiens eCognition 2.0 software was used. Digital forestry map database operations were carried out using ESRI ArcView GIS 3.1.

3 METHODOLOGY

3.1 Image geometry evaluation and enhancement

Both geometric accuracy and price of Ikonos images depend on product level. In our study, lowest-price Ikonos Carterra Geo images were purchased and post-processed in order to make practical investigations on geometric accuracy and to improve cost-efficiency.

Raw Ikonos images can contain errors due to

- angular tilt
- the Earth curvature
- atmospheric refraction
- surface topography.

(Guienko, 2001.)

The image supplier offers multiple product levels, with different levels of geometric correction. It is not possible to order raw images. During pre-processing of Ikonos Geo images, errors are mostly eliminated, except those due to surface topography.

According to Guienko (2001), Ikonos images acquired with 26 degrees cross track pointing view angle can contain errors up to 243,8 pixels for areas with surface topography variations of 500 m.

Space Imaging claims that all Ikonos Carterra Geo images have the following geometric accuracy, *not including topographic displacements* (*Table 3*):

Circular Error for the 90% of points (CE90)	50.0 m
Mean RMS error	23.6 m

Table 3. *Geometric errors of Ikonos Carterra Geo images.*

It is important to note here that Space Imaging doesn't provide any information about the Ikonos satellite sensor nor orbit parameters. It is therefore impossible to carry out normal orthorectification on Ikonos images.

3.1.1 Evaluation of geometric accuracy

Actual geometric accuracy of both Ikonos Geo images was measured by comparing control points selected on both the reference map database and each of the images.

The images were reprojected from UTM/WGS84 projection to EOV (Hungarian Uniform National Projection System) using the ERDAS Imagine EOV definition add-on module developed at the Institute of Geodesy, Cartography and Remote Sensing. Rigorous transformation with bicubic resampling was used in order to obtain maximal reprojection accuracy. The maximal errors of the reprojection process using rigorous transformation don't exceed some decimetres.

After reprojection, the topographic map database and the images became directly comparable. Check point pairs were selected manually on both of the images and the topographic map reference database. All the points were carefully selected and positioned on surface elements that can be clearly identified both on the satellite images and the topographic maps. Using digital elevation model of the area, points were taken at multiple different heights to examine the effect of surface topography on the errors of original Ikonos Geo images.

On the larger left image (69566), 38 points, on the right image (69567), 16 points were selected. Most of these points were also used as ground control points in geometric correction processes. Difference of meter-based EOV coordinates measured on the reprojected original images and on the reference map database were used to calculate X and Y errors for each point. RMS errors were calculated using a simple distance equation.

Visual examination of geometric errors was also performed using line features (roads, railways) vectorized from the 1 : 10,000 topographic maps.

3.1.2 Geometric accuracy enhancement

As the Ikonos Geo images were found not to be accurate enough for large-scale (1 : 10,000) mapping, multiple methods were tested to improve geometric accuracy.

Since all of the tested methods require highly accurate reference data, carefully selected ground control points were used. These thoroughly tested points were already used for geometric accuracy evaluation.

Original georeferenced Ikonos Geo images were used for GCP collection. After importing the original GeoTIFF files, image map model was deleted, and raw image (pixel) coordinates were stored for each GCP as input coordinates.

The same ground control points were identified with high accuracy on the reference map database. EOV X and Y coordinates were stored as reference coordinates.

During image resampling, cubic convolution resampling method was used in each case to ensure good quality results.

3.1.2.1 Simple polynomial method

This method doesn't require sensor nor topographic information, so the digital elevation model couldn't be used. This simple but method is generally used to rectify images of smaller areas with relatively flat terrain. Distortions that can be described or approximated using polynomial functions can be eliminated using this method.

Simple 2nd, 3rd and 4th order polynomials were used for image transformation. For both of the images, RMS errors for all of the GCPs were calculated. GCP input coordinates were raw image pixel coordinates (column/row number), reference coordinates were given in EOV map projection system.

3.1.2.2 Rational polynomial method

Rational function model is essentially a generic form of polynomials. It defines the formulation between a ground point and the corresponding image point as ratios of polynomials. (Tao and Hu 2001). As this method uses Z coordinates (elevation), elevation data for all control points is needed. Rational function model doesn't require sensor information. According to Toutin and Cheng (2001), this method can be useful for areas with gentle terrain.

Rational polynomial function (RPF) rectification process was carried out on both of the images using the same GCPs as for geometric accuracy evaluation and for every other methods. GCP input coordinates were raw image pixel coordinates (column/row number), reference coordinates were given in EOV map projection system. Elevation data was for each point was given using digital elevation model. RMS errors for all of the GCPs were calculated for comparison.

3.1.2.3 Satellite orbital modeling

Satellite orbital modeling is an implementation of the Canada Centre for Remote Sensing (CCRS) rigorous model.

This model reflects the physical reality of the complete viewing geometry, and corrects distortions due to the platform, sensor, Earth, and sometimes the deformation due to the cartographic projection. It then takes into consideration the satellite-sensor information. (Toutin and Cheng 2001).

The method is based upon principles related to orbitography, photogrammetry, geodesy and cartography. Except some basic considerations, all data required for the calculations is available in the metadata file provided by Space Imaging with every Carterra product (see *Table 1*).

This method requires digital elevation model and takes elevation data into consideration in every point of the image.

The orthorectification process requires GCP reference coordinates and elevation data in the same map projection/coordinate system as the original images have. Reference coordinates were therefore recalculated from EOV into UTM/WGS84 coordinates. Digital elevation model was also reprojected using rigorous transformation and bicubic resampling. GCP input coordinates were, as usual, raw image pixel coordinates (column/row number).

3.2 Thematic data extraction

3.2.1 Image characteristics

Both Ikonos images are pan-sharpened products, provided by fusioning 4-m resolution multispectral and 1-m resolution panchromatic data. The data for each band is stored on 11 bits, providing dynamic images with high radiometric resolution. The values of all of the four bands (red, green, blue, near infra-red) have Gaussian distribution on both images. Despite panchromatic-multispectral data fusion, the values are radiometrically meaningful, as dynamic range adjustment (DRA) was not applied.

3.2.2 Visual interpretation

The extremely high spatial and radiometric resolution of Ikonos images facilitates visual interpretation. Visual quality of both the original and resampled images is very good.

Colour composites of the images were created. As the images contain all of the three visible wavelengths (red, green and blue), it is possible to create true-colour composites. However, these composites don't contain the additional information provided by the NIR band that is very useful for vegetation mapping purposes.

"Standard" composites containing green, red and near infrared wavelengths (R: NIR, G: green, B: blue) were created to evaluate Ikonos images in order to obtain detailed information about woodland area.

Digital forestry map database and orthorectified false-colour composite were displayed to examine map updating possibilities (see *Figure 1*). Selected attributes such as tree species, age class, closure and mixture rate were estimated and updated by visual interpretation.

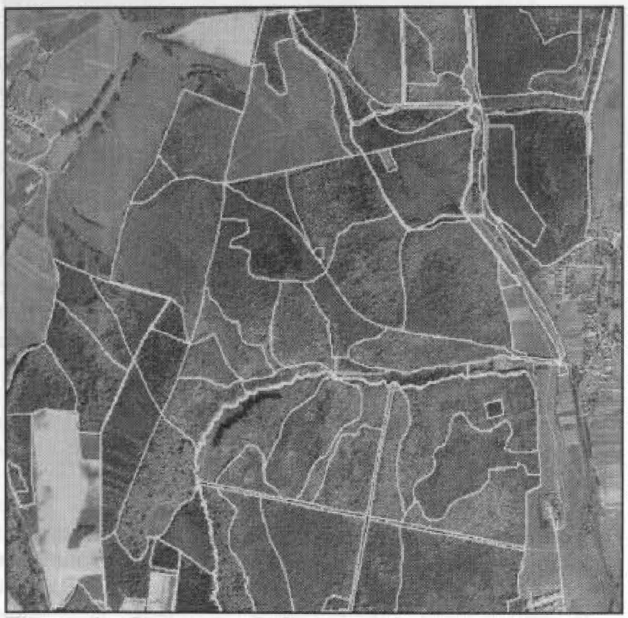


Figure 1. *One part of the research area. False colour composite (R: Ikonos NIR band, G: Ikonos green band, B: Ikonos blue band). Orthorectified image with digital forestry map database polygons. Image copyright Space Imaging 2001. Image resolution was digitally degraded.*

3.2.3 Vegetation indices

Vegetation indices were calculated in order to have meaningful quantitative results depending on vegetation cover. Normalized Difference Vegetation Index (NDVI) calculation was performed on both images. In order to insure data comparability and to eliminate sensor effects, image DN values were recalculated to provide radiance images. DN/Radiance coefficients were obtained from Space Imaging website. Resulting NDVI images were also included in colour composites in order to improve visual interpretation accuracy.

3.2.4 Classification

Two classification methods were used in order to examine digital data extraction possibilities. Classifications were carried out on a subsetted image covering one part of the research area. The classifications were realized on two types of images. In the first case, raw image pixel (DN) values were classified directly. Alternatively, in the second case, image segmentation was carried out on the subsetted image. After that, the resulting group of segments (image objects) with each object containing the mean value of pixels it was created from, was classified.

The two methods were compared for accuracy using ground truth data.

3.2.4.1 Ground truth data collection

Ground truth data was collected during two field check and ground truth data collection campaigns.

For the first campaign, field check sites and ground truth data collection areas have been selected during visual comparison of Ikonos images, non-supervised classification results and digital forestry map database. Ground truth data collected during the first campaign was then used to identify classes of non-supervised classifications and also to delimitate training areas for supervised classifications. It was also useful to improve visual interpretation reliability. Global positioning system and field datalogger were used to store every relevant information *in situ* provided by local forestry experts.

The principal aim of the second field check campaign was the evaluation of classification results. Field check areas were selected at sites where evident contradictions were found between classification results and the digital forestry map database. Data acquired during this campaign served primarily to validate classification results and to improve classification accuracy.

3.2.4.2 Unsupervised classifications

Unsupervised classifications were carried out using Isodata clustering method (Tou and Gonzalez 1974) with 30 classes. Classes of the resulting classified image were identified using ground-truth data and digital forestry database.

3.2.4.3 Supervised classifications

Training areas for supervised classifications were selected using ground truth data and visual image interpretation. Classifications were carried out using Maximum Likelihood method (Hord 1982). Classes were selected primarily based on tree species. Multiple different age class training samples were selected for each species to evaluate species and age class separability together.

3.2.4.4 Image segmentation

Image segmentation is an object-based approach to image analysis. Instead of pixel representation of the Earth's surface, usually called "image", segmentation provides image objects from groups of pixels according to specific homogeneity criteria. These image objects, represented for example by the mean values of the pixel group they were created from, can be classified based on spectral properties and/or spatial relationships. Image segmentation was performed on the above-mentioned subsetted image. Scale parameter was adjusted in order to create image objects of appropriate size.

Each resulting image object contained mean pixel values.

4 RESULTS AND DISCUSSION

4.1 Geometric accuracy of original Ikonos Geo images

Visual comparison of reprojected original Ikonos Geo images with some vectorized line features of the topographic reference maps makes it clear that the images contain distortions (*Figure 2a*). The distortions vary depending on surface topography.

The results of more accurate examinations based on selected check point pairs are presented in *Table 4*.

Image	69566	69567
Number of check points	38	16
Minimal error	2.81 m	3.61 m
Mean error	16.11 m	14.43 m
Maximal error	96.91 m	29.25 m

Table 4. *Errors of the original Ikonos Carterra Geo images reprojected to EOV map projection system, based on check points from the 1 : 10,000 topographic map database as reference.*

Although this sample cannot be considered representative due to the relatively low number of points, it shows clearly that, for both images, the distortions make it impossible to use these images for large-scale mapping. Maximal RMS error for the image 69566 reaches almost 100 m that can be explained by topographic effects. The point with the largest RMS error is situated on a hilltop.

Topographic effects can be important for both of the images not only in cross-track direction due to scanner geometry but also in other directions due to collection elevation values (see *Table 1*).

4.2 Geometric accuracy enhancement

Results of geometric accuracy enhancement examinations of multiple methods are shown in *Table 5*.

Image	69566			69567		
	Min	Mean	Max	Min	Mean	Max
Poly2	1.58	13.82	70.71	2.06	7.13	19.32
Poly3	1.41	13.26	41.35	0.68	5.98	20.13
Poly4	2.07	13.68	42.25	0.25	5.98	22.24
RPF	5.09	18.93	35.24	1.51	6.32	15.54
Ortho	0.98	6.61	20.08	0.20	4.38	9.38

Table 5. RMS errors of control points using different models of geometric correction.

Polyx: Polynomial transformation of xth order

RPF: Rational polynomial function of 2nd order

Ortho: Orthocorrection using satellite orbital modeling

If we compare *Table 4* and *Table 5*, we can have some information about the overall effects of the methods tested for geometric accuracy enhancement.

Simple polynomial transformations cannot eliminate distortions due to surface topography. However, these simple and robust methods can slightly decrease these effects. After the 3rd order, RMS errors begin to increase for the 4th order polynomial transformation.

Rational polynomial functions didn't perform as well as expected with our imagery. This is maybe due to non-perpendicular image acquisition angle and high surface relief.

It is clearly shown that orthorectification based on satellite orbital modeling provided the best results, however, for some points it didn't work as well as expected. We can state that the mean RMS error of 4..7 meters of the control points makes it possible to use the geometrically corrected images for large-scale mapping purposes (*Figure 2b*). It is possible to directly compare the images with other existing large-scale databases i.e. the digital forestry map database. Geometric accuracy obtained by orthorectification is sufficient for large-scale forest mapping, management and database updating.

The accuracy of the corrected Geo images are between the accuracy of Ikonos Carterra Map (5.7 RMSE) and Pro (4.8 m RMSE) product levels.

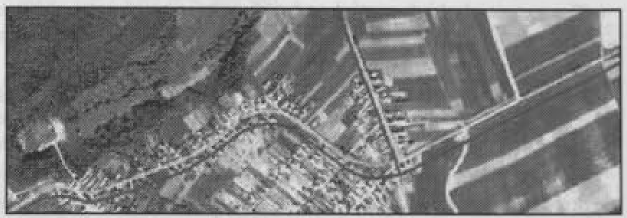


Figure 2a

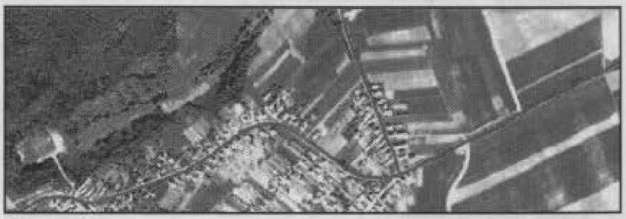


Figure 2b

Figure 2. *Vectorized linear features (roads, in red) shown on the original (2a) and orhtorectified (2b) Ikonos images. Image copyright Space Imaging 2001. Image resolution was digitally degraded.*

4.3 Thematic data extraction possibilities

The high spatial and radiometric resolution and geometric accuracy of the orthocorrected images ensure good visual interpretation possibilities. Trained image interpreters can easily interpret near-infrared false-colour composites as the colours and resolution are similar to false-colour infrared photographs. Visual interpretation and data comparison can be used efficiently for forestry map database updating.

However, it is important to note that the lack of middle infrared band sometimes makes the discrimination of different vegetation types difficult.

Vegetation indices such as NDVI may help to interpret vegetation types more clearly. These indices are also useful for vegetated area delimitation and for classification mask band creation. When displayed in colour composites, vegetation indices improve the efficiency of visual interpretation. However, direct relationship wasn't found between any of the forestry map attributes and NDVI.

Both unsupervised and supervised pixel-based classification results were very poor on the original images. The confusion matrix for the supervised classifications showed 31..58% of well-classified pixels for the control areas of the different classes. The poor results can be explained by the relatively poor spectral information content of the images and by the extremely high spatial resolution (*Figure 3a*). The visible bands (red, green, blue) are highly correlated and don't give many information for the discrimination of vegetation types (species, closure, mixture rate etc.) The near infrared band clearly improves classification accuracy, containing one of the maxima of vegetation reflectance curves. Some additional bands including middle infrared would be needed to improve pixel-based classification results.

It is also important to note that high spatial resolution doesn't facilitate spectra-based classification (see Franklin et al., 2000). Medium-resolution satellite images, such as SPOT HRVIR or Landsat TM images, have the advantage of "self-calculating" mean spectral values. This means that, from the forestry point of view, the coarser spatial resolution makes "average" pixels, where for example brighter and darker sides of individual tree crowns aren't visible. Instead, there are "tree crown averages" in each 10- or 28-m pixel of the image. The 1-meter pixel size of Ikonos makes spectral variations much greater.

Segmentation, however, improves largely the accuracy of both the supervised and unsupervised classifications. Confusion matrices of supervised classifications performed on previously segmented images showed 73..95% accuracy. The improvement is very important compared to the classifications on pixel-based images (*Figure 3b*).

The results show that segmentation, developed initially for microwave remote sensing image analysis, can be used and should be used on very high-resolution optical imagery to obtain good classification results. Spectral and textural information can be treated simultaneously during segmentation and subsequent classification.

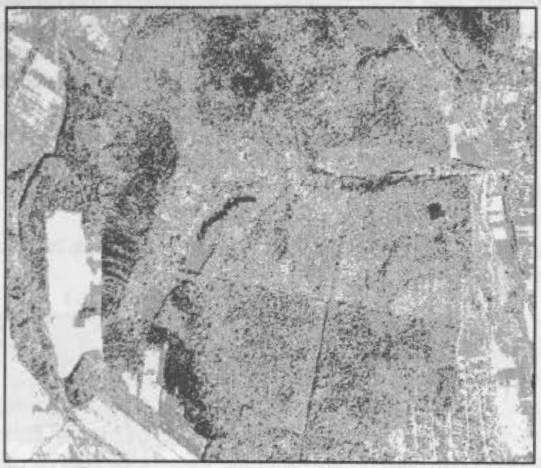


Figure 3a

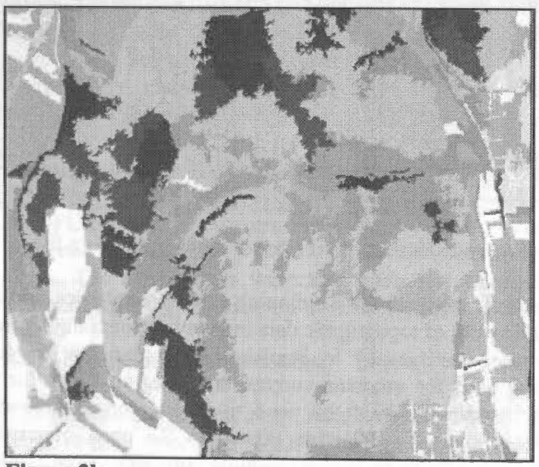


Figure 3b

Figure 3. *Unsupervised classification results on original (3a) and segmented image (3b) Similar colours indicate similar classes.*

5 CONCLUSION

Ikonos Geo images, as provided by the image supplier, don't have sufficient geometric accuracy for large-scale mapping. Price-sensitive users such as forestry sector are not able to purchase geometrically more accurate Ikonos Carterra image products. It is possible to improve geometric accuracy of Ikonos Geo images, and a higher accuracy level can be reached depending largely on the original image acquisition circumstances and topography of the area represented by the image. Orthorectification using satellite orbital modeling method seems to be the most effective and accurate modality to obtain high geometric accuracy. Visual interpretation works well on Ikonos images. Pixel-based classification techniques provide poor results due to extremely small pixel size and relatively poor spectral content. Image segmentation and subsequent classification improves largely the efficiency and accuracy. Tree species data can be well derived from Ikonos satellite imagery.

6 ACKNOWLEDGEMENTS

This research was supported by IKTA 3 / KÉPI 2000 programme of the Hungarian Ministry of Education. Authors wish to thank FABICAD Kft. for the cooperative collaboration, and PCI Geomatics support team for the help provided during orthorectification. Special thanks to János Szuhanyik (FABICAD Kft.) who took an active part in the geometric correction efforts and always provided useful help.

References from books:

Tou, Julius T., and Gonzalez, Rafael C., 1974, *Pattern Recognition Principles.* Reading, Massachussets, United States, Addison-Wesley Publishing Company.

Hord, R. Michael, 1982, *Digital Image Processing of Remotely Sensed Data.* New York, United States, Academic Press.

References from journals:

Franklin, S. E., Hall, R. J., Moskal, L. M., Maudie, A. J. and Lavigne, M. B., 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing*, Vol. 21, No. 1, pp. 61-79.

References from websites:

Davis, C. H., Wang, X., 2001. Planimetric accuracy of ikonos 1-m panchromatic image products
ASPRS 2001 paper.
http://www.pcigeomatics.com/tech-papers/
ASPRS_2001_ikonos.pdf

Guienko, G. A., 2001. Geometric accuracy of Ikonos: Zoom In. Submitted to IEEE Transactions on Geoscience and Remote Sensing
http://www.technion.ac.il/~guienko/

Toutin, T., Cheng, P. Demystification of Ikonos.
Earth Observation Magazine, 2000.
http://www.eomonline.com/Common/Archives/
July00/toutin.htm

USING LANDSAT-7 IMAGES TO UPDATE CANADIAN NATIONAL TOPOGRAPHIC DATA

M.-E. Martin, E. Loubier

Natural Resources Canada, Centre for Topographic information
010-2144 King Street West, Sherbrooke J1J 2E8, Québec Canada
www.ctis.nrcan.gc.ca
eloubier@RNCan.gc.ca

KEY WORDS: Cartography, Mapping, Change Detection, Acquisition, Production, Updating

ABSTRACT:

Natural Resources Canada's Centre for Topographic Information has a significant collection of topographic data at the 1:50 000 scale. The extent of the Canadian landmass makes it very difficult to update these data on a continuing basis. Following production of national Landsat-7 orthoimages coverage produced in partnership with more than 20 Canadian organisations, a project for updating topographic data using these orthoimages was initiated in June 2001. This article presents the operational constraints related to implementing an updating project of this magnitude (nature of topographic entities, image heterogeneity, size of the territory, production rate, absence of recent ground truth, etc.). Moreover, it describes why production could not be completely automated and why some visual interpretation had to be used.

RÉSUMÉ:

Le Centre d'information topographique de Ressources naturelles Canada possède une importante collection de données topographiques à l'échelle de 1/ 50 000. L'étendue du territoire canadien rend la mise à jour continue de ces données très difficile. Suite à la mise en production d'une couverture nationale d'orthoimages Landsat-7 via un partenariat de plus de 20 organisations canadiennes, un projet de mise à jour des données topographiques utilisant ces orthoimages a débuté en juin 2001. Cet article présente les contraintes opérationnelles liées à la mise en production d'un projet de mise à jour d'une telle ampleur : nature des entités à cartographier, hétérogénéité des images, étendue du territoire, rythme de production, absence de vérité-terrain récente, etc. Il explique pourquoi la mise en production n'a pu être totalement automatisée et pourquoi l'interprétation visuelle a dû être en partie utilisée.

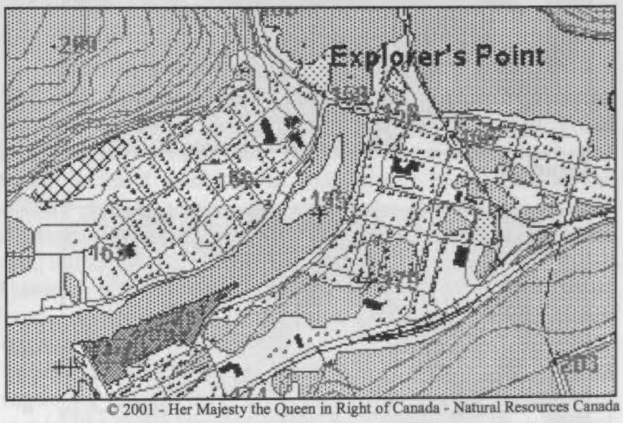
1. TOPOGRAPHIC DATA OF CANADA

The Centre for Topographic Information (CTI), part of Natural Resources Canada, is responsible for providing the digital representation of the Canadian landmass required by users of geographical information in support of the protection and development of the country, which includes its resources and their use. The main type of information managed by the CTI is topographic data at the 1:50 000 scale. These data yield more than 14 000 maps based on the National Topographic System (NTS). They cover nearly the entire Canadian landmass, with the exception of certain more northern regions that haven't been mapped at this scale. The data are available in vector form (National Topographic DataBase or NTDB) or on paper (Canadian Topographic Maps).

The vector data contain many topographic features divided into 14 themes: hydrography, vegetation, manmade features, hypsography, designated areas, road network, power network, rail network, relief and landform, water saturated soils, roads, general, and toponymy. These data were originally acquired by photogrammetry and ground survey, and constitute a unique resource of topographic data for all of Canada.

2. CONTEXT

Canadian topographic data were acquired over the last 50 years. Many focused update projects have been completed in recent years, such as updating the road network, major towns, and more densely populated regions. Yet no updating project of national scope has been carried out, since the costs and time involved are too great. The validity of some maps can be as old as 1948.



© 2001 - Her Majesty the Queen in Right of Canada - Natural Resources Canada

Figure 1: Extract of a 1:50 000 topographic data file

The current context for developing new technologies and the increasing use of topographic data in many sectors ranging from Web-based cartography to precision agriculture has led to a growing need for accurate, current, and integrated data for the entire landmass. While the need to update topographic data continued to grow, the costs related to the photogrammetric process and high-resolution satellite imagery were simply prohibitive.

The flexible distribution rights policy and the lower acquisition costs of imagery of the Landsat-7 satellite launched in 1999 opened the door to new possibilities for data providers and users. As such, a project funded by GeoConnections with more than 20 Canadian federal, provincial, and territorial organisations working in partnership focused on creating national Landsat-7 orthoimage coverage.

Although the Landsat-7 images do not have optimal spatial resolution for working with data at the 1:50 000 scale, they are adequate for producing a partial update. Moreover, being able to produce national orthoimage coverage brings a Canada-wide update into the realm of possibility.

3. OBJECTIVES

CTI initiated a national program to update topographic data at the 1:50 000 scale using Landsat-7 orthoimages.

The objectives of this project are:
- Enhance the planimetric accuracy of existing data;
- Update topographic data;
- Restructure the data in a format that complies with international standards in a new database: the Geospatial Database (GDB).

This article deals mainly with the first two project objectives. Interested individuals can find additional information about the Geospatial Database and its standards by consulting CTI website at www.cits.rncan.gc.ca.

This project is an audacious production project both in scope and the number of operational constraints involved. Indeed, thousands of topographic data must be updated homogeneously while the terrain's conditions vary greatly.

4. METHODOLOGY

4.1. Landsat-7 Orthoimages

The orthoimages used for the project were produced from Landsat-7 images. The objective is to produce orthoimages with an accuracy of 50 m or better with most of them under 30m. To date the accuracy of the orthoimages produced varies from 11 m to 26 m. To cover the entire Canadian landmass, more than 700 Landsat-7 images will be acquired over a period of 5 years. The images, usually collected from early May to mid-October, are selected in order to minimize cloud cover and maximize vegetation activity. For more details about Landsat-7 orthoimages characteristics and production methods, see Bélanger & Benoit 2002.

The resulting orthoimages (15m for the panchromatic band and 30m for the multispectral bands) were fused using ©PCI's PANFUSE tool. They were then subdivided according to the National Topographic System for correspondence with NTS maps. The orthoimages were not mosaicked even when two orthoimages were necessary to cover a mapsheet.

4.2. Planimetric Enhancement

The planimetric accuracy of Canadian topographic data varies from 10 m to more than 100 m. When the orthoimage accuracy is better than the accuracy of the topographic data, the topographic data are corrected planimetrically. Altought the data shown on map could be out of date, amelioration of its accuracy is a great improvement for customers.

The planimetric correction is outsourced to private companies. Although we recommend that contractors use a method based on local linear correction to carry out this task, no specific method is imposed. CTI quality control ensures that the quality of the work complies with our accuracy requirements.

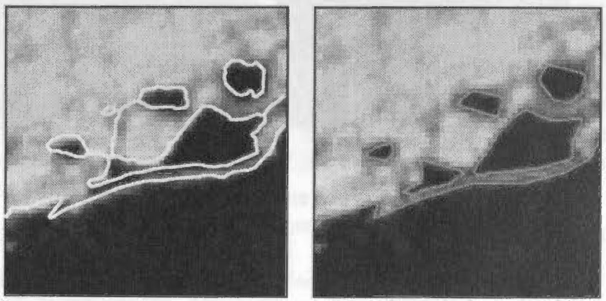
The maximal relative accuracy allowed between orthoimages and vector data is 20 m with a degree of confidence of 90%. The relative accuracy is calculated by measuring the displacement between the orthoimage and a representative sampling of the resulting planimetrically corrected vector data. The absolute accuracy obtained from the resulting vector data is computed based on relative accuracy and orthoimage accuracy. To date, more than 350 mapsheets have been planimetrically improved.

4.3. Other available datasets

Orthoimages and topographic data are not the solely data available for this project. Aerial photographs used to originally produce the topographic data are still available. Although they are as old as the data to update, they can be very useful to understand the environmental context in some regions. Canadian updated national road network and Canadian digital elevation data are under completion and can also help interpreting the orthoimages when they are available.

4.4. Selection of a update method

Different approaches can be use to update topographic information at scale 1:50 000. On one hand, it would have been possible to use the Landsat-7 orthoimage to create a totally new map coverage of Canada. However, this option would have meant quality deterioration, or even loss, of many entities that did not change over time. Data like the hydrographic network would have been difficult to recreate with Landsat-7 imagery. (Figure 2) For this reason, this approach was not selected.


Original hydrography network

Example of the hydrographic network extracted from Landsat-7 imagery

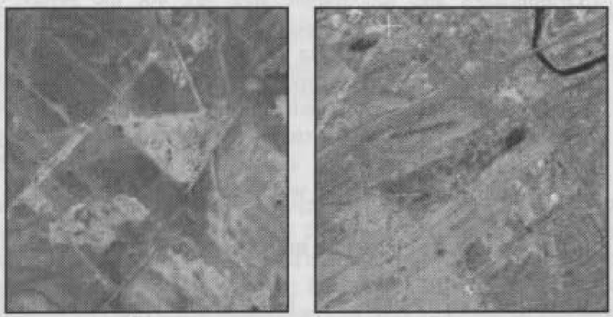
Figure 2: Comparison of quality difference between original data and data capture on Landsat-7 orthoimages.

On the other hand, a change detection approach was deemed more appropriate. The change detection process makes it possible to validate the existence, the geometry, and the description of features visible in the image, whether they have been modified or not. New features are captured as seen on the image and features that are not visible are preserved and transferred if there is no proof that they should be eliminated.

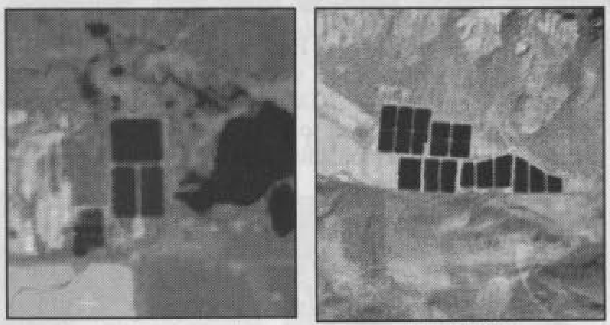
The main advantage of a change detection approach is to combine, both in terms of resolution and content, the currency of Landsat-7 data with the heritage data provided by years of photogrammetric data capture and ground surveys. Moreover, this approach leaves the door open for future updates with better tools or imagery.

A number of avenues are open for detecting change and updating data using satellite imagery. They include methods based on classification, texture recognition, snakes, and so on. Nevertheless, the operational constraints in this project significantly limited the choices available. The main constraints encountered in implementing production are provided below.

Constraint 1 - Entities not suitable to the Landsat-7 context: Entities to update are very heterogeneous and their definition does not always suit to Landsat-7 images. For example, topographic mapping normally distinguish different types of mining operations and indicate whether it is a pit, quarry, or mine. Landsat-7 orthoimages may show the extent of the mining area but can hardly shows its type.



Pit Quarry



Sewage disposal ponds Water basins

Figure 3: Examples of entity definition not suitable to the Landsat-7 image

Moreover, due to the spatial resolution of Landsat-7 images, certain entities, such as hydroelectric lines and buildings, cannot be updated at all. It was therefore decided to limit the entities to be updated to major themes generally visible in the images such as vegetation, hydrography, water-saturated soils, built-up areas, and certain designated areas. As much as possible, the definition of entities was adjusted to comply with the orthoimages possibilities. For example, the entity ''mining operation: type unknown'' was added to the existing entities set. Other entities, like the roads, are acquired through other projects or are obtained through partnerships with specialised partners.

As it will be explained later, the adjustments made were not sufficient to allow the implementation of a completely automated change detection method.

Constraint 2 – Image heterogeneity and size of the territory: More than 700 images are required to cover the entire country. The coverage should be completed within five years. In order to do so, images from various seasons will be used. Despite this constraint, the objective is to conduct the map updating process as homogeneously as possible. As such, interpretation rules, entity definition and output format characteristics are the same all over the Canadian territory. An entity that should be fairly easy to interpret, such as vegetation, can be a problem when the quantity of variables involved is taken into consideration: variation of the type of vegetation, variation of the images acquisition date, lack of ground truth and so on.

Constraint 3 – Production rate: CTI is committed to provide the most up-to-date data possible. In doing so, it is extremely important to have a production rate that prevents the data offered from becoming obsolete. This problem takes critical proportions when the ongoing updating process must be carried out for a territory as large as Canada. The required production rate limits the amount of time that can be spent on a single image. For example images preparation is conducted using fully automated methods, such as main component analysis and unsupervised classification.

4.5. Description of the updating process

Updating data through change detection process is done by studying the evolution of a given entity. This study leads to important questions and decisions that can not easily be undertaken by automated systems. The main questions set by the change detection concerns visibility, entity identification, importance of the change, minimal dimensions, generalisation and identification of the change detected. Figure 4 summarises the change detection decision tree through which every entity has to go. Some negatives branches (no capture, entity destruction) have been omitted to simplify the figure.

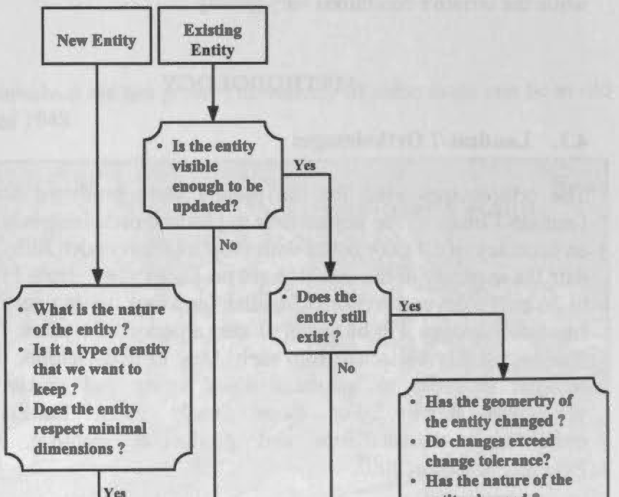

New Entity
Existing Entity
• Is the entity visible enough to be updated?
Yes
No
• What is the nature of the entity ?
• Is it a type of entity that we want to keep ?
• Does the entity respect minimal dimensions ?
Does the entity still exist ?
Yes
No
• Has the geomertry of the entity changed ?
• Do changes exceed change tolerance?
• Has the nature of the entity changed ?
• Is it a type of entity that we want to keep ?
• Does the entity respect minimal dimensions ?
Yes
Capture the entity
Keep the original entity
Remove entity
There is a change : Capture the entity
No Change : Keep the original entity
Situation : New Entity
Situation : Transfer
Situation : Destruction
Situation : Modification
Situation : Confirmation

Figure 4: Change detection decision tree.

Entity visibility: The evaluation of the visibility of an entity may seem easy but the difference between an entity that has disappeared and an entity that is not visible can be difficult, specially for entities like wetlands. An entity may not be visible and still continue to exist. The first decision to make is therefore to decide if the entity is visible enough to be updated. Non-visible entities are preserved as they were in the old data set and information indicating that they were not updated is registered (Transfer situation).

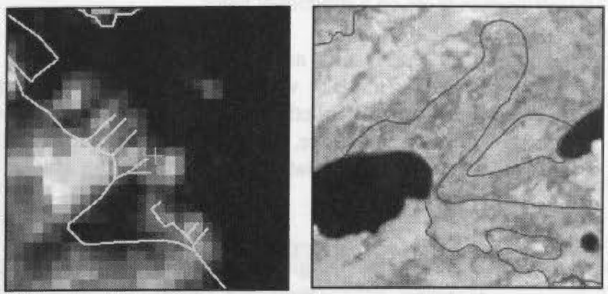
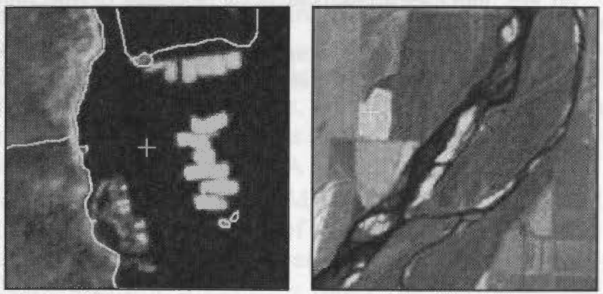

Wharfs Wetlands

Figure 5: Example of non-visible entities. Entities are too small or do not contrast enough with the surroundings to be updated

Entity nature: Entity nature, or identification, determines if a visible entity should or should not be mapped. Identification of an entity does not rely uniquely on its composition but also on its utilisation. Some entities, like roads, should not be mapped although similar entities like runways should. Sometimes, entity nature can only be deducted for the surrounding context.


Timber rafts Sand in water

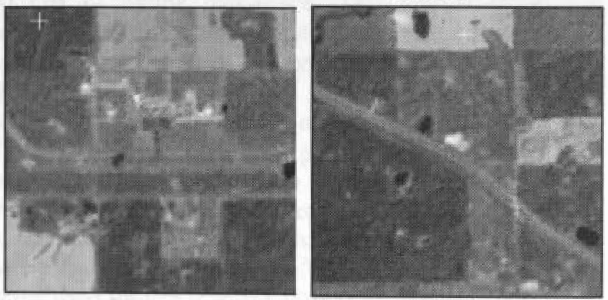

Runway Road

Figure 6: Examples where the surrounding context is necessary to identify an entity

Change versus entity nature: If the entity visible on the image has been geometrically modified compared to the original data set, a decision has to be taken whether or not the change should be mapped. For example, lakes are shown on topographic maps with their normal water level. Half-drained lakes or inundation should not be mapped if those water levels are only temporary. The normal water level may not have changed.

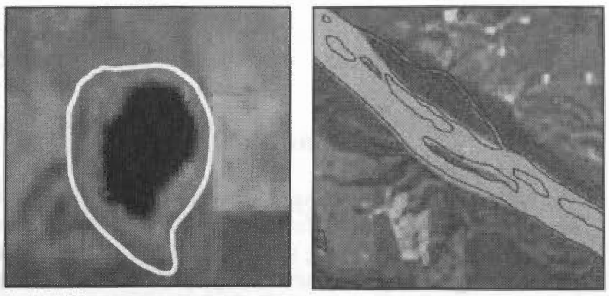


Figure 7: Temporary draining and inundation

Change versus change tolerances: The extents of the visible change also need to be evaluated against change tolerances. A change that is under the change tolerances is not considered significant and should not be mapped. The use of change tolerances gives more control on the change detection process, as only significant changes are mapped. The change tolerances have been defined to fit the Landsat-7 resolution and vary according to entity visibility.

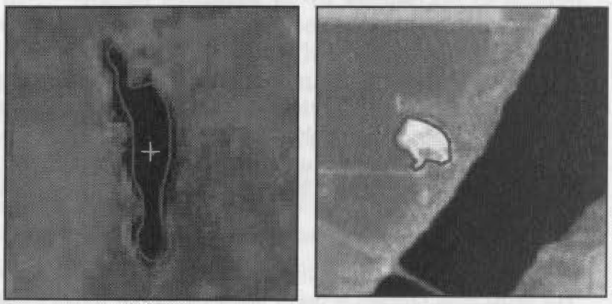

Lake Quarry

Figure 8: Are those changes significant enough to be mapped?

Minimal dimensions: The extent of the entities should also be evaluated according to the 1:50000 scale needs. Some small entities will not be mapped and others will be generalised. A river too narrow to be represented as a surface should for example be mapped as a line. Sparse forest entities could be represented as one big forest if the clearings between them are too small.

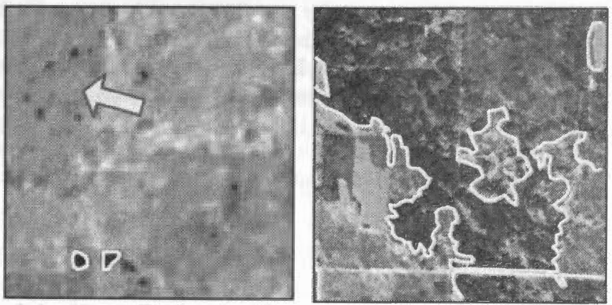

Lakes too small to be mapped Generalised forest area

Figure 9: Examples where minimal dimensions affect mapping

Metadata: The change detection process leads to different situations: destruction of entities, confirmation, modification, transfer or new mapping. Each of those situations comes from decisions taken while going over the decision tree. They lead to different information on the entity: its precision, its validity, or its historic evolution. All this information should be registered as metadata. Therefore, for every entity, the right metadata have to be identified and saved.

5. PRODUCTION

The objective of the updating process is to obtain an up-to-date mapping coverage. In order to do so, a change detection approach based on existing topographic data sets was chosen over creating a completely new cartography. The chosen change detection process implies that the evolution of every entity is studied and preserved. This evolution is deducted from decisions taken while going through a decision tree.

There are different ways to update data. Traditional methods like classification, filtering, vectoring, are some of those methods. Used with GIS systems, they could give interesting automated results for specific usage. However, given the constraints of a national coverage using existing data, it was shown that it is very difficult to automate the topographic map updating process. The nature of the entities to map was too difficult to discriminate by an automated method. Given the various constraints listed above (image heterogeneity, territory size, production rate, and definition of entities) and the complexity of the decision process, visual inspection had to be partly maintained as it is the only method that enables us to update the data adequately. As the process could no be completely automated, the updating work was given to private companies. Companies are free to use the method they prefer to conduct the map updating.

CIT responsibility is brought down to inspection of the results. The inspection process is faster and simpler but implies that the same decision tree is used for quality control. Some decisions can be automated, like those referring to minimal dimensions and generalisation. The rest of the decisions could not be automated with satisfactory results. For instance identifying the nature of an entity depends on the image but also on the context. As for today, minimal dimension, generalisation and image preparation process have been automated. Work is still done to automate other inspection process like the automatic identification of the metadata.

5.1. Resulting data

The production process for correcting and updating began in June 2001; 200 maps were updated during fiscal 2001-2002. These maps represent 17 Landsat-7 images. In the years to come, the production rate should accelerate to attain an objective of more than 1000 maps per year. Planimetric correction of the data is also proceeding; 350 were corrected during fiscal 2001-2002 and nearly 1500 maps will be during the next fiscal year.

Canadian topographic data with enhanced accuracy will be distributed to users as accurate data and correction matrices. The updated maps will be used and distributed as a new product.

6. CONCLUSIONS

The purpose of this article was to present a production project using Landsat-7 images and the operational constraints involved in implementing practical applications on a large scale.

Updating data for territory covering more than 9 900 000 km^2 involves interpreting more than 700 images and represents a colossal challenge. The change detection process and the project's operational constraints, such as production rate, image heterogeneity, absence of recent ground truth, and territory size, obliged us to use methods partly based on visual interpretation. The change detection process of topographic data is too complicated to be totally automated. Existing automated process and remote sensing tools may help human operators but still can't replace them on a functional basis.

As for today, the results are satisfactory. Nevertheless, the objective is still to replace visual interpretation to the degree possible by automated methods, subject to the operational constraints mentioned above, which would reduce or eliminate the use of human interpretation and enhance the quality and homogeneity of the results.

REFERENCES

Bélanger D., Benoit A., 2002. Production of Landsat-7 orthoimages covering the Canadian Landmass. Presented to 95th Annual Geomatics Conference of the Canadian Institute of Geomatics (CIG), Ottawa , Canada.

Boutin D., Massé F. 2001. The new age of topographic Data – Management and Access In : GIS 2001 proceedings, Vancouver Canada.

CTIS, 2001. The Geospatial Database ... The national Topographic Data Base of Tomorrow, *CTIS Newsletter,* 4(1), pp. 4-6. http://www.cits.rncan.gc.ca/ (accessed 15 Apr. 2002)

CTIS, 2001. CTIS and international standards, *CTIS Newsletter,* 4(1), pp. 7-8. http://www.cits.rncan.gc.ca/ (accessed 15 Apr. 2002).

A HIGH QUALITY DIGITAL ORTHOPHOTO OF THE MT. EVEREST AND ITS SURROUNDING

Baral Toya Nath[a], Michael Hahn[a], Frank Neidhart[b]

[a] University of Applied Sciences, Stuttgart
Faculty: Surveying and Geo-informatics, Schellingstr. 24, 70174 Stuttgart
m.hahn.fbv@fht-stuttgart.de
www.fht-stuttgart.de
[b] Z/I Imaging, Oberkochen, Germany
f.neidhart@ziimaging.de
www.ziimaging.com

Commission IV, WG IV/7

KEY WORDS: Orthophoto, Digital Terrain Model, Mount Everest, Nepal, Orthomosaic, Aerial triangulation, Color-depth, Image fusion, Elevation.

ABSTRACT:

The project area is situated in an antique-mountainous country, Nepal where the elevation varies from 58 m to 8850 m above sea level. It is located approx. in between 26N to 30N latitude and 80E to 88E longitude. Nepal is known as having the highest mountains in the world including Mount Everest. As the nature of the terrain is mostly mountainous with extremely irregular topography, generation of DTM and orthophoto mosaic is a challenging task. The specific area under study is Mount Everest and its surroundings, where the elevation varies from 3000m to 8850m. Most of the area is fully covered with snow and steep slopes. A

A digital terrain model (DTM) of the Mount Everest and its surroundings and a set of orthorectified aerial image data products is being created by this mission using photogrammetric workstation (Image Station) with some 70 GB of online disk-space. This unique DTM and subsequently digital orthophotomap is being created from aerial images data using commercially available software (fully automatic) from Z/I Imaging a joint venture company between Intergraph and Zeiss GmbH.

The aerial photographs were produced using a jet aircraft and RC 20 aerial camera equipped with GPS/INS installation onboard. A total of 82 photo frame (distributed in 9 flight lines having 80% forward overlap and 30% lateral overlap) of approx.1:50000 scale are being used for this research project.

Scanning was carried out from second generation transparent originals in 12 bit color-depth using PhotoScan 2001. Orientation procedures shall be fully automatic. Aerial triangulation shall be carried out using ISAAT2001-Image Station Automatic Aerial Triangulation software. Similarly DTM and orthophoto production software from the same provider shall be used for the final products.

Performance of an image fusion of high resolution monochrome orthophoto and a color low resolution satellite image of the same area shall be carried out to obtain a high resolution color orthophoto of the region that can be used for visualization and 3-0 animations for further use.

This paper will report on the issues of automatic orientation problems of images due to the lack of well-known features in the whole image, extreme height differences and its effect in aerial triangulation. It will address the automatic matching quality using images with large differences in scale and poor contrast. It will also report on the issues of orthorectification including tone balancing and mosaicing for a final orthophotomap where the image has extremely bright and extremely dark parts in it due to snow and shadows. Assessment of the quality of DTM and orthophotomap thus produced shall be another part of the report.

This project approach heralds a new age for aerial imaging and remote sensing whereby orthorectified image maps can be produced for anywhere without the need for any expensive and time-consuming Ground Control Point acquisition.

This paper briefly describes all activities pertaining to the photgrammetric processing of high mountainous area where the elevation difference is extremely high and the lack of good contrast and textures in the images to be used. There is then some discussion on improving the quality of final products in a fully automatic digital photogrammetric environments.

HIGH ACCURACY HANDHELD MAPPING SYSTEM FOR FAST HELICOPTER DEPLOYMENT

J. Skaloud [a],*, J. Vallet [b]

[a] Geodetic Engineering Laboratory,
[b] Laboratory of Photogrammetry,
Swiss Federal Institute of Technology (EPFL), 7, Avenue Piccard, CH - 1015 Lausanne, Switzerland
jan.skaloud@epfl.ch, julien.vallet@epfl.ch

Commission IV, WG IV/7

KEY WORDS: Aerial, Mobile, Disaster, Mapping, GPS, Precision, Sensor, Fusion.

ABSTRACT:

This paper presents a self-contained, light and flexible mapping system that can be quickly deployed into inaccessible areas. Although designed to measure wind-transported snow volumes and the snow avalanche runoff over an experimental site, the system is suitable to any large-scale 3-D terrain mapping. The system comprises of a supporting electronic that is loosely linked to a light but ridged sensor block containing a camera, an IMU and a GPS antenna. The relatively small size and weight of the sensor block permits manual pointing of the camera (film-based or digital) towards either the mountain face or the valley bottom. Such hand-held steering allows mapping of the avalanche release and deposit zones during the same flight and also dampens the engine-induced vibration. The exterior orientation (EO) parameters of the camera are determined directly by GPS/IMU integration. The orientation performance of the navigation solution is improved by integrating the data from a second GPS antenna placed on the helicopter tail. Once the system is calibrated and with EO determined the mapping procedure implies only the forward resection without the need of aerial triangulation (AT) and ground control points. The practical experiences have demonstrated a mapping accuracy of 10cm and 15cm in the horizontal and vertical, respectively.

RÉSUMÉ:

Cet article présente un système de cartographie indépendant, léger et flexible qui peut être rapidement utilisé dans des zones incessibles. Bien que conçu à l'origine pour mesurer des volumes de neige transportés par le vent ou par des avalanches sur un site expérimental, le système convient également à toute cartographie 3D à grande échelle du terrain. Le système comporte une structure rigide et légère, supportant une caméra, une unité inertielle et une antenne GPS, qui est liée à l'électronique d'acquisition et de traitement. Le faible poids et l'encombrement restreint de la structure permet d'orienter manuellement la caméra (analogue ou digitale) de manière oblique ou verticale. Ce maintien manuel du système permet de saisir la zone de dépôt et la zone de déclenchement de l'avalanche lors du même vol et d'amortir les vibrations induites par l'hélicoptère. Les paramètres de l'orientation externe (EO) de la caméra sont directement déterminés par l'intégration IMU/GPS. La précision d'orientation de la solution de navigation est améliorée par l'intégration d'une seconde antenne GPS sur la queue de l'hélicoptère.
Une fois le système calibré et l'orientation externe déterminée, la procédure de cartographie nécessite uniquement une résection inverse sans avoir recours à une aérotriangulation (AT) avec des points d'ajustage au sol. Des tests pratiques est ressortie une précision de 10cm en planimétrie et 15cm en altimétrie sur la mesure de points au sol.

1. INTRODUCTION

1.3 Motivation

The sporadic and erratic occurrence of avalanches or landslides requires mapping systems that are flexible and can be quickly deployed into inaccessible areas. Furthermore, periodic mapping of avalanche test sites or natural hazard areas is sometimes required as in the particular case of Alps (Issler, 1999). The relatively small size of these areas, their spatial characteristics and their seasonal use put particular prerequisites on the mapping system:

1.4 System Requirements

- fast set-up and availability (days or hours)
- relative independence from a particular carrier
- possibility to map near vertical (mountain faces) and horizontal (valley bottoms) features during the same flight
- high relative and absolute mapping accuracy (~20cm)
- no assistance of ground control points.

1.5 Choice of a System Concept

Considering the above mentioned requirements, two appropriate concepts can be considered:

- large scale photogrammetry using handheld light aerial camera coupled with a GPS/INS sensors
- laser scanner (i.e. LIDAR) integrated with a GPS/INS.

The differences between these two concepts are highlighted in the Table 1 bellow.

Handheld Camera, GPS/INS	LIDAR, GPS/INS
accuracy - see Section 5	
+ fast set-up time (<20 min)	– long set-up time
+ independent from a carrier	– particular requirements
– contrast dependent	+ illumination independent
+ mountain faces mapping	–steep slopes mapp. limited
– semi-automatic mapping	+ automatic DTM generation
+ direct cost: $100K	- direct cost: $1500K
+ low cost for small areas	- high cost for small areas

Table 1. Comparison between different mapping approaches.

The use of an airborne laser scanning system may be appealing mainly due its automatic generation of DTM and its independence of surface illumination. However, its acquisition cost cannot usually be justified by the small application scale and its seasonal use. Moreover, the portability of the laser scanning system between different carriers is limited because of specific demands (e.g., floor view) and the long set-up time. The cost of having a designated system carrier is therefore another prohibiting factor for such type of application. An alternative solution by mandating a third-party service provider is not suitable due to the need of system availability on a short-time notice. To conclude this discussion, although both concepts are most likely comparable in terms of accuracy, the photogrammetric approach was chosen as the better suited due to its improved flexibility, higher versatility, lower acquisition costs and relative independence from a particular rescue helicopter.

The choice of a helicopter as the mapping system carrier is justified by its capability to fly close to the ground at low speed. This allows capturing photographs in large-scale and provides better flight line navigation. In following, particularities of the system will be described together with an analysis of its performance.

2. AIRBORNE DATA COLLECTION SYSTEM

The system can be divided into three essential parts:

- light and rigid sensor block (Figure 1), comprising of a handheld aerial camera (film or digital), GPS antenna and an IMU
- a second GPS antenna placed on the tail of the helicopter
- the supporting electronic, i.e. the GPS receivers, GPS/INS integrator and data logger (Figure 2)

2.1 Imagery Component

Two film-based aerial cameras have been successfully tested, while implementation of a digital one is on the way. The system prototype used a light handheld Linhof Aerotechnika camera that stores up to 200 colour, large format photographs (4x5 inches) and has a 90mm wide angle lens. Its total weight reaches 8kg. The most recent testing have been conducted with the Tomtecs HIEI G4 with 370 colour pictures capacity, 5x5 inch format and 90mm lens. Both cameras fulfil the required flexibility and provides images of high quality. While the Tomtecs camera offers an internal synchronisation with the GPS time, the employment of Linhof camera required additional installation of photodiodes and a RC circuit that sense the shutter open and sends an event-input pulse towards the data logger.

2.2 Navigation Component

The system currently employs two Leica SR530 GPS receivers on the helicopter and additional GPS receiver/s on the ground. The IMU is a customised tactical-grade strapdown inertial system (LN-200) with fibre-optic gyros. A real-time GPS/IMU integrating component was provided by Dynamic Research Corporation (DRC, 2001). It integrates the IMU data with the Trimble Force 5 military receiver over the VME (Versa Module Eurocard) bus in real-time and at 400Hz. The navigation solution is stored concurrently with the IMU data. When high orientation accuracy is required the recorded IMU data are used in a post-mission integration with the differential carrier phase GPS data.

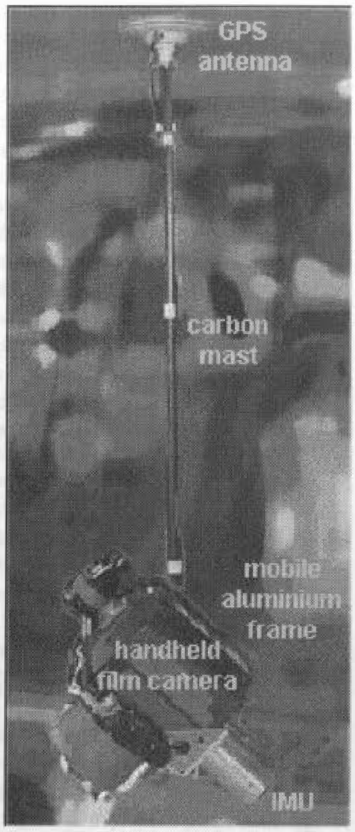


Figure 1. The sensor block.

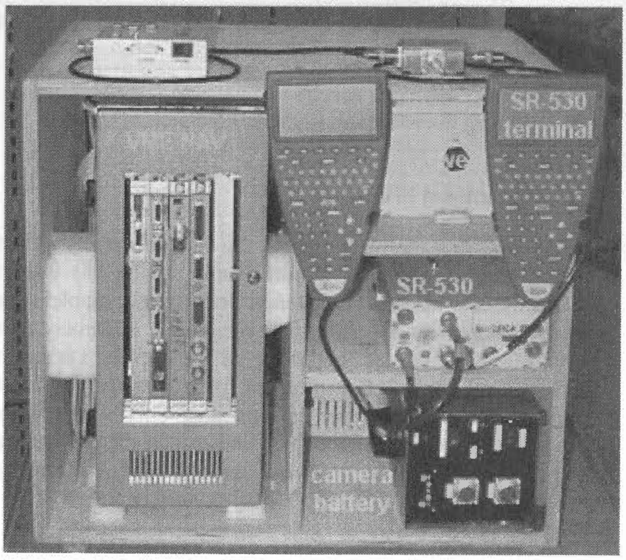


Figure 2. The sensor supporting electronic that is quickly installable.

2.3 Helicopter Mount

The placement of the sensors in the helicopter was chosen to comply with the following requirements:

- prevent any differential movements between sensors
- reduce the disturbing effect of the vibrations

- enable manual orientation of the camera to capture oblique as well as vertical imagery
- offer a firm attachment of the sensor block that can facilitate long transition flights and can permit the use of a GPS-derived azimuth during the period of IMU-alignment

Addressing the first objective, a light and rigid steel-aluminium-carbon holder was designed to connect all sensors within a rigid block as shown in Figure 1. Its relatively small size permits handholding. This responds to the other two requirements as the vibrations are dampened through the body of an operator (Figure 3) who can also aim the camera to a desirable direction. The last requirement is fulfilled by a steel socket-joint by which the system can be quickly attached to the side of a helicopter.

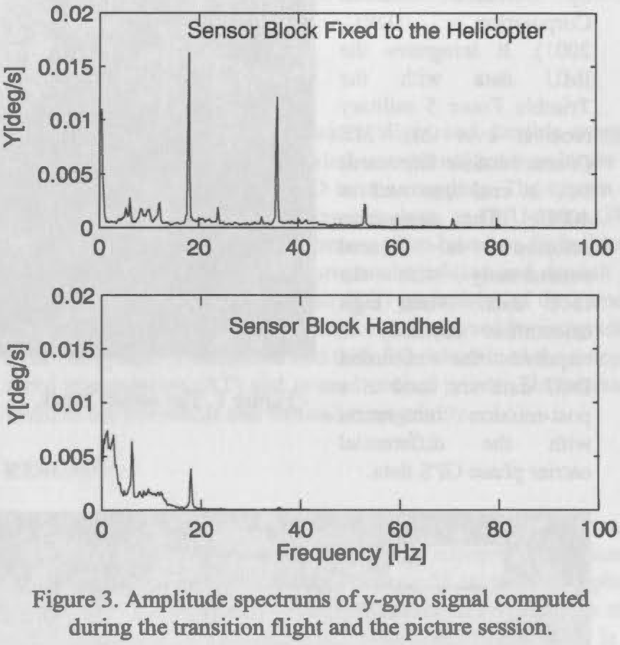


Figure 3. Amplitude spectrums of y-gyro signal computed during the transition flight and the picture session.

2.4 Operational Procedure

At the transition period of a flight the system is stiffly mounted outside the helicopter on a steel frame. During this time, the data from a second GPS antenna provide the supplementary azimuth important for achieving the precise alignment of the inertial system. At the beginning of the picture session, the camera-IMU-GPS block is removed from its holder through the side door to handheld by the operator. This allows manual orientation towards the mountain face around the omega angle. The sensor block can be fixed again during another flight transition period, or in a long flight line.

3. NAVIGATION DATA PROCESSING

Simulations and practical experience have shown that camera EO parameters need to be determined with an accuracy of 5-15cm in position and 0.005-0.008deg in orientation to satisfy the highest requirements of 10-30cm on the mapping accuracy (Vallet et al., 2000).

As the real-time GPS/INS navigation solution is accurate to 10-20m in position and ~0.1deg in orientation, following data processing scheme is adopted either entirely, or partially to rise up the EO estimate to an appropriate level according to the mapping requirements.

3.1 Carrier Phase Differential GPS (CPDGPS)

The processing of the GPS code and carrier phase data is the determining factor of positioning accuracy. Depending on the number and the distribution of the base stations, the superior accuracy limit is most likely 5-10cm in the position. *GrafNav/GrafMov* (Waypoint Consulting Inc., 2001) processing engine is used for that purpose and a search is on the way to supply the user with an automated and reliable data handling. As the measurement frequency of the employed GPS receivers is 10Hz, the trajectory is sufficiently over-sampled to accurately interpolate the events of camera exposure.

3.2 Post-Mission GPS/INS Filtering and Smoothing

The approximate orientation accuracy at real-time is at 0.1deg. When higher accuracy is required, the recorded inertial measurements need to be integrated with the position and velocity data based on the CPDGPS estimate computed in the previous step. The *POSProc* (Applanix Corp., 2000) represents a fine tool for such a purpose as it optimally blends the inertial data with different kinds of navigation data via centralised Kalman filtering/smoothing configuration. The versions prior to 3.0 are furthermore ideal for a developer, as they permit to fine-tune the Kalman filter parameters to specific scenarios and IMUs. Automated data treatment and evaluation can also achieved by the batch processing and is planned to be implemented in future.

3.3 GPS-Azimuth Aiding

It is a well-known fact that updating an inertial system with navigation information of better quality prevents the unbounded growth of position and attitude errors. As usually, GPS provides a mean of 'in-flight alignment' of the inertial system, removing the need for performing the 'north seeking' process prior to the flight. However, the accuracy of the in-flight alignment is strongly affected by the dynamic of the carrier. Since the accelerations induced by helicopter manoeuvres are

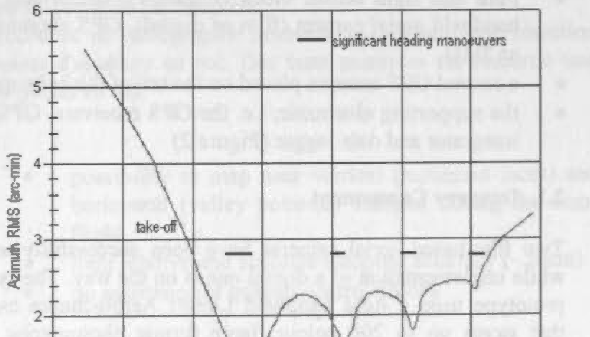


Figure 4. The effect of aircraft manoeuvres and GPS-derived azimuth aiding on the estimated azimuth accuracy.

considerably smaller than those of an aircraft, this problem needs to be circumvented by other means.

Here, the necessary information is derived from a second GPS antenna placed at the tail of the helicopter. Thanks to a relatively long (5 meter) separation between the two GPS antennas, the derived azimuth is sufficiently accurate and can be used as an additional information aiding the IMU. Again, the *GrafNav/GrafMov* software package is employed for calculating the relative vector between the two moving GPS antennas.

The theoretical effect of GPS-azimuth aiding on the INS alignment is depicted in Figure 4, while the practical experience showed that this helps in achieving alignment accuracy of better than 0.01deg in the azimuth. Furthermore, this accuracy can be maintained even during long flight lines but requires the camera to remain fixed in its holder.

4. SYSTEM CALIBRATION

The calibration of the sensors involves determining the relative displacement and orientation differences (boresight) between the camera and the inertial system, the parameters of the interior orientation as well as the constant synchronisation offsets inherently present due to data transmission and internal hardware delays.

4.1 Camera Calibration

The calibration of the camera consisted of two phases. First, the focal length, the principal point and the radial distortion were calibrated during a preliminary flight over a well signalised test field using the GPS-assisted AT. They were primary introduced together as unknowns into the *Bingo-F* bundle adjustment software (Kruck, 2001), while their final estimate was performed separately by iterations.

A second calibration phase was performed when large residual systematic effects were observed under extremely cold winter conditions. Investigation revealed a significant shift of the camera projection centre, 200um and 50um, respectively, and also detected defaults in the flatness of the vacuum plate in the range of -45um to +17 um. The latter was compensated by introducing additional non-radial deformation parameters as depicted in Figure 5.

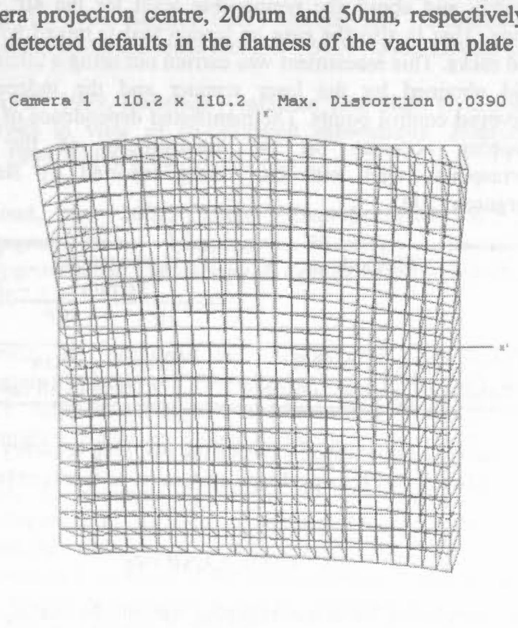


Figure 5. Non-radial deformations due to the flatness defaults of the vacuum plate.

4.2 Sensor Placement

The relative displacements among the individual components of the sensor block were measured terrestrially in the image frame and with sub-centimetre precision. A good knowledge of these parameters subsequently produces better boresight calibration and also allows reducing the size of a Kalman filter for the leaver-arm components in GPS/INS integration.

4.3 Boresight Calibration

The determination of the relative orientation displacement between the camera's image frame and the IMU's body frame requires the use of a well-designed block of images of a strong geometry. As the usual implementation of the AT concept results in a strong correlation between the parameters of exterior orientation, the introduction of GPS/INS observations for the camera perspective centre positions into the bundle adjustment is recommended (Skaloud, 2002). Such approach also permits the concurrent calibration of the parameters of interior orientation (Schmitz et al., 2001).

The evolution of the estimated boresight angles is depicted in Figure 6. This method also provides good indication of the relative orientation accuracy of the navigation system when the AT-estimated orientation is significantly more precise. The 1-σ values of ~0.01deg are certainly fulfilling the demanding expectations.

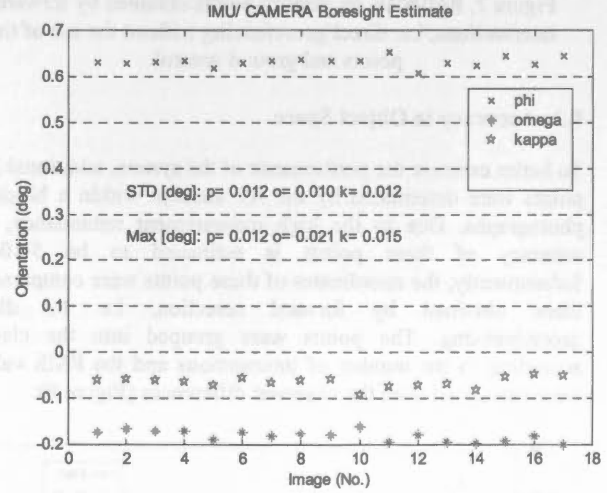


Figure 6. The evolution of the boresight parameters.

5. PRACTICAL EXPERIENCES

The results of snow surface restitution by photogrammetric means were reported previously (Vallet et al., 2000). The following evaluation will therefore focuses on the system absolute accuracy at discrete points. A test field of about 25 Ground Control Points (GCP's) will serve for this purpose.

The scale of the images that were taken over this test field varies between 1:3000-1:4000 and the accuracy of ground control points is at 2cm level. Some GCP's are not specially signalised and therefore the measurement of their image coordinates may introduce additional error between 7-20um (i.e. 3-8cm in the object space).

5.1 Residuals in Object Space

Figure 7. depicts the 3D residuals on all ground control points when neither GCP nor tie-points are used. In other words, the EO parameters for each image are used as supplied by the navigation system and only forward intersection is carried out. It can be concluded from the plot that about 95% of the residuals lies between the ±0.15cm and ±25cm in the horizontal and vertical components respectively. The size of the residuals depends partially on the number of intersections implemented and will be further analysed.

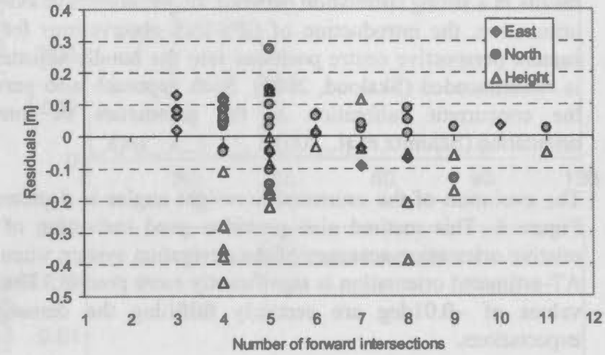


Figure 7. Residuals on control points obtained by forward intersections, i.e. direct georefencing without the use of tie points and ground control.

5.2 Accuracy in Object Space

To better estimate the performance of the system, additional 300 points were determined by the AT concept within a block of photographs. Due to the high measurement redundancy, the accuracy of these points is estimated to be 5-10cm. Subsequently, the coordinates of these points were compared to those obtained by forward resection, i.e. by direct georeferencing. The points were grouped into the classes according to the number of intersections and the RMS values were calculated from the observed differences (Figure 8).

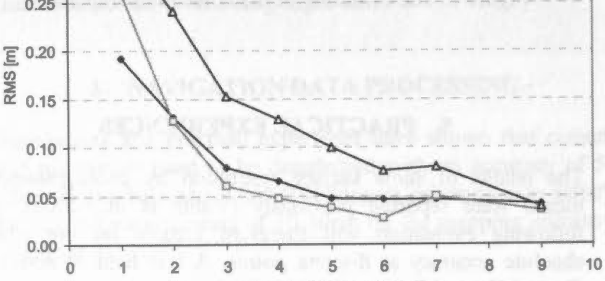


Figure 8. Direct georeferencing accuracy as a function of number of intersections.

As can be seen from this plot, the accuracy improves by a factor of two when moving from the minimum of 1 to 3 intersections (i.e. from 30cm to 15cm). It is also interesting to observe that the higher number of intersections does not longer improve the horizontal accuracy, as the precision of our reference is reached. This takes slightly longer for the vertical component that is influenced by the weaker geometry of the intersecting arrays.

5.3 Comparison to AT

The indirect (AT, AT/GPS) and direct (GPS/INS) approaches to photogrammetric mapping are compared in Table 1 in terms of empirically estimated accuracy. Although, the RMS values for the direct method are slightly bigger they still comply with the demand on the system to deliver ≤20cm mapping accuracy. Adopting the direct approach, however, avoids the difficulties encountered in performing the automated AT in mountainous areas. This considerably increases the operational flexibility for natural disaster mapping.

Method	Constrains		RMS at GCP's [cm]		
	GCP	Block	East	North	Height
AT	●	●	5	4	4
AT/GPS		●	6	6	8
Direct			8	9	20

Table 2. Comparison of empirically obtained mapping accuracy for different approaches to EO determination as well as their operational constrains.

5.4 Comparison to a Laser Scanning System

The mapping accuracy of the proposed method is practically independent from the slope angle as the direction of camera can be adjusted accordingly. Such an optimal orientation is lost when mapping a highly inclined terrain with LIDAR. Table 3 illustrates the decrease in LIDAR mapping accuracy as a function of the steepness of the slope and the homogeneity of the terrain. We conclude that if the accuracy of a laser scanner on a homogenous slope of 10° is satisfying, the errors rise quickly and above the permissible level for the 40° inclined slope. That is also the case on terrain that is mixed with snow and rocks. This assessment was carried out using a 20cm terrain grid obtained by the laser scanner and the independently surveyed control points. The manifested dependence of LIDAR mapping accuracy on the homogeneity of the terrain corresponds well with the values reported by Berg and Ferguson, 2001.

Method	Homogenous Terrain		Mixed Terrain
	10°	40°	
LIDAR, GPS/INS	17cm	30cm	70cm
Handheld Camera, GPS/INS	≤20cm in all cases		

Table 3. Comparison between the mapping accuracy of a LIDAR and the developed system on different terrain.

6. CONCLUSIONS

The developed handheld mapping system for natural hazard mapping is now operational and following conclusions can be drawn with respect to its characteristics:

- The system can be quickly deployed into inaccessible areas as its concept is independent of a carrier.
- Use of GPS/INS onboard provides an adequate accuracy in the object space without any ground control or tie points.
- The terrain is mapped with a consistent accuracy of ≤20cm regardless is steepness and character.

When compared to LIDAR, the constructed system seems to be better suited for small mapping tasks in mountainous regions in terms of accuracy, flexibility and cost.

ACKNOWLEDGMENTS

Flight time and expertise was supplied by Air Glacier Switzerland. Their support and friendliness is greatly acknowledged. The Tomtecs Corp. is thanked for providing the camera and Altex Corp. for offering their laser scanning services free of charge. Many thanks also belong to Mr. Didier Jacquemettaz for his excellent work in the field end the evaluation of the control data. Finally, we thank Dr. W. Ammann, Dr. V. Gruber and Mr. F. Dufour of the Swiss Federal Institute for Snow and Avalanche Research for their active participation in this research grand.

REFERENCES

Applanix Corp., 2000. POSProc User Manual v2.1, 85 Leek Crescent, Ontario, L4B 3B3, http://www.applanix.com.

Berg, R., Ferguson, J. 2001. Mapping Ontario's Highways with Lidar, *GIM Inernational*, Vol. 15, Issue 11, November 2001, pp. 12-15.

DRC, 2001. VME-NAV User Manual (Force 5 version), Dynamic Research Corporation, Test Equipment Division, 93 Border Street, West Newton, MA 02165, April 2001, http://www.drc.com.

Kruck, E., 2001. Bingo-F User's Manual, GIP - Gesellschaft für Industriephotogrammetrie mbH, D-73430 Aalen, Germany.

Issler, D. ,1999. European Avalanche Test Sites. Overview and Analysis in view of co-ordinated experiments. Mitteilungen #59, 1999. SLF Davos.

Skaloud, J., 2002. Direct Georeferencing in Aerial Photogrammetric Mapping, PE&RS, *Journal of Photogrammetric Engineering & Remote Sensing*, March 2002, pp. 207-210.

Schmitz, M., Wübbena, G., Bagge, A., 2001. Benefit of Rigorous Modeling of GPS in Combined AT/GPS/IMU-Bundle Block Adjustment, OEEPE-Workshop on Integrated Sensor Orientation, Hannover, Germany, Hannover, Germany.

Vallet, J., Skaloud, J., Koelbl, O., Merminod, B., 2000. Development of a Helicopter-Based Integrated System for Avalanche Mapping and Hazard Management, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Amsterdam, Vol. XXXIII, Part B2, pp. 565-572

Waypoint Consulting Inc., 2001. GrafNav/GrafMov Operating Manual v603a, Suite 210, 200 Rivercrest Drive S.E., Calgary, Alberta T2C 2X5, August 2001, http://www.waypnt.com.

RIGOROUS GENERATION OF DIGITAL ORTHOPHOTOS FROM EROS A HIGH RESOLUTION SATELLITE IMAGES

Liang-Chien CHEN and Tee-Ann TEO

Center for Space and Remote Sensing Research, National Central University, 320 Chung-Li, TAIWAN
Tel: +886-3-4227151ext7622 Fax: +886-3-4255535 Email: lcchen@csrsr.ncu.edu.tw

Commission IV, WG IV/7

KEY WORDS: EROS A Satellite Images, Orbit Adjustment, Least Square Filtering, Orthorectification

ABSTRACT:

As the resolution of satellite images is improving, the applications of satellite images become widespread. Orthorectification is an indispensable step in the processing for satellite images. EROS A is a high resolution imaging satellite. Its linear array pushbroom imager is with 1.8meter resolution on ground. The satellite is sun-synchronous and sampling with asynchronous mode. The main purpose of this investigation is to build up a procedure to perform orthorectification for EROS A satellite images. The major works of the proposed scheme are:(1) to set up the transformation models between on-board data and respective coordinate systems, (2) to perform correction for on-board parameters with polynomial functions, (3) to adjust satellite's orbit using a small number of ground control points, (4) to fine tune the orbit using the Least Squares Filtering technique and, (5) to generate orthoimage by using indirect method. The experiment includes validation for positioning accuracy using ground check points.

1. INTRODUCTION

The generation of orthoimages from remote sensing images is an important task for various remote sensing applications. Nowadays, most of the high resolution satellites are using linear pushbroom arrays, such as IKONOS, Quickbird, EROS and others. From the photogrammetric point of view, base on the collinearity condition equations; bundle adjustment may be applied to model the satellite orientation (Guaan and Dowman 1988, Chen and Lee 1993). This approach needs a large number of ground control points (GCPs). Chen and Chang (1998) used on-board data and a small number of GCPs to build up a geometric correction model for SPOT satellite images. Similarly, we will propose to use the on-board orbital parameters and GCPs to calibrate the satellite orbit. After orbit modelling, we will develop an indirect method to do the image orthorectification.

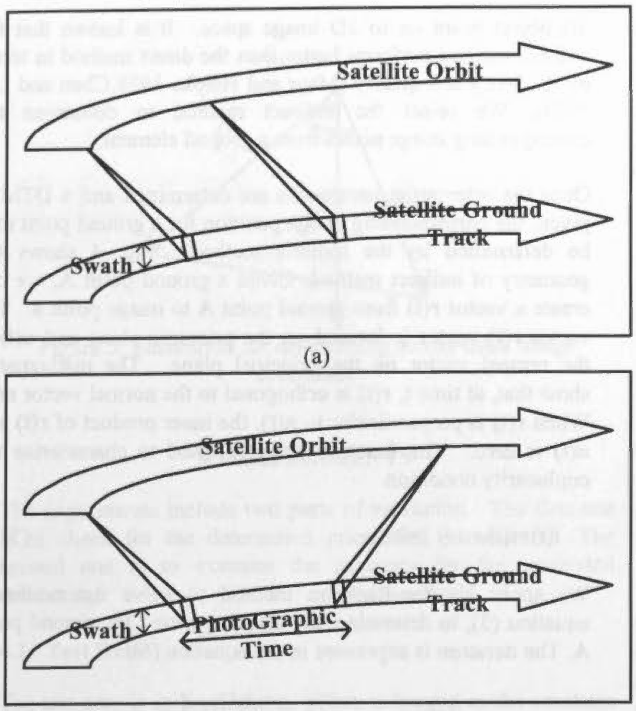
2. CHARACTERISTICS OF EROS A SATELLITE

EROS A was launched by ImageSat International (ISI) on the 5th of December, 2000. It is expected to have at least four years of lifetime. Its orbit altitude is 480km with 97.3 degrees orbit inclination, which make the satellite sun-synchronous. Using its body rotation function, the satellite is able to turn up to 45 degrees in any direction. Its linear array pushbroom imager is with 1.8meter resolution on ground and 1.5degree of field of view. Table 1 shows the characteristics of EROS A satellites.

EROS A satellite takes images with asynchronous mode. It allows the satellite to move in a faster ground speed than its rate of imaging. The satellite actually bends backwards to take its images in an almost constant, predetermined angular speed, enabling its detectors to dwell longer time over each area. In this way, it will be able to get lighter, and improve contrast and conditions for optimal imaging. Fig 1(a) illustrates the synchronous mode vs. asynchronous one. Referring to fig. 1(b), the satellite orbit is longer than the sampling area. In the best condition, the ratio of satellite orbit to sampling area is about 1 to 5

Table I. The characteristics of EROS A satellites

ITEM	EROS A Specification
Orbit Altitude	480km
Orbit Inclination	97.3 Deg
Orbit Pass rate	15.3 orbits/day
Body Rotation	Yes
Mode of Operation	Push Broom Scanning
Scanning Mode	Asynchronous (up to 750 lines/sec)
Stereo pairs	In Track, Cross Track
Sensor Type	CCD
Swath Width	12.5km
Ground Sampling Distance	1.8m (PAN)
Focal Length	3.4 m
Slant Angles	45 Deg
Field of View	1.5 Deg
Pixels-in-line	7800
Spectral Band	Panchromatic:0.5 to 0.9 μm
Sampling Depth Transmitted	11 Bits



(a)

(b)

Figure 1. Illustration for scanning modes (a) Synchronous (b) Asynchronous

3. METHODOLOGY

The proposed method comprises two major parts. The first part is to build up the satellite orientation by using the ground control points. The second one is to use the orbit parameters to perform the orthorectification.

3.1 Orientation Modeling

The major step in validating the positioning accuracy for an image is to model the orbit parameters and the attitude data. Once those exterior orientation parameters are modeled, the corresponding ground coordinates for an image pixel can be calculated. Due to the extremely high correlation between two groups of orbital parameters and attitude data, we only correct the orbital parameters. That means, we will use the attitude information in the on-board ephemeris data as known values.

Three steps are included in this investigation. The first step is to initialize the orientation parameters using on-board ephemeris data. We then fit the orbital parameters with second-degree polynomials using GCPs.

Once the trend functions of the orbital parameters are determined, the fine-tuning of an orbit is performed by using Least Squares Filtering (also called "Least Squares Collocation") technique.

3.1.1 Initialization Of Orientation Parameters: The on-board ephemeris data and GCPs are in different coordinate systems. Before the orbit adjustment, it is essential to build up the coordinate transformation, so that the orbit adjustment will be performed in WGS84 as unified coordinate system. Those coordinate systems include inertial frame, WGS84, GRS67, Geodetic Coordinate System, TWD67, Orbital Reference Frame and EROS A Body Frame. In which, GRS67 used the Kaula ellipsoid (a=6378160m, f=1/298.247). The Geodetic Coordinate System is in longitude and latitude. Based on the datum of GRS67, the TWD67 system is a Transverse Mercator Projection using λ =121E as the central meridian. There will be three steps between TWD67 and WGS84 transformation. First, we project TWD67 into the Geodetic Coordinate System, then the Geodetic Coordinate System is projected into GRS67. Finally, the GRS67 system is transformed into WGS84. The camera model was provided by ISI International.

3.1.2 Preliminary Orbit Fitting: Because the on-board data includes errors to a certain degree, GCPs are needed to adjust the orbit parameters. Referring to fig. 2, the observation vector (Ua) provided by the satellite will not pass through the corresponding GCP due to errors the on-board data. Thus, correction of the orbit data from (x_0, y_0, z_0) to (x, y, z) may be performed under the conditions

$$X_i = x(t_i) + S_i u_{xi} \quad (1a)$$

$$Y_i = y(t_i) + S_i u_{Yi} \quad (1b)$$

$$Z_i = z(t_i) + S_i u_{Zi} \quad (1c)$$

$$x(t_i) = x_0 + a_0 + a_1 \cdot t_i + a_2 \cdot t_i^2 \quad (2a)$$

$$y(t_i) = y_0 + b_0 + b_1 \cdot t_i + b_2 \cdot t_i^2 \quad (2b)$$

$$z(t_i) = z_0 + c_0 + c_1 \cdot t_i + c_2 \cdot t_i^2 \quad (2c)$$

where

X_i, Y_i ,Z_i are object coordinates of the ith GCP,
u_{xi}, u_{yi} ,u_{zi} are components of the observation vector,
$x(t_i)$, $y(t_i)$,$z(t_i)$ are the satellite's coordinates of the ith GCP after correction,
x0,y0 ,z0 are the satellite's coordinates before correction,
a_i, b_i and c_i (i=0,1,2) are coefficients for orbit corrections,
t represents sampling time, and
S_i is the scale factor.

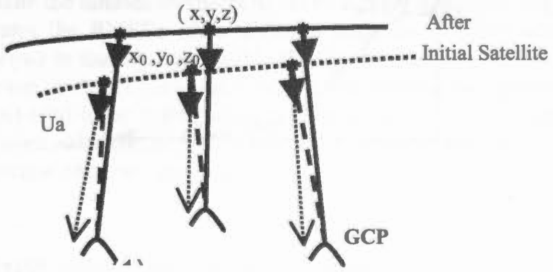


Figure 2. Preliminary fitting for satellite orbit

3.1.3 Least Squares Filtering: Because the least squares adjustment is a global treatment, it cannot correct for the local errors. Therefore, the least squares filtering (Mikhail and Ackermann, 1982) has to be used to fine tune the orbit. By doing this, we assume that the x, y, z-axis are independent. Thus, we use three one-dimensional functions to adjust the orbit. The model of least squares filtering is shown as

$$\rho_k = \vec{V}_k \, [\Sigma_k]^{-1} \, \vec{\varepsilon}_k \tag{3}$$

where

k is x,y,z axis

ρ_k is the correction value of the interpolating point ,

$\vec{V}_k$ is the row covariance matrix of the interpolating point with respect to GCPs,

Σ_k is the covariance matrix for GCPs, and

$\vec{\varepsilon}_k$ is the residual vectors for GCPs.

The basic consideration in this investigation is to minimize the number of required GCPs. Thus, using a large amount of GCPs to characterize the covariance matrix is not practical. In this paper, we use a Gaussian function (shown as fig. 3) with some empirical values as the covariance function. The Gaussian function is shown as

$$Covariance = \begin{cases} c e^{-(2.146 \frac{d}{d_{max}})^2} & , \ if \ d \neq 0 \\ \mu_k & , \ if \ d = 0 \end{cases} \tag{4}$$

$$c = (1 - r_n)\, \mu_k$$

where

d is the distance between an intersection points and a GCP,

d_{max} is the distance between an intersection point and the farthest GCP,

μ_k is the variance of GCPs' residual, and

r_n is the filtering ratio, in which we use r_n=0.1 in experiment.

The empirical value 2.146 is selected so that the covariance limit is 1% of 0.01(1- r_n)μ_k when d=dmax (Chen & Chang, 1998).

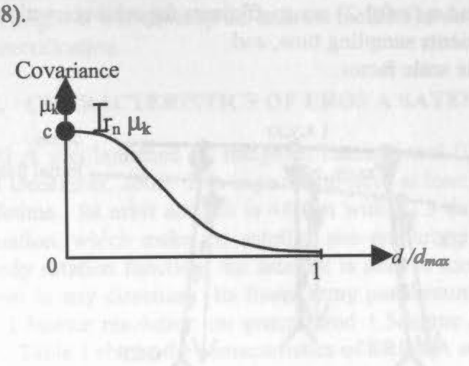


Figure 3. Covariance function of least squares filtering

3.2 Orthorectification

There are two ways to do the orthorectification. The first one is the direct method. A technique called Ray-Tracing (O'Neill and Dowman, 1988) was developed to solve the problem by direct method. It projects a 2D image point on to a 3D object model. The second one is indirect method, which projects the 3D object point on to 2D image space. It is known that the indirect method performs better than the direct method in terms of efficiency and quality (Mayr and Heipke 1988,Chen and Lee 1993). We select the indirect method to determine the corresponding image pixels from a ground element.

Once the orientation parameters are determined and a DTM is given, the corresponding image position for a ground point may be determined by the indirect method. Fig. 4 shows the geometry of indirect method. Given a ground point **A**, we can create **a** vector **r(t)** from ground point **A** to image point **a**. The vector **r(t)** vector is located on the principle plane and **n(t)** is the normal vector on the principal plane. The mathematics show that, at time t, **r(t)** is orthogonal to the normal vector **n(t)**. When **r(t)** is perpendicular to **n(t)**, the inner product of **r(t)** and **n(t)** is zero. The function **f(t)** is defined to characterize the coplanarity condition

$$f(t) = r(t) \cdot n(t) = 0 \tag{5}$$

We apply Newton-Raphson method to solve the nonlinear equation (5), to determine the sampling time t for ground point **A**. The iteration is expressed in the equation (6),

$$t_{n+1} = t_n - \frac{f(t_n)}{f'(t_n)} = t_n - \frac{f(t_n)}{[f(t_n + \Delta t) - f(t_n - 2\Delta t)]/2\Delta t} \tag{6}$$

when n=0,1,2….. until $|t_{n+1}-t_n|<10^{-5}$ sec is satisfied.
For an image point sampled at time t, we can decide a principal plane, the along track image coordinate can be calculate by

$$Line=(t-t0)/(integration\ time) \tag{7}$$

where t0 is sampling time for the first scan line. Integration time is the sampling interval.

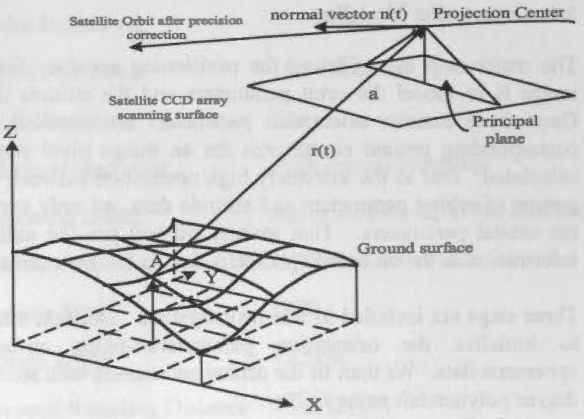

Satellite Orbit after precision correction
normal vector n(t)
Projection Center
Satellite CCD array scanning surface
a
Principal plane
Z
r(t)
Ground surface
A
Y
X

FIGURE 4. Illustration of indirect method

The across track image coordinate may be determined, as shown in fig. 5. In the figure, V_f is the pointing vector of first CCD in line, and V_l is the pointing vector of last CCD in line. The across track coordinate for the pixel is

$$Sample=(S/FOV)*7043 \tag{8}$$

where

S is the angle between V_f and **r(t)**.

FOV is field of View angle, and

7043 is number pixel in a line.

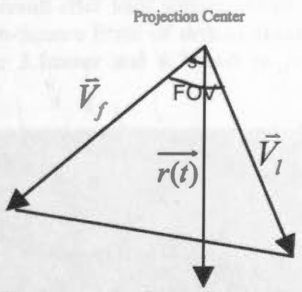


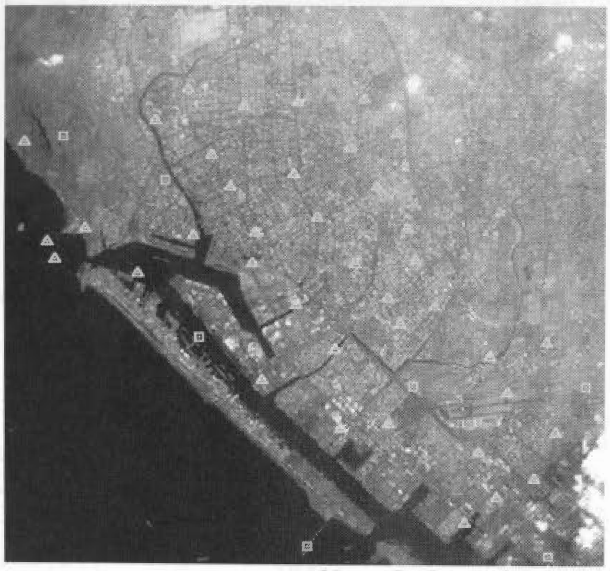
Figure 5. Illustration for determining across-track image coordinate

4. EXPERIMENTS

The experiments include two parts of validation. The first one is to check for the determined orientation parameters. The second one is to examine the accuracy for the generated orthoimage

4.1. Test Data

The test area is in KaoHsiung, which is located in the southern part of Taiwan as shown in fig.6. Scene ID is TAW1-e1019903, which were sampled on Apr. 15, 2001. The asynchronous ratio of the satellite orbits to sampling area is 1:13. The GCPs and check points (CHKPs) were measured from 1:1000 scale topographic maps. The position accuracy is better than 50 centimeters. The distributions of those points are shown in fig. 6. In the figure, triangles represent the GCPs while boxes are the CHKPs. The total number of GCPs and CHKPs are 53. Other related information is shown in Table II.



©ImageSat International 2001.

Figure 6. The test image.

Table II. Related information of test images

ITEM	Parameters
Scene ID	TAW1-e1019903
Date	2001/04/15
Integration Time	3.7msec
Ground Sampling Distance	1.90m
Test Area	13.38km * 12.48km
Image size	7043* 6572 pixel
Place	KaoHsiung, Taiwan
Pointing Angle	11.60 Deg
Orbit Arc	170KM(about 1:13)

The DTM used in the orthorectification was acquired from the Topographic Data Base of Taiwan. The pixel spacing of DTM is resample from 40m to 2m. Fig. 7 illustrates the terrain variation. The elevation ranges from 0m to 340m.

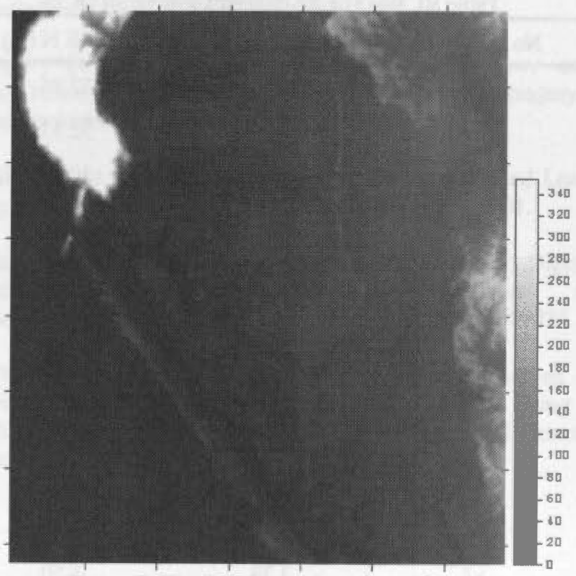


Figure 7. The DTM used in orthorectification

4.2. Accuracy VS. Number of GCPs

The Ray-Tracing method is applied to evaluate the orbit accuracy. Given the satellite orientation and image point, we calculate the intersection point of DTM and ray direction. Fig.8 indicates the RMSEs when different numbers of GCPs were employed in the test data. Table III lists the figures to indicate the trend in detail. It is obvious that the RMSEs i.e., (3.53m, 4.70m) tend to be stable when 9 or more control points were employed. Notice that the coordinate system is in TWD67 with Transverse Mercator projection.

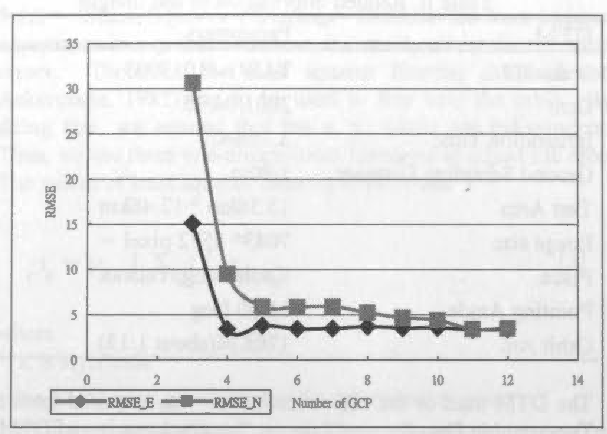


Figure 8. RMSEs for the different number of GCPs

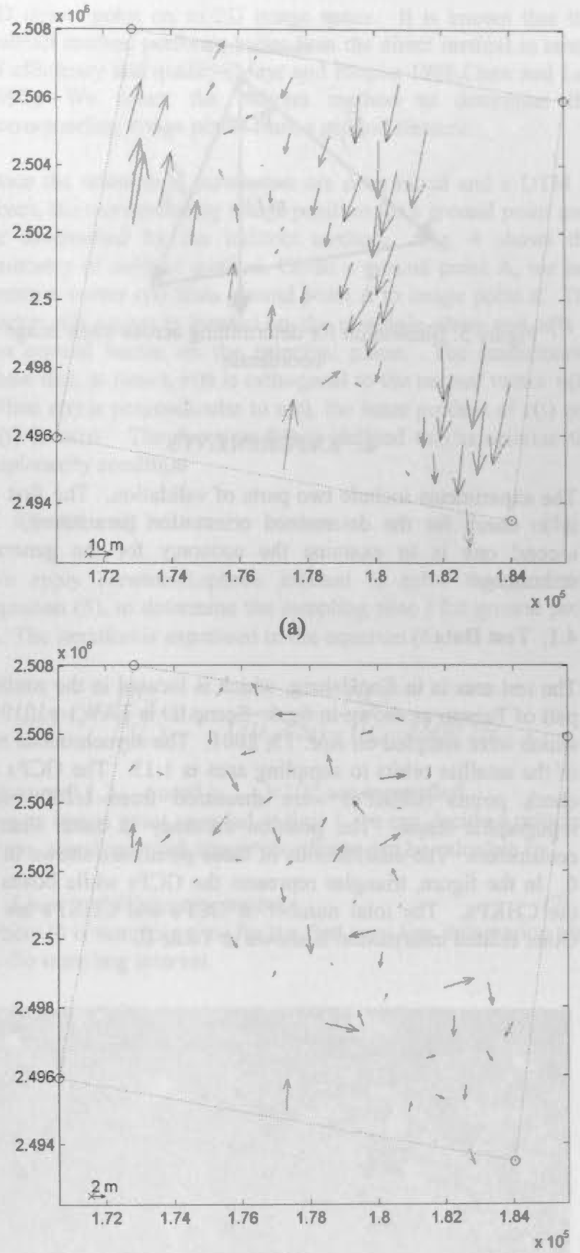
Table III. RMSEs for different number of GCPs

No. of GCPs	RMSE E (m)	RMSE N (m)
1	22.17	37.85
2	10.96	32.31
3	15.11	30.64
4	3.41	9.51
5	3.80	5.90
6	3.45	5.86
7	3.48	5.90
8	3.70	5.24
9	3.53	4.70
10	3.55	4.42
11	3.34	3.43
12	3.38	3.50

4.3. Accuracy Analysis Of Orbit Modeling

We further evaluate the error behaviours in the two different phases. Fig.9(a) depicts the error behaviour after preliminary orbit fitting in TWD67 coordinate system. We could see that the system errors are obvious. Fig.9(b) illustrates the results after precision correction, i.e., least squares filtering. The coordinate system is also in TWD67. After using least squares filtering to fine tune the orbit, the major system errors have been compensated.

We provide Table III for the summary of accuracy. Table IV illustrates the accuracy performance of GCPs and CHKPs, when 9 GCPs were employed. After preliminary orbit fitting, the RMSE of CHKPs is about 6meter and 24meter in two directions. After least squares filtering, the RMSE of CHKPs are reducing to 3.3meter and 4.3meter respectively.



(a)

(b)

Figure 9.Error vectors of orbit modeling (a)Error vectors of the preliminary orbit fitting (b) Error vectors after least squares filtering

Table IV. Root-Mean-Square Error of orbit modeling

	RMSE E (meter)	RMSE N (meter)
Preliminary orbit fitting		
GCPs (9)	6.17	30.47
CHKPs (44)	5.63	23.57
Least Squares Filtering		
GCPs (9)	1.47	2.72
CHKPs (44)	3.34	4.37

4.4. Accuracy Evaluation Of Orthorectification

The generated orthoimage is shown in fig. 10. In order to evaluate the quality of orthoimage, we check it manually. Fig. 11 illustrates that the RMSE of ground check point is slightly better than 2 pixels. It is observed that the error vectors are similar to the result after least square filtering. Table V shown the Root-Mean-Square Error of orthoretification. The RMSEs of CHKPs are 3.1meter and 3.7meter in E and N directions respectively.

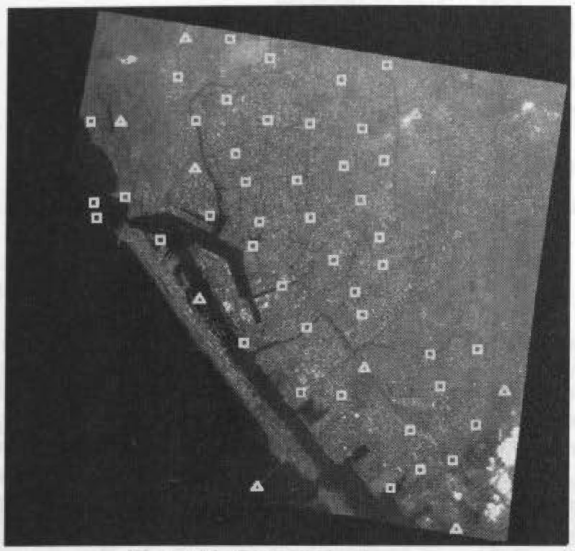


Figure 10. Generated orthophoto

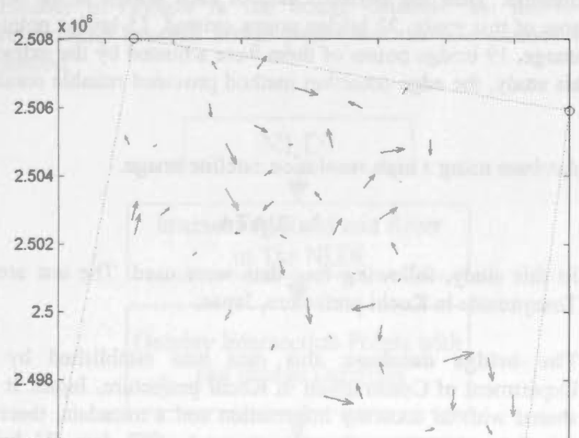


Figure 11.Error vectors for the generated orthophoto

Table V. Root-Mean-Square Error of orthoretification

	Orthorectification	
	RMSE E (meter)	RMSE N (meter)
GCPs (9)	1.80	3.10
CHKPs (44)	3.13	3.74

5. CONCLUSIONS

We have proposed a procedure to perform geometric correction for EROS A satellite images using a small number of GCPs. The corrections for orbital data are modeled as functions of time. The GCPs are applied to correct the on-board data to maintain the geometrical relationship between image space and object space. After that, we use least squares prediction to fine-tune the orbit. Finally, we use indirect method to generate the orthophotos. Experimental results indicate that the proposed scheme may reach an accuracy of better than two pixels in the image scale for an image sampled with an asynchronous ratio of 13. Because of the measurement error, the results of orthorectification and least squares filtering are slightly different. The DTM used in the orthorectification was resampled from the one with 40m resolution. Due to its error, the precision of the proposed method could be underestimated.

6. REFERNENCES

Chen, L.C., and Chang, L. Y., 1998, "Three Dimensional Positioning Using SPOT Stereostrips with Sparse Control", Journal of Surveying, ASCE 124(2): pp.63-72.

Chen, L.C., and Lee, L.-H. 1993, "Rigorous Generation of Digital Orthophoto from SPOT Images." Photogrammetric Engineering and Remote Sensing, 59(5), 655-661.

Gugan, D.J., and Dowman, I. J. 1988, "Accuracy and completeness of topographic mapping from SPOT imagery." Photogrammetric Record, 12(72), 787-796.

Mikhail E.M and F. Ackermann, 1982, Observation and Least Squares, University Press of America, New York, pp.401.

Mayer, W. and C. Heipke, 1988, "A Contribution to Digital Orthophoto Generation", International Archives of Photogrammetry and Remote Sensing, 27(B11):IV430-IV439.

O'Neill, M.A., and Dowman, I. J., 1988, "The Generation of Epipolar Synthetic Stereo Mates for SPOT Images Using a DEM", International Archives of Photogrammetry and Remote Sensing, Kyoto, Japan, 27(B3): pp.587-598.

AN ACCURACY ADJUSTMENT BY FUSION METHOD WITH GIS DATA AND REMOTE SENSING DATA

Jong Hyeok JEONG*, Takagi MASATAKA**
*Master Student, Department of Infrastructure System Engineering
**Associate Professor, Department of Infrastructure Systems Engineering
Kochi University of Technology
*045032w@gs.kochi-tech.ac.jp
**takagi@infra.kochi-tech.ac.jp

Working Group IV/7

KEY WORDS: Accuracy Adjustment, Data Fusion, IKONOS

ABSTRACT:

The Department of Construction in Kochi prefecture, Japan established and shared a bridge database without accuracy information. When the data overlay with a commonly used GIS data, some errors appeared. To adjust the accuracy of the bridge database, the intersections of roads and rivers of the National Land Digital Information (NLDI) of Japan were used in the accuracy adjustment, and 64% of bridges over rivers obtained accuracy adjustment (Jong Hyeok JEONG, Masataka TAKAGI, 2001). In the adjustment, 36% bridges could not be adjusted because of the limitation of the NLDI. The NLDI did not have enough information for the adjustment. To cover the limitation of the NLDI in the accuracy adjustment of the bridge database, an IKONOS satellite image was used. Expected bridges were extracted from an ISO unsupervised classification of the IKONOS image and edge detection of the IKONOS image, and then the extracted bridges were converted to polygons. Each bridge polygon in the bridge database matched with the nearest extracted bridge point to make a one-to-one relationship. Then the attribute data of each bridge in the bridge database added to the coordinate of the nearest polygon. In the test area of this study, 23 bridge points existed, 15 bridge points of them were adjusted by the extracted bridge position from classified image, 19 bridge points of them were adjusted by the extracted bridge position from the edge detection of the IKONOS image. In this study, the edge detection method provided reliable result of accuracy adjustment.

1. INTRODUCTION

GML was established by Open GIS Consortium and G-XML was established by Data Promotion Center in Japan. They are XML based GIS data format, and were adapted as a standard GIS format. Nowadays, major GIS softwares support the XML based GIS data. XML based GIS data format provides various channels for GIS users to access a GIS data, and it enables to use various GIS. Since the advantages of XML based data format, companies, governments and organizations started to convert their GIS data into XML based GIS data. Currently, various GIS data are based on different scales. When those GIS data are overlaid with other GIS data in different scales, many problems will be encountered in the accuracy domain.

The Department of Construction in Kochi prefecture, Japan established and shared a bridge database without accuracy information and a metadata. When the bridges on river in the bridge database overlaid with road data and river data in a commonly used GIS data, the bridge points did not matched with the intersections of them.

To adjust the positional accuracy of the bridge database, the attribute data of each bridge point added to the coordinate of the nearest intersection of a road data and a river data in a reference GIS data (Jong Hyeock JUNG, Masataka TAKAGI, 2001). But in the result of the accuracy adjustment, 36% bridges could not be adjusted. The reason came from the low accuracy of the bridge database and the limitation of the reference data. In case of used reference data based on a 1:25000 scale map, narrow river streams and roads were not described. Therefore enough intersections were not generated. The objective of this investigation is positional accuracy adjustment for the bridge database using a high-resolution satellite image.

2. DATA

In this study, following four data were used. The test area is Tosayamada in Kochi prefecture, Japan.

The bridge database: this data was established by the Department of Construction in Kochi prefecture, Japan. It was shared without accuracy information and a metadata, therefore this data was assumed as an uncertain GIS data. 23 bridge points of the bridge database existed and used on the test area.

The National Land Digital Information (NLDI): this data was established by the Ministry of Land, Infrastructure and Transport, its data type is line, and it based on a 1: 25000 scale map, used as a reference data in this study, the intersections of roads and rivers of the NLDI were used.

The Disaster Prevention Information (DPI): Kochi prefecture government in Japan established this data, and it based on a 1: 2500 scale map, its data type is polygon, road and river data were used for generating training area for land cover classification and image processing.

The IKONOS image: This image was composed of 4 spectral bands including infrared band, and it was used for extraction of bridge positions. The coverage of this image is the test area in this study.

3. METHODOLOGY

For the accuracy adjustment, a data fusion method was used. The data fusion method is combining the bridge database with a high accuracy reference data. In this study, bridge positions in the bridge database were adjusted by following three methods.

- To extract expected bridge positions by intersecting of the roads and rivers of the NLDI.
- To extract expected bridge positions from ISO unsupervised classification of the IKONOS image
- To extract expected bridge positions from the edge detection of the IKONOS image.

3.1 Bridge Extraction with The NLDI

In the case of accuracy adjustment of the bridge database with the NLDI, the intersections of roads and rivers were used. Figure 1 shows the flow chart of the accuracy adjustment of the bridge database with the NLDI.

Firstly, the intersections of roads and rivers in the NLDI were generated. The intersections were assumed expected bridges. Not all bridges on the test area recorded in the bridge database. Because the only bridges managed by government were recorded. The number of intersection was larger than the number of bridges in the bridge database. So one-to-one relationship between bridge points and the intersections were needed.

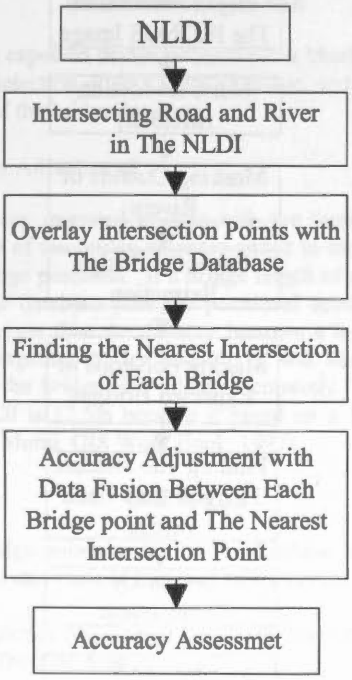

NLDI
Intersecting Road and River in The NLDI
Overlay Intersection Points with The Bridge Database
Finding the Nearest Intersection of Each Bridge
Accuracy Adjustment with Data Fusion Between Each Bridge point and The Nearest Intersection Point
Accuracy Assessmet

Figure 1. The flow chart of the accuracy adjustment of the bridge database with the NLDI

To make one-to-one relationship between bridge points and the intersections, the nearest intersection of each bridge point was searched. The attribute of each bridge added to the coordinate of the nearest intersection, and then they were considered as expected bridge position.

3.2 Bridge Extraction with ISO Unsupervised Classification of The IKONOS Image

The ISO unsupervised classification method was used to classify the land cover of the test area on the IKONOS image that was composed of four spectral bands including infrared band. After the classification, 256 classes were generated (Figure 2).

The classification result overlaid with polygon of road and river data in the DPI. Then the histogram of classified categories on roads and rivers were generated to find unique categories of bridges in the IKONOS image. In this study, category number 82,83,84,85,92 could be used as unique categories on road. Then classified roads on rivers were filtered with median filter (3x3) to eliminate small noises. The out sides of rivers were masked by the river data of the DPI to consider only classified roads on rivers. The classified roads on rivers were assumed expected bridge positions. The expected bridges converted to polygons (Figure 4). After that, the polygons of expected bridges overlay with the bridge points in the bridge database (Figure 4).

Figure 5 shows the flow chart of the accuracy adjustment of the bridge database with the ISO unsupervised classification of the IKONOS image.

Figure 2. An example of ISO unsupervised classification result

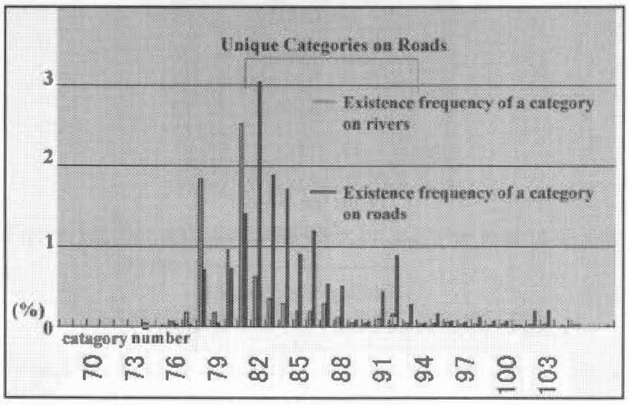


Figure 3.Histogram of classified categories on road and river

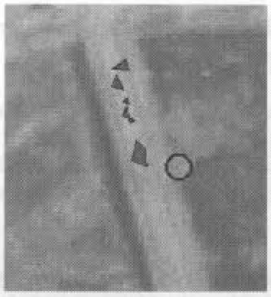


Figure 4. An expected bridge from the ISO unsupervised classification result (red polygons) and a bridge point (green point) in the bridge

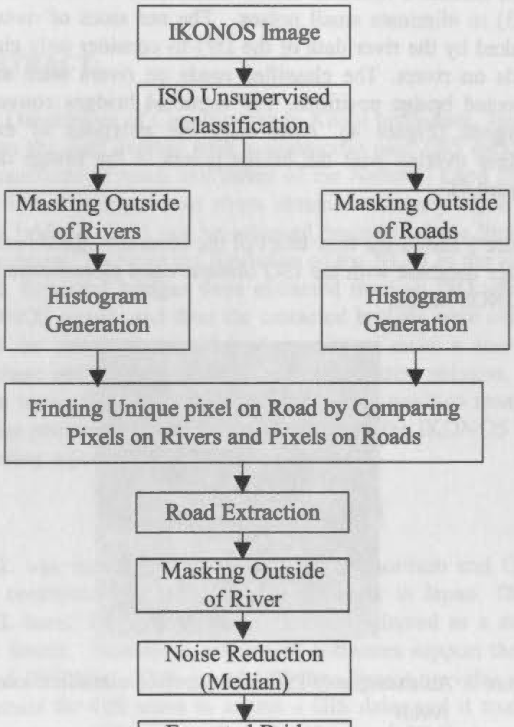

IKONOS Image
ISO Unsupervised Classification
Masking Outside of Rivers
Masking Outside of Roads
Histogram Generation
Histogram Generation
Finding Unique pixel on Road by Comparing Pixels on Rivers and Pixels on Roads
Road Extraction
Masking Outside of River
Noise Reduction (Median)
Expected Bridge Extraction
Making Polygons of Expected Bridges
Overlay Polygon with the Bridge database
Finding The Nearest Polygon from Each Bridge
Accuracy Adjustment
Accuracy Assessment

Figure 5. The flow chart of the accuracy adjustment of the bridge database with the result of ISO unsupervised classification of the IKONOS image

3.3 Bridge Extraction with The Edge Detection of The IKONOS Image

Figure 6 shows the flow chart of the accuracy adjustment of the bridge database with the edge detection of the IKONOS image. To extract the edges, the spatial filter (Table 2) was used because the most bridges on the test area are located from north to south and from northeast to southwest. Figure 7 shows an example of edges around a bridge. After filtering the IKONOS image, the linear feature of edge extracted clearly (Figure 8). After that, the edge detection image on rivers was filtered with median filter to reduce small noises. Then the outsides of the rivers of test area on the IKONOS image were masked to consider only linear feature position on rivers from the edge detection of the IKONOS image. Then the expected bridges were extracted. And they were overlaid with the bridge points of the bridge database (Figure 9).

Table 2. The spatial filter to edge the IKONOS image

0	1	1
0	0	0
-1	-1	0

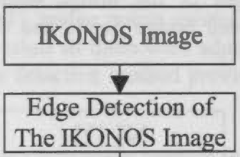

IKONOS Image
Edge Detection of The IKONOS Image
Noise Reduction (Median)
Masking Outside of Rivers
Expected Bridge Extraction
Making Polygons of Expected Bridges
Finding The Nearest Polygon from Each Bridge
Accuracy Adjustment
Accuracy Assessment

Figure 6. The flow chart of the accuracy adjustment of the bridge database with the edged IKONOS image

Figure 7. Band1 of the IKONS image

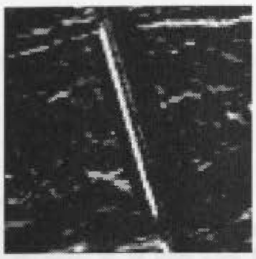


Figure 8. The edged image of IKONOS band1 image

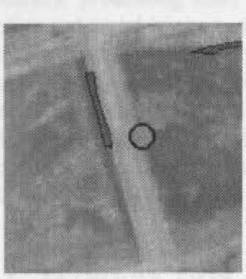


Figure 9. An expected bridge polygon (dark blue) from the edge detection of the IKONOS image, and a bridge point of the bridge database

3.4 Accuracy Adjustment

After extracting expected bridges with the three methods, the attribute data of the bridge database added to the coordinate of expected bridge positions. If a bridge length of a bridge points in the bridge database plus the positional accuracy of NLDI (12.5m) is longer than the distance between a bridge point and the nearest expected bridge position, it was assumed that the accuracy of the bridge was adjusted correctly. The positional error of NLDI is 12.5m because it based on a 1: 25000 scale map (Shunji Murai, GIS Work Book, 1997).

4.RESULTS

Total 23 bridge points in the bridge database were used, and table 3 shows the result of accuracy adjustment.

Table 3. Accuracy Adjustment Results by Each Reference Data on The Test Area

Reference Data	Number of Adjusted Bridge
NLDI	16/23
Classified IKONOS image	15/23
Edged IKONOS Image	19/23

In the result of accuracy adjustment by the three methods, 19 (83%) bridge points on the test area adjusted by the edge detection method. This method provided the most reliable reference data among the three methods.

In the result of the ISO unsupervised classification, bridge and gravel were classified in the same category (Figure 10), therefore bridges on gravel area in a river stream were not extracted.

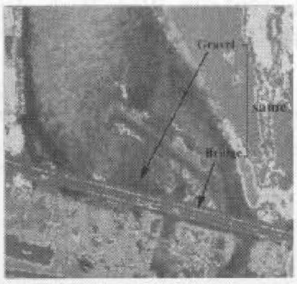


Figure 10. The ISO unsupervised classification of a bridge on a gravel area

On the other hands, bridges are not on a gravel area (Figure 11) in a river stream, they were extracted clearly by the classification method (Figure 12).

Figure 11. A bridge not on a gravel area in a river stream

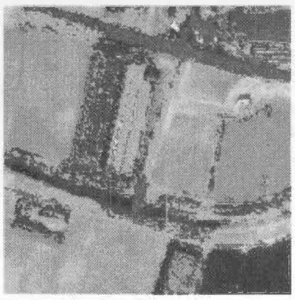


Figure 12. An extracted bridge not on a gravel area in a river stream

5.CONCLUSIONS AND FURTHER STUDY

From this study, the accuracy adjustment of the bridge database by using a high-resolution IKONOS image was successfully achieved, and the method of the edge detection of the IKONOS image provided most reliable reference data among the three methods. Although the result of ISO classification provided low quality of a reference data, bridges were not on a gravel area, they were extracted clearly in a river stream. In the result of extraction, many noises existed in the result of extraction of the expected bridges, and they prevented us from extracting clear bridge positions from the classified IKONOS image and the edge detection of the IKONOS image. As a further study, to extract more clear linear feature of bridges from the edge detection of the IKONOS image, Hough transform will be used. And the result will be compared with the results of other methods.

REFERENCES

Shunji Murai, 1997.GIS Work Book. Japan Association of Surveyors, Japan, p. 35.

Jong Hyeok JEONG, Masataka TAKAGI, *An Accuracy of Uncertain GIS Positional Data by Data Fusion Method,* The proceedings of ACRS 2001, Volume 2, pp.1498-1500.

A NEW EPIPOLARITY MODEL BASED ON THE SIMPLIFIED PUSHBROOM SENSOR MODEL

Hae-Yeoun Lee [a], Wonkyu Park [a]

[a] Image and system department, SaTReC Initiative (SaTReCi),
18F, Sahak Bldg., 929, Dunsan-dong, Seo-gu, Taejon, Korea – (hytoiy, wpark)@satreci.co.kr

Commission IV, WG IV/7

KEY WORDS: Epipolar geometry, Linear pushbroom imagery, Simplified pushbroom model

ABSTRACT:

This paper addresses the epipolar geometry of linear pushbroom imagery. Two images of a single scene are related by epipolar geometry which contain all geometric information and is essential for the three dimensional reconstruction of the scene in computer vision and remote sensing. It is told that the epipolar geometry of the linear pushbroom sensor is different from that of the perspective one. In this paper, we propose an epipolarity model which does not require the ephemeris data but shows high performance (in accuracy, processing time, etc.). We also quantitatively analyse various epipolarity models such as the epipolar geometry of perspective and aerial imagery, the one by Gupta and Hartly and the one based on the Orun and Natarajan sensor model. To analyze the accuracy of the proposed epipolarity model and others, we quantitatively measure the distance between the truth point and epipolar lines on two types of linear pushbroom images; SPOT and KOMPSAT. The results show that the epipolarity model based on the Orun and Natarajan sensor model is more accurate than that of perspective sensor and by Gupta and Hartly because the ephemeris data of the satellite image is applied. The proposed epipolarity model shows a high accuracy similar to that of the Orun and Natarajan sensor model without the ephemeris data. Our epipolarity model will be very useful when the ephemeris data are not available such as IKONOS images or are not accurate.

1. INTRODUCTION

In general, two or more images of a single scene are related by the so-called epipolar geometry. Since the epipolar geometry contains all geometric information that is necessary for establishing correspondence, it is commonly used for the extraction of three-dimensional information in computer vision, photogrammetry and remote sensing.

The epipolar geometry means that a point (*a*) in the image is mapped to the point on the known linear line (epipolar line) or non-linear curve (epipolar curve) in the other image (refer figure 1). In case of aerial and perspective imagery, the epipolar geometry is mathematically well founded and widely used in computer vision and aerial photogrammetry [Zhang, 1998]. In case of linear pushbroom imagery, however, the epipolar geometry is modelled as very complex non-linear equations and, based on the reviews, depends on the sensor model. It is also told, but not proved, that the epipolar geometry of perspective imagery cannot be applied to linear pushbroom imagery [Kim, T. 2000]. Details of the epipolar geometry are described in section 2.

In this paper, we propose a new epipolarity model based on the simplified pushbroom sensor model, proposed by Gupta and Hartly, which does not require the ephemeris data but show high performance (processing time, accuracy, etc.). We also verified the accuracy of the proposed epipolarity model in comparison with other models for linear pushbroom imagery; (1) the epipolarity model of perspective and aerial imagery, (2) the one by Gupta and Hartly and (3) the one based on the Orun and Natarajan sensor model.

For the quantitative analysis of the epipolarity models, we used 20 ground control points, measured using GPS receiver, as modelling points and 30 conjugate pairs, accurately extracted by an experienced operator, as independent checking points on two types of linear pushbroom imagery, SPOT and KOMPSAT stereo image pairs.

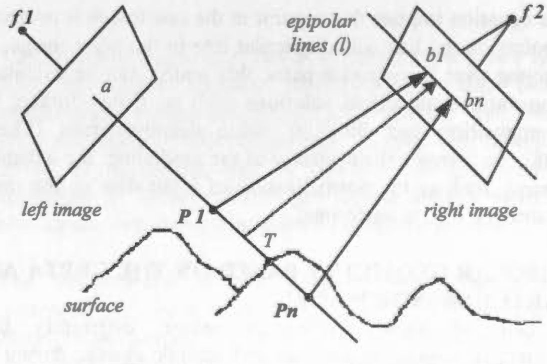


Figure 1. The relation among sensor, image and objects.

Based on the results, the epipolarity model of perspective imagery and by Gupta and Hartly show the mean accuracy below 1 pixel although the error on several checking points was large. The epipolarity model based on the Orun and Natarajan sensor model is more accurate than that of perspective imagery and by Gupta and Hartly. The accuracy of the proposed epipolarity model is considerably high and similar to that of the epipolarity model based on the Orun and Natarajan sensor model although the ephemeris data is not applied. It can be effectively applicable to the imagery which do not provide the ephemeris data such as IKONOS or when the quality of the ephemeris data is low.

In section 2, the epipolarity models for perspective imagery and linear pushbroom ones are reviewed. Section 3 describes our proposed epipolarity model. In section 4, the results of experiments are shown and discussed.

2. VARIOUS EPIPOLARITY MODELS

EPIPOLAR GEOMETRY OF PERSPECTIVE AND AERIAL IMAGERY

The epipolar geometry of perspective and aerial imagery is mathematically well founded and has been extensively studied in computer vision and aerial photogrammetry. In these images, the epipolar geometry is modeled as a 3x3 singular matrix called as a fundamental matrix shown below.

$$(x_r \quad y_r \quad 1)\begin{pmatrix} f_{11} & f_{12} & f_{12} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix}\begin{pmatrix} x_l \\ y_l \\ 1 \end{pmatrix} = 0$$

Where, (x_l, y_l) and (x_r, y_r) is the coordinates of each image points and the matrix has only seven degrees of freedom.

The matrix can be represented as below.

$$k_1 x_r + k_2 y_r + k_3 = 0$$

Where,
$k_1 = f_{11}x_l + f_{12}y_l + f_{13}$,
$k_2 = f_{21}x_l + f_{22}y_l + f_{23}$,
$k_3 = f_{31}x_l + f_{32}y_l + f_{33}$.

This equation implies that a point in the one image is projected to points on the line called epipolar line in the other image. By selecting over 7 conjugate pairs, this matrix can be calculated using various numerical solutions such as, Gauss-Jordan, LU decomposition and Singular value decomposition [Zhang, 1998]. To increase the accuracy of the modelling, the extensive research such as the normalization of input data or the rank2 constraints, etc., is performed.

EPIPOLAR GEOMETRY BASED ON THE GUPTA AND HARTLY SENSOR MODEL

In case of linear pushbroom sensor, differently from perspective sensor, its position and attitude change during the acquisition moment. Hence, its modeling called as a sensor model is very difficult and computationally expensive. By assuming the linear movement and the constant attitude of sensor, Gupta and Hartly propose a simplified pushbroom model[1], which can be calculated from ground control points without the ephemeris data. They also described its epipolarity model.

The epipolar geometry by Gupta and Hartly is represented as a 4x4 singular matrix called as LP fundamental matrix as against that of perspective imagery. This matrix is shown below.

$$(x_r \quad x_r y_r \quad y_r \quad 1)\begin{pmatrix} 0 & 0 & f_{13} & f_{14} \\ 0 & 0 & f_{23} & f_{24} \\ f_{31} & f_{32} & f_{33} & f_{34} \\ f_{41} & f_{42} & f_{43} & f_{44} \end{pmatrix}\begin{pmatrix} x_l \\ x_l y_l \\ y_l \\ 1 \end{pmatrix} = 0$$

Where, (x_l, y_l) and (x_r, y_r) is the coordinates of each image points and it contains no more than eleven degrees of freedom. Similar to that of perspective imagery, the solution can be acquired using over 11 corresponding points.

This matrix can be represented as below.

$$k_1 x_r + k_2 x_r y_r + k_3 y_4 + k_4 = 0$$

Where,
$k_1 = c_1 y_l + c_2$,
$k_2 = c_3 y_l + c_4$,
$k_3 = (c_5 x_l + c_6) y_l + (c_7 x_l + c_8)$,
$k_4 = (c_9 x_l + c_{10}) y_l + (c_{11} x_l + c_{12})$.

From this equation, we can certain that the epipolar geometry by Gupta and Hartly is represented as a hyperbola curve, called as an epipolar curve. This means that a point in the image is mapped to points on the non-linear curve, differently from that of perspective imagery, in the other image. We must note the fact that it is represented as first-order polynomials for along-track and across-track, respectively. Similar to that of perspective imagery, the solution can be calculated from a set of corresponding points using numerical solutions. In general, this non-linear equation is approximated as a piece-wise linear-line for the practical use.

EPIPOLAR GEOMETRY BASED ON THE ORUN AND NATARAJAN SENSOR MODEL

As explained previously, the position and attitude of linear pushbroom sensor changes during the acquisition moment. Orun and Natarajan model the position as second-order polynomials, the yaw variations as second-order polynomials, and the pitch and roll angles as constants in the attitude[2]. Differently from the Gupta and Hartly sensor model, the Orun and Natarajan sensor model necessarily needs ground control points and the ephemeris data for the calculation. Although this sensor model is mathematically complex and computationally expensive, its accuracy is high. Most of commercial software packages are based on this sensor model as well [Orun and Natarajan, 1994].

The epipolar geometry based on the Orun and Natarajan sensor model is represented as a mathematical equation shown below [Kim, 2000].

$$y_r = \frac{k_1 x_l + k_2 y_l + k_3}{(k_4 x_l + k_5 y_l + k_6)\sin Q(x_r) + (k_7 x_l + k_8 y_l + k_9)\cos Q(x_r)}$$

Where, (x_l, y_l) and (x_r, y_r) are the coordinates of each image points, respectively, k_1~k_9 are constants and $Q(x_r)$ is a quadratic polynomial of x_r.

---

[1] In this paper, we will call this model as the Gupta and Hartly sensor model.

[2] In this paper, we will call this model as the Orun and Natarajan sensor model

In this equation, we can find that the epipolar geometry of the Orun and Natarajan sensor model is a hyperbola-like equation. Although this model is considerably accurate, the ephemeris data is necessarily required. Therefore, this epipolarity model cannot be applied to images if the ephemeris data is not provided, for example IKONOS or its accuracy is low. Similar to that of the Gupta and Hartly sensor model, it is approximated as a linear line for the practical use.

3. PROPOSED EPIPOLARITY MODEL

In this section, we propose another epipolarity model based on the Gupta and Hartly sensor model, which does not require the ephemeris data but show high performance (processing time, accuracy, etc.). To derive the epipolarity model between two images (called as left and right image, respectively), let the Gupta and Hartly sensor model for each images be shown as below.

$$\begin{aligned} x_l &= A_0X + A_1Y + A_2Z + A_3 \\ y_l &= \frac{A_4X + A_5Y + A_6Z + A_7}{A_8X + A_9Y + A_{10}Z + A_{11}} \end{aligned} \quad (a)$$

$$\begin{aligned} x_r &= B_0X + B_1Y + B_2Z + B_3 \\ y_r &= \frac{B_4X + B_5Y + B_6Z + B_7}{B_8X + B_9Y + B_{10}Z + B_{11}} \end{aligned} \quad (b)$$

In the left pusbroom model (*a*), equations can be represented about *X* and *Y* as below.

$$\begin{aligned} X &= \frac{(k_1y_l + k_2)Z + (k_3x_ly_l + k_4x_l + k_5y_l + k_6)}{i_1y_l + i_2} \\ Y &= \frac{(j_1y_l + j_2)Z + (j_3x_ly_l + j_4x_l + j_5y_l + j_6)}{i_1y_l + i_2} \end{aligned} \quad (c)$$

Where, $k_1 \sim k_6$, $i_1 \sim i_2$ *and* $j_1 \sim j_6$ are constants.

By applying (*c*) to the right pushbroom model (*b*), we can write the right pushbroom model as below.

$$(m_1y_l + m_2)Z = m_3x_ry_l + m_4x_r + m_5x_ly_l + m_6x_l + m_7y_l + m_8$$

$$y_r = \frac{(n_1y_l + n_2)Z + (n_3x_ly_l + n_4x_l + n_5y_l + n_6)}{(n_7y_l + n_8)Z + (n_9x_ly_l + n_{10}x_l + n_{11}y_l + n_{12})}$$

Where, $m_1 \sim m_8$ and $n_1 \sim n_{12}$ are constants.

Combining two equations, we can derive a new equation as shown below.

$$y_r = \frac{(c_1y_l^2 + c_2y_l + c_3)x_r + (c_4x_l + c_5)y_l^2 + (c_6x_l + c_7)y_l + (c_8x_l + c_9)}{(d_1y_l^2 + d_2y_l + d_3)x_r + (d_4x_l + d_5)y_l^2 + (d_6x_l + d_7)y_l + (d_8x_l + d_9)}$$

Where, (x_l, y_l) and (x_r, y_r) are the coordinates of each image points, respectively and $c_1 \sim c_9$ and $d_1 \sim d_9$ are constants.

This is our proposed epipolarity model and can be written as below.

$$k_1x_r + k_2x_ry_r + k_3y_4 + k_4 = 0$$

Where,

$$\begin{aligned} k_1 &= c_1y_l^2 + c_2y_l + c_3, \\ k_2 &= d_1y_l^2 + d_2y_l + d_3, \\ k_3 &= (d_4x_l + d_5)y_l^2 + (d_6x_l + d_7)y_l + (d_8x_l + d_9), \\ k_4 &= (c_4x_l + c_5)y_l^2 + (c_6x_l + c_7)y_l + (c_8x_l + c_9). \end{aligned}$$

The proposed epipolarity model is a non-linear hyperbola curve and similar to that by Gupta and Hartly in section 2.2. However, differently from that by Gupta and Hartly, the along-track (y_l) is represented as second-order polynomials in the proposed model. If we ignore the second term (y_l^2), we can sure that the proposed model becomes the one by Gupta and Hartly. As against that of Gupta and Hartly, the proposed epipolarity model can be represented as a 4x6 matrix shown below.

$$(x_r, x_ry_r, y_r, 1)\begin{pmatrix} 0 & f_{12} & 0 & f_{14} & f_{15} & f_{16} \\ 0 & f_{22} & 0 & f_{24} & f_{25} & f_{26} \\ f_{31} & f_{32} & f_{33} & f_{34} & f_{35} & f_{36} \\ f_{41} & f_{42} & f_{43} & f_{44} & f_{45} & f_{46} \end{pmatrix}\begin{pmatrix} x_ly_l^2 \\ y_l^2 \\ x_ly_l \\ y_l \\ x_l \\ 1 \end{pmatrix} = 0$$

Where, (x_l, y_l) and (x_r, y_r) are the coordinates of each image points, respectively. The proposed epipolarity model can be calculated using only conjugate pairs without ground control points and the ephemeris data.

4. EXPERIMENTAL RESULTS

The proposed epipolarity model is verified using two types of linear pushbroom imagery; SPOT "Taejon" and "Boryung" panchromatic images and KOMPSAT "Taejon" and "Nonsan" EOC (Electro-optical) images over Korea. The resolution of SPOT is 10 meters and its swath is 60 kilometers. The resolution of KOMPSAT is 6.6 meters and its swath is 17 kilometers. Details of scenes are summarized in table 1.

The performance of the proposed epipolarity model is compared with those of three other models; (1) the epipolarity model of perspective and aerial imagery, (2) the one by Gupta and Hartly and (3) the one based on the Orun and Natarajan sensor model described in section 2.1, 2.2 and 2.3, respectively.

20 conjugate pairs taken from 20 ground control points are used to calculate the epipolarity model of perspective and aerial images and the one by Gupta and Hartly. The proposed

Table 1. The information of SPOT and KOMPSAT stereo image pairs

		SPOT Boryung	SPOT Taejon	KOMPSAT Taejon	KOMPSAT Nonsan
Left Scenes	Acquisition time	Mar. 1 1997	Nov. 15 1997	Mar. 9 2000	May 1 2000
	Viewing angle	-25.8	4.2	26.0	19.456
Right Scenes	Acquisition time	Nov. 15 1998	Oct. 14, 1997	Mar. 1 2000	April 28 2000
	Viewing angle	0.6	25.8	-4.0	-12.305

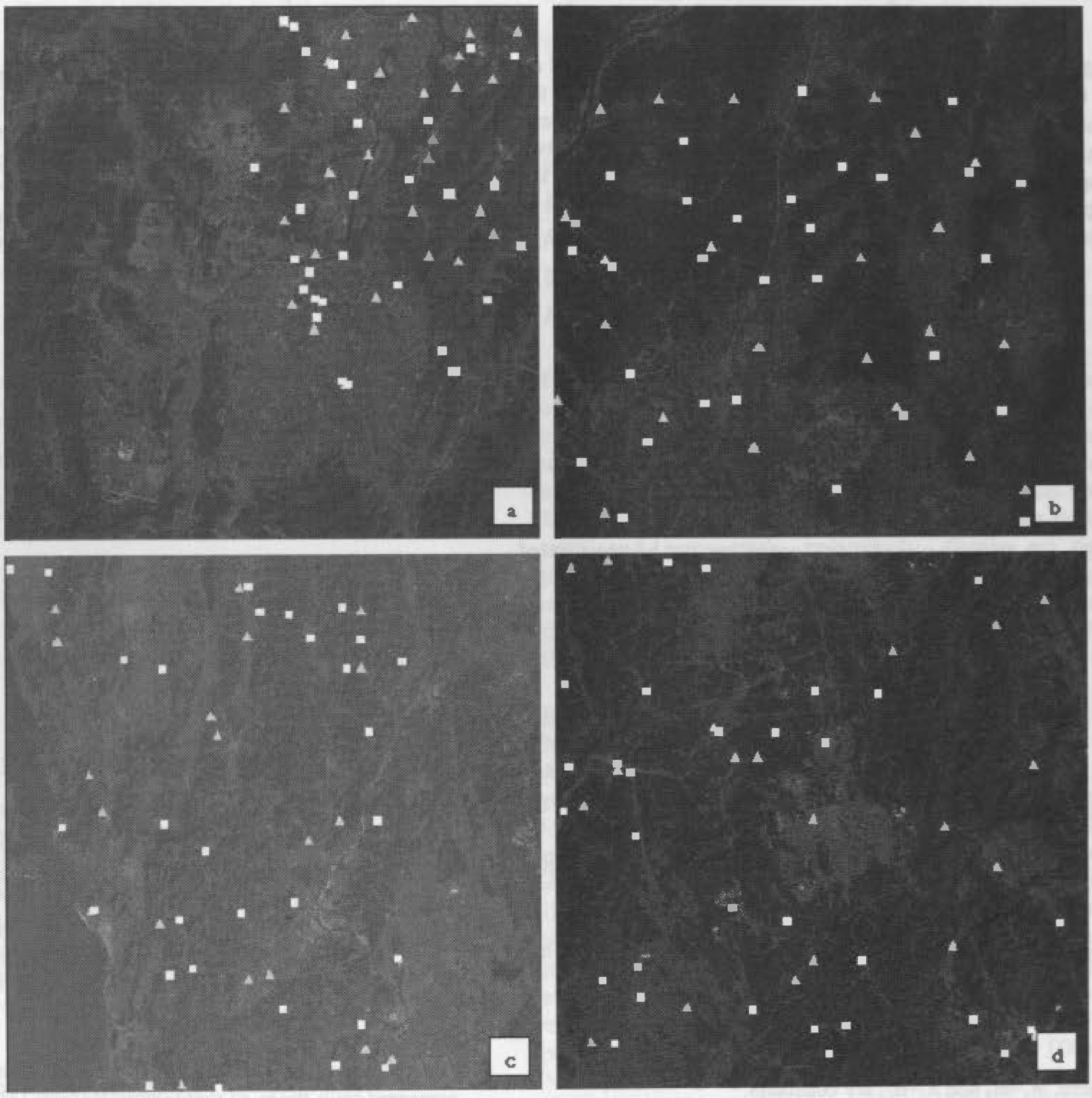


Figure 2. The distribution of modeling and checking points (Modeling : triangle, Checking, rectangular). (1) KOMPSAT Taejon, (2) KOMPSAT Nonsan, (3) SPOT Taejon, (4) SPOT Boryung

epipolarity model is computed using 20 ground control points and the one based on the Orun and Natarajan sensor model is acquired from the ephemeris data of the satellite images and 20 ground control points.

For the quantitative assessment of the accuracy, we devise a robust method which measure the minimum distance between the truth point and the epipolar lines. Except the epipolar geometry of perspective imagery, the epipolar geometry is represented as a non-linear curve in other models. To apply this measure, we estimated a non-linear curve as a linear one because a non-linear curve can be regarded as a linear line in locally [Kim, 2000]. We take 30 corresponding points, accurately extracted by an experienced operator, as independent checking points. The checking and modelling points are distributed to entire images as shown in figure 2.

The results are summarized in table 2. The errors of 30 independent checking points are shown in figure 3, 4, 5 and 6. As shown in the results, the epipolarity model of perspective and aerial imagery and by Gupta and Hartly show the average accuracy below 1 pixel although the errors on several checking points are large. We think that the epipolarity model of perspective imagery can be applicable to linear pushbroom imagery. To clarify the results, more experiments are necessary using various linear pushbroom images. The epipolarity model based on the Orun and Natarajan sensor model show high accuracy than that of perspective images and by Gupta and Hartly. However, this model cannot be computed without the ephemeris data. The accuracy of the proposed epipolarity model is considerably high and similar to that based on the Orun and Natarajan sensor model although the ephemeris data is not applied. The proposed model is also not so computationally expensive. It means that the proposed epipolarity model can be effectively applicable to the imagery which do not provide the ephemeris data such as IKONOS or when the quality of the ephemeris data is low. In this experiments, the proposed epipolarity model is derived using ground control points. However, it can be acquired from only conjugate pairs as described in section 3.

5. CONCLUSIONS

In this paper, we proposed a new epipolarity model which does not require the ephemeris data but show high performance (accuracy, processing time, etc.). We also quantitatively analyzed various epipolarity models to verify the applicability for linear pushbroom imagery.

The analysis of the proposed epipolarity models and others are performed using two types of linear pushbroom imagery; SPOT and KOMPSAT. The results show that the proposed epipolarity model can model the epipolar geometry of linear pushbroom images although the ephemeris data is not used. It means that the proposed epipolarity model is effectively applicable to the linear pushbroom imagery which does not provide the ephemeris data because of various reason for example IKONOS, etc.

In this paper, we verified the proposed epipolarity model on two types of linear pushbroom imagery; SPOT and KOMPSAT. However, we think that the proposed model will work in other types of linear pushbroom images. Our future research is focused on applying the proposed and other models to the reconstruction of the digital elevation model. We will report results of such experiments in the future.

REFERENCES

Gugan, D.J. and Dowman, I.J., 1988, Accuracy and completeness of topographic mapping from SPOT imagery, Photogrammetric Record, 12(72), pp. 787-796.

Gupta, R. and Hartly, R. 1997, Linear Pushbroom Cameras, IEEE Transaction on pattern analysis and machine intelligence, 19(9), pp. 963-975.

Kim, T., 2000, A Study on the epipolarity of linear pushbroom Images, Photogrammetric Engineering and Remote Sensing, 66(8), pp. 961-966.

Orun, A.B. and Natarajan, K., 1994, A Modified Bundle Adjustment Software for SPOT Imagery and Photography: Tradeoff, Photogrammetric Engineering and Remote Sensing, 60(12), pp. 1431-1437.

Zhang, ZY., 1998, Determining the epipolar geometry and its uncertainty: A review, The international journal of computer vision, 27(2), pp. 161-195.

Table 2. Performance analysis for each epipolarity models (in pixels)

		Perspective Epipolarity model	Gupta and Hartly Epipolarity model	Proposed Epipolarity model	Orun and Natarajan Epipolarity model
SPOT Boryung	MEAN	0.309	0.358	0.275	0.240
	STD. D	0.190	0.223	0.178	0.177
	RMS	0.363	0.422	0.327	0.298
SPOT Taejon	MEAN	0.716	0.924	0.299	0.504
	STD. D	0.669	0.701	0.222	0.392
	RMS	0.981	1.160	0.373	0.639
KOMPSAT Nonsan	MEAN	0.509	0.519	0.528	0.505
	STD. D	0.377	0.452	0.340	0.336
	RMS	0.634	0.689	0.629	0.607
KOMPSAT Taejon	MEAN	0.537	0.568	0.521	0.498
	STD. D	0.267	0.365	0.302	0.244
	RMS	0.600	0.676	0.602	0.554

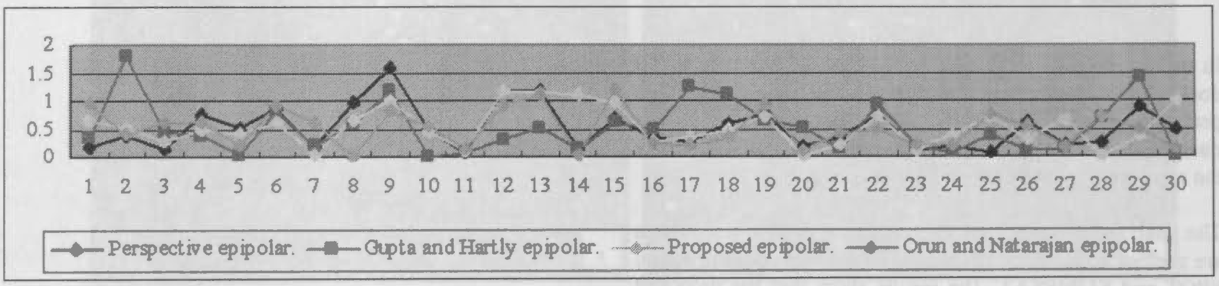


Figure 3. The errors for each independent checking points in SPOT Boryung area (in pixels)

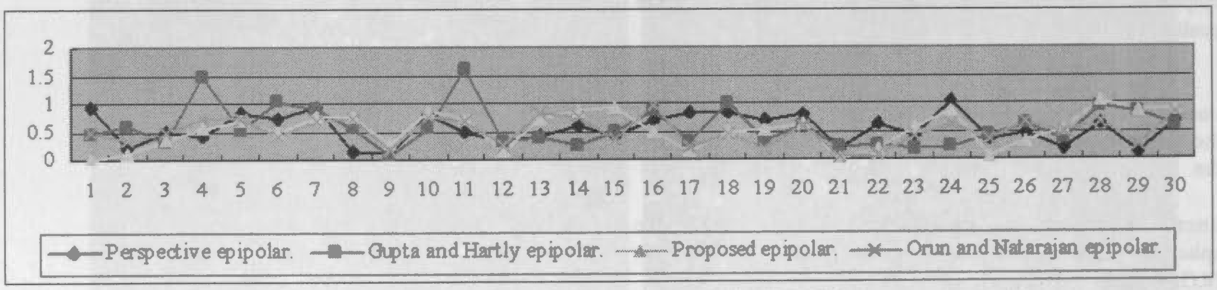


Figure 4. The errors for each independent checking points in SPOT Taejon area (in pixels)

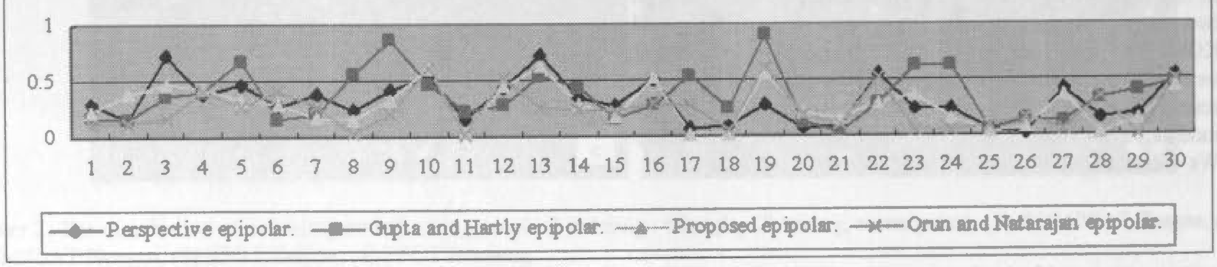


Figure 5. The errors for each independent checking points in KOMPSAT Nonsan area (in pixels)

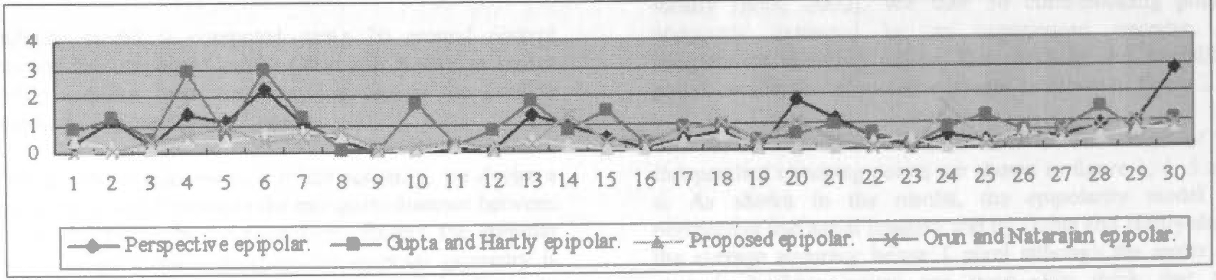


Figure 6. The errors for each independent checking points in KOMPSAT Taejon area (in pixels)

3D RECONSTRUCTION FROM VERY HIGH RESOLUTION SATELLITE STEREO AND ITS APPLICATION TO OBJECT IDENTIFICATION

J. R. Kim , J-.P. Muller

Dept. of Geomatic Engineering, University College London, Gower Street, London, WC1E 6BT UK
jkim@ge.ucl.ac.uk, jpmuller@ge.ucl.ac.uk

Commission IV, WG IV/7

KEY WORDS: High resolution satellite image, DEM, DSM, Lidar, automated Building and tree detection

ABSTRACT:

New high-resolution satellite images such as those from IKONOS provide potentially useful information for the identification of individual surface objects such as trees and buildings compared with previous satellite image sources (e.g. SPOT, Landsat). However, the detailed structure of buildings and trees is still ambiguous for the application of existing machine vision algorithms, mainly due to the resolution limit, currently about 1 metre, the existence of strong shadows due to the large oblique angles and the local time of overpass. Other data sets such as Lidar, aerial photographs, GIS information and pre-existing DEMs, can be used to complement these IKONOS image drawbacks. In particular, owing to the current paucity of IKONOS stereo-pairs, it may be possible to utilise Lidar data as an alternative source of 3D information, where available. This hypothesis has been tested with a 1m pan sharpened multispectral image of East London using a geometrically corrected IKONOS image and coincident 2m Lidar data using a new algorithm which identifies regions of interest (ROIs) for landscape objects (such as trees & buildings). This new algorithm was developed using a Bald Earth Digital Terrain Model (DTM) derived from Lidar and NDVI derived from IKONOS. As a result, excellent building outlines and individual tree extents were detected. However, results shown here applied to an IKONOS stereo-derived DEM indicate that the existing quality of such DEMs is not yet good enough for the detection of small objects.. A new processing system to refine IKONOS DEM quality is currently being developed and higher resolution satellite images (such as Quick-Bird) have recently appeared leading to the expectation that high resolution satellite based GIS construction may be realised in the near future.

1. INTRODUCTION

IKONOS 1m resolution is superior to all previous civilian satellite data. Unfortunately, although 11-bit images are recorded, the image quality and resolution appear to make them unsuitable for the direct application of machine vision algorithms and the large size (11 x 11km) makes the search space too large for object detection.
One alternative approach is to use segmentation of DSMs (Digital Surface Models) in a focusing strategy to reduce the search space into more manageable discrete units. There are several very efficient algorithms available for Lidar and photogrammetrically derived DSMs. We have assessed stereo IKONOS DEM quality from commercially available software for this purpose, such as SocetSet, as well as our own in-house stereo-matching system (Kim & Muller, 1999) but the output DSM quality appears to be insufficient for the automated delineation of individual buildings in urban areas apart from very large structures so that such focusing strategies don't work. Also the positioning error of IKONOS Geo product is too large to register correctly multi-source data-sets. Two possible solutions are here tested. The first is to exploit secondary information such as Lidar and the second is to update the positioning accuracy and DEM quality of IKONOS stereo. Currently very good results have been obtained using the first approach. In the case of IKONOS stereo DEM updating, this work is still in progress and only preliminary results are shown here.

1.1 Data set description

Two data sets have been utilised to date:

1) East London area comprising a 12 km by 12 km IKONOS Precision 1m Pan sharpened image, 2m resolution Lidar which is provided by Infoterra Limited, 4,600 kinematic GPS (7 tracks) (kGPS) measurements and Landline® plus GIS vector data for some part of the area.
2) Barton Bandish test area in Eastern England comprising a pair of IKONOS stereo panchromatic 1m and 4m multispectral images covering an 11 x 11km area, 40,000 kGPS measurements and Landline® plus data. Field survey work was performed in the test areas (Barton Bendish, East London) by collecting k-GPS and static GPS point measurements as well as using a digital camera to obtain "ground truth".

1.2 Previous research work

The first significant research work concerning focusing strategies was published by Baillard and Maitre (1999). They proposed as their focusing strategy segmenting a scene into two different kinds of components: "at ground" and "above-ground" objects. The set of 3-D points is segmented into homogeneous altitude regions by a classical region growing and merging algorithm. In coarse DEM areas, they suggested graph based interpolation and the application of this is used as part of our gridding component development. (see later)

Haala (1997) employed a DSM segmentation technique based on local surface normals. The normal direction from a Lidar DEM was approximated using the derivatives of a local bivariate polynomial fit. The direction of the unit normal vector of a possible roof plane emerging from the ground plan is then

perpendicular to the segment of it. So within a building outline, the correct extent of heights can be determined and segmentation carried out.
Both methods are unsuitable for our use, due to the lack of sufficiently good quality and/or sufficiently high enough resolution DEMs.

External data such as maps and digital map data stored in a GIS, can also be used to define a ROI as in Roux (1997), Brennerand and Haala (1998). However, such additional data is not always available.

One simple method to detect a building or a tree is a morphological opening filter. By applying a "top hat' filter to a DEM, building area height points can be removed. This method is usually applied to low resolution imagery since around very large buildings or clusters of buildings, the filter often fails to derive the correct ground DEM.

We used a modified version of this filter for the initial stage to define the ROI.

According to our end user specifications, bald surface DTMs are also required. There is some research about bald surface DEM construction (often referred to as a Digital Terrain Model or DTM) from a DEM. Usually such methods are based on a hierarchical fitting scheme such as the one described by Kraus & Pfierfer (1998).

Another example is the method described by Tao et al (2001) using Lidar and very high resolution airborne SAR. The authors initially fitted some ground seed points starting at low resolution. "Above ground" height points were detected by their height difference and the SAR amplitude information, and applied to a next level fitting stage with a low weighting value. Lu et al (1998) developed a method, which is very similar to Tao's method but used image classification results instead of SAR amplitude.

1.3 Fundamental Approach

A DSM segmentation based approach cannot produce correct landscape object detection with low density height points and/or poor quality heights, because the gridding error in object boundaries prevents correct building definition.

Therefore, the identification of a correct building and tree extent in our method is dependent on the accuracy to which we can retrieve a Bald Earth DEM. By splitting all height points into "above-ground" and "at ground' components using this information, a new gridding scheme with only "above-ground" height points has been developed to obtain relatively clear building and tree boundaries. Moreover, the "above-ground" points can also be separated by an NDVI mask so that the trees, which are near to a building boundary don't impinge on the building ROI quality. Through a hierarchical scheme, the quality of the ROI can be updated, because the slope is smoother at coarser resolution. A tension factor based gridding method has been employed to take this effect into account.

1.4 Structure of this Paper

In section 2, the overall processing scheme and some outline details of the algorithmic method are described for object detection including bald surface extraction, gridding of lidar points and the extraction of a normalised DEM. Assessments of this scheme were made with the two input data-sets, viz. 1) IKONOS and multispectral pair, 2) Lidar and IKONOS pan sharpened multispectral image combination. In both cases, a positioning accuracy check was performed and an additional processing step to update the IKONOS Geo stereo based product's positioning quality was applied. This step was crucial to update the registration accuracy to compensate IKONOS stereo derived DEM quality with other data sources through, for example, the pan-sharpened multi-spectral images.

The quality of the ROIs for building and tree areas was assessed by comparing them against the Landline® plus digital map vector data.

2. ALGORITHM DESCRIPTION

The overall approach is shown schematically in Figure 1. To save CPU time and construct a more accurate object boundary, the "above-ground" and "ground' points are separately gridded. The most crucial factor in this algorithm is the selection of "ground seed points". If the seed points are correctly selected, the constructed bald surface should be accurate enough to define a sensible ROI.

2.1 Bald Earth seed points selection

The simplest method for this step is to use a local minimum operator (Kilian et al., 1996). However, the local minimum based Bald Earth reconstruction is very sensitive to noise and the reconstructed surface based on these seeds points is usually lower than the actual bald Earth surface. Therefore, we developed a more robust technique based on the local height distribution. At the start, every height point is divided by the local minimum value to remove the effect of any slope bias. Using a moving window, whose size is determined by an initial estimate of maximum building size, the local height distribution was analysed.

If the moving window size is large enough, a height distribution should consist of two Gaussian functions, so that the mean value of a lower distribution can be used for reliable seed points. To fit two normal distributions, we use Kittler and Illingworth (1986)'s criterion.

$$\begin{aligned} J(t) = 1 + 2.0(P_1(t)\log\sigma_1(t) + P_2(t)\log\sigma_2(t)) \\ - 2.0(P_1(t)\log P_1(t) + P_2(t)\log p_2(t)) \end{aligned}$$

(where P(t) is the sum of Probability Density Functions (PDFs) and σ is the standard deviation)

A threshold value for two normal distributions is then minimized using the above function. After finding this threshold value, two mean values can then be calculated.

Now,
If $\mu_2(t) - \mu_1(t) <$ the estimated object height, then $\mu_1(t)$ is the mean value height of a larger object surface or it is assumed to be flat earth so that the window size then needs to be extended and the estimation repeated.
If $\mu_2(t) - \mu_1(t) >$ the estimated object height, then $\mu_1(t)$ is selected as a seed point within the window

Applying this method, suitable seed points can then be selected.

2.2 Gridding of Bald earth

From the initial seed points, an approximate Bald Earth DTM can be constructed by grid-point interpolating the seed points selected in the previous step. However, the interpolation interval needs be too large enough to avoid interpolating noise. Hence, we use Smith & Wessel (1990)'s method to interpolate bald earth seed points. It's one kind of fitting method for the solution of the following function:

$$(1-T)\,\nabla^2 Z - T\nabla Z = 0$$

(where Z height points, T : Tension factor between 0.0 to 1.0)

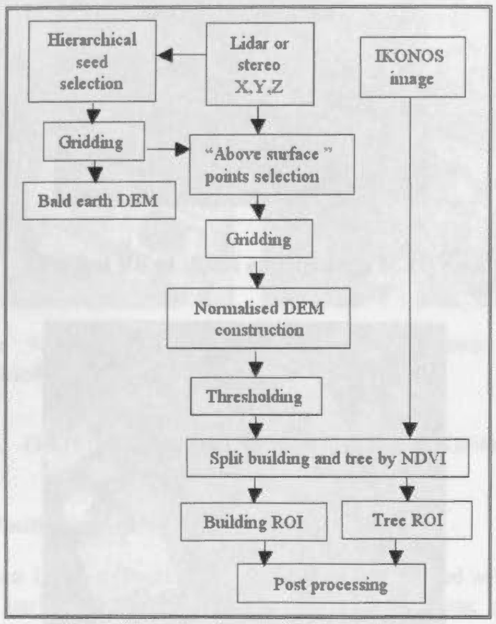

Hierarchical seed selection
Lidar or stereo X,Y,Z
IKONOS image
Gridding
"Above surface" points selection
Bald earth DEM
Gridding
Normalised DEM construction
Thresholding
Split building and tree by NDVI
Building ROI
Tree ROI
Post processing

Figure 1 ROI definition scheme from height and NDVI

Using this method, a higher tension factor will produce a higher curvature surface. To construct a smoother surface at a low hierarchical stage, a low tension factor is required. In the first stage, the lowest tension factor and largest window size are used to save CPU time. From experience, they do not appear to seriously affect the quality of the derived bald earth surface. This implies that the accuracy of the derived Bald Earth is relatively free from user defined parameters in our suggested processing method if and only if, the seed point selection is appropriate and not subject to gross errors.

Then the height difference between the bald surface and Lidar measurements can be derived and the height points which are higher than some critical value are excluded. This critical value is calculated from the assumption that most natural Earth slope values are always lower than 25 degrees. These points, which are derived from this step will consist of the next level Bald Earth DEM seeds points and these are then interpolated using a higher tension factor and finer interpolation interval. The iteration continues until a termination criterion is reached.

2.3 Normalised DEM construction and ROI definition

Comparing the constructed Bald Earth DTM and height values, "above-ground" points are then selected. After "filling" gaps with a reasonable search area, all points are re-gridded using the highest tension factor (≅1.0). This step is useful for preserving edges. Then by overlapping with the IKONOS NDVI, these normalised DEM patches can be divided into tree and building areas using any thresholding algorithm, because the NDVI distribution within "above-ground" areas have clearly two Gaussian distributions. We expect a better result by isolated point removal, such as morphological pruning, because if trees are removed close to a nearby building, the building boundary line will be distorted.

2.4 Correspondence with IKONOS image features

Building roofs in images are sometimes not very well matched with the corresponding DEM patches in the case of very tall buildings because of local height relief (Figure 2).

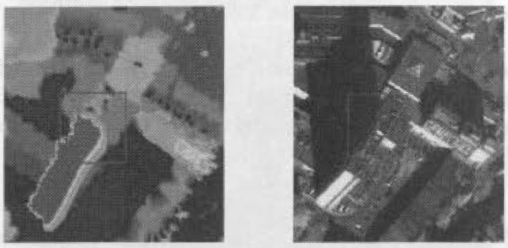


Figure 2. Planimetric shift due to local height relief

If an ortho-image is created using a Bald Earth DTM, this shift can be found without detailed camera information (Baltsavias et al., 2001) using the relationship

$$\Delta X_i = -\Delta Z \sin(a)/\tan(e),\ \Delta Y_i = -\Delta Z \cos(a)/\tan(e)$$

(where ΔZ : normalised height points-Bald earth height
a : sensor azimuth
e : sensor elevation)

Now, by merging a building area's normalised heights and roof features, the correct building structure can then be reconstructed.

3. OBJECT DETECTION IN IKONOS STREO PAIRS

3.1 Positional and radiometric properties' check

Initially, the positioning accuracy of IKONOS was assessed by using GCPs from GPS measurements and Landline® plus data. In this case about 10 metres of planimetric shift was observed (Figure 3).

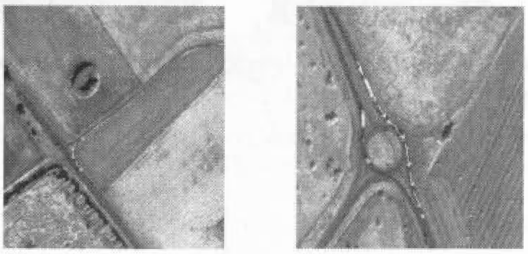


Figure 3 Error of ortho-image positioning error by IKONOS RPC (red: original, white: corrected)

Table 1 shows the positional error in the ortho-image measured using 24 static GPS measurements before and after updating the RPCs.

	X direction (metre)		Y direction (metre)		X,Y RMS (metre)	
	A*	B*	A	B	A	B
Average	9.733	0.002	-3.366	0.002	10.749	0.002
Std dev	2.449	3.352	3.264	3.791	2.614	5.061

Table 1 24 GPS points' positional error in the IKONOS ortho image which is produced by the initial (A) and updated (B) RPC parameters

These positional errors will be in WGS84 ellipsoid co-ordinates and are a serious problem for multi-data registration, which is one solution for object detection.

Another problem of IKONOS images are there radiometric quality (Figure 4). Edges are not clearly detected in IKONOS images. That's one reason why the correct ROI definition using 3D information is necessary.

3.2 Positional accuracy updating for DEM construction

Using a simple RPC analysis, we updated positional accuracy to within 3-4 m (Figure 5), so that a better registration between data sets is possible. However, not from Table 1 that although all the bias is removed, there is still some 3m of random error which we believe to be due to higher order RPC parameters not currently taken into account

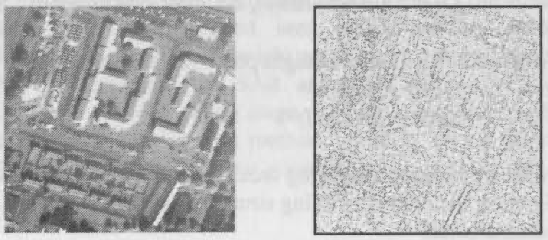


Figure 4 Edge detection in IKONOS image

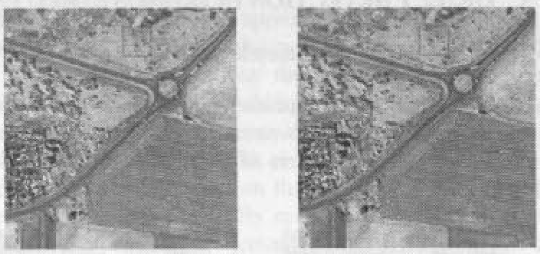


Figure 5 Positioning error removing (left : before updating, right : after updating)

The DEM (Figure 6) and ortho-images (Figure 7) were constructed with this updated RPC and our in-house P-Gotcha matcher (Kim & Muller, 1999). The 3 metres of vertical error which was previously reported by (Muller et al., 2001) is removed, so that a more accurate Bald Earth DTM construction is now possible.

3.3 ROI definition result

We have tested our ROI method using the previously described method. Because the target area is very flat, the region growing based method, and the multi-channel classification (4 pan-sharpened multispectral + DSM) was applied to the flat area (Muller et al., 2001) with the height distribution based method. The Bald Earth construction result (Figure 8) appears reasonable but the ROI boundary is not correctly defined, because of the delineation problem with the building boundaries. The solution of this problem might be an improved image matcher, which can treat features and can address problems caused by high order RPC errors.

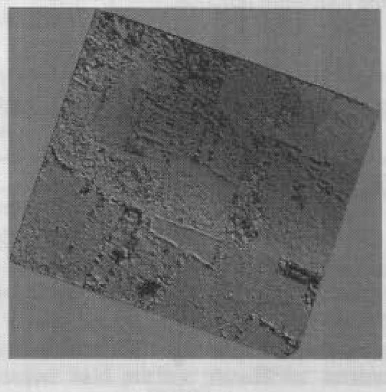


Figure 6 DEM construction result in BB test area

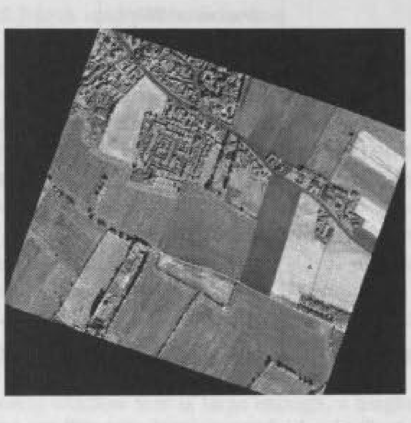


Figure 7 1 metre resolution IKONOS pan-sharpened multispectral ortho-image

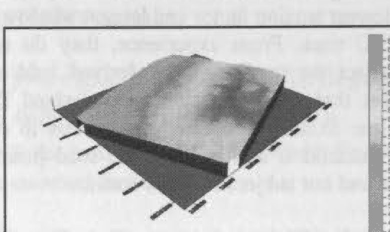


Figure 8 Bald Earth construction result compared to the OS® PANORAMA® DTM

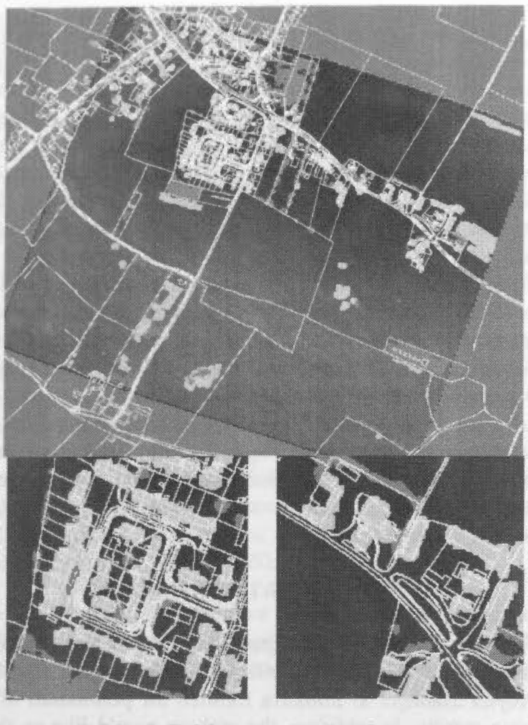


Figure 9. ROI definition and visual assessment with Landline® plus

4. OBJECT DETECTION IN LIDAR & IKONOS PRECISION PRODUCT

4.1 Positioning

The East London (Precision) image was orthorectified with a 2-metre horizontal and 3-metre vertical accuracy using ground control points. In this case, no RPC exists. However, the horizontal and vertical accuracy using k-GPS points was checked (Figure 10). As observed, the horizontal accuracy is within 2 metres as per the Space Imaging specification. Also the edge detection quality was assessed compared to Landline® plus.

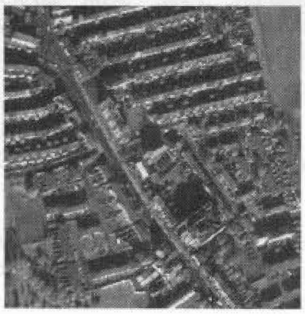


Figure 10. Example of Positioning and radiometric accuracy of IKONOS Precision by k-GPS tracks

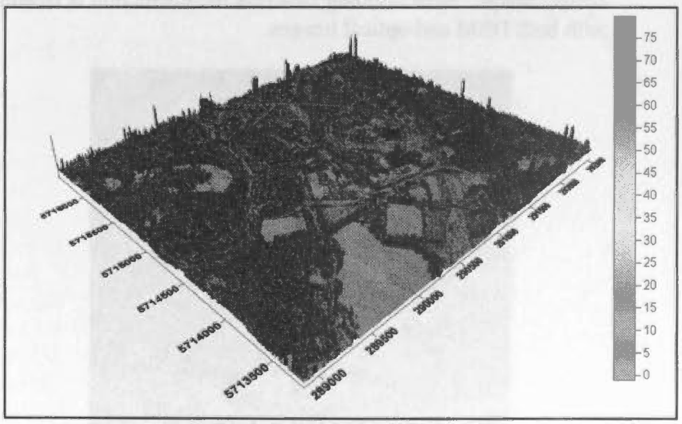
4.2 Bald Earth construction

The maximum estimated building size and target object area height were set as 100 metres and 4 metres respectively. So the moving window size and threshold value are set following these values. Figure 11 (a)-(d) shows the initial input data and results. This figure shows that the Bald Earth construction appears to be more accurate so that in the normalised DEM, the detailed roof structures over the buildings are now well observed.

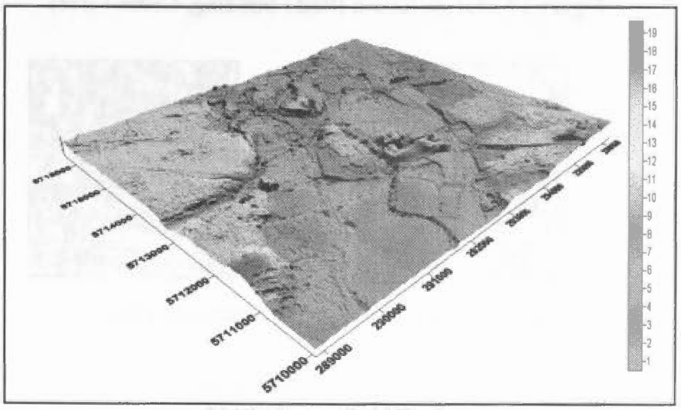
4.3 ROI definition by thresholding and NDVI splitting

The tree and building ROI segmentation is produced using optimal thresholding and shows good agreements with the IKONOS image (Figure 12). If the normalised DEM is accurate enough, in other words, global slope effects are completely removed, a higher threshold value should give a clearer building boundary. We used a 4 metres threshold value. Because after segmenting trees around buildings, the boundary is distorted, these ROIs must be rectified by a morphological filter.

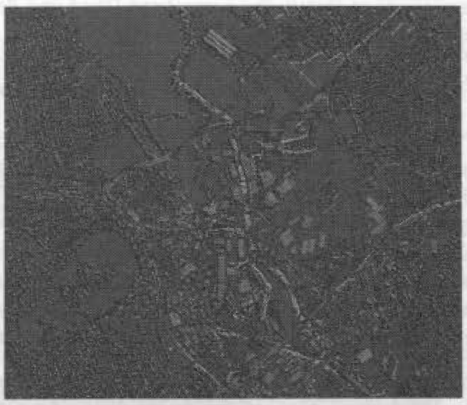
(a) IKONOS Precision image in test area (R:IR,G:Green,,B:Red)



(b) Original Lidar DEM



(c) Final stage Bald Earth



(d) Re-constructed normalised DEM

Figure 11 Input Lidar data, final Bald Earth DTM and normalised DEM

4.4 ROI assessment

Landline® plus data was used to do a visual assessment of the building ROI. Even for very small buildings, it appears to detect buildings very well (Figure 13). The outlines of buildings can now be directly input to building reconstruction

4.5 Connection to the IKONOS image

The average normalised heights of each building shape can be extracted and applied to find out the correct roof location in the optical image. Now building structure reconstruction is feasible with both DEM and optical images.

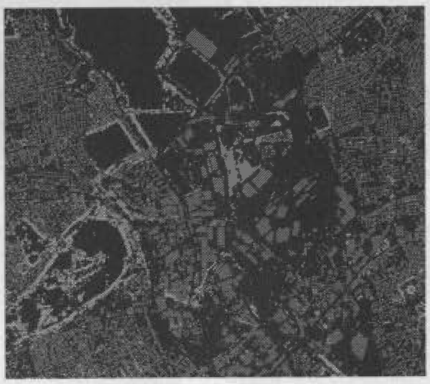


Figure 12 ROI definition (Red : building, Green : tree)

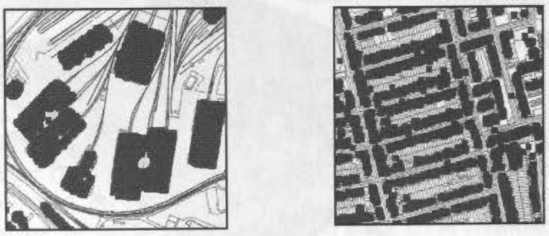


Figure 13 Assessment of ROI by Landline® plus

5. CONCLUSIONS

Because of the 1m resolution limit, the existence of strong shadows, the radiometric characteristics and the large search space, object detection in IKONOS image is very difficult to achieve directly. We have developed a new technique to define the building and tree extents using 3D height points. It is robust and fast, if the DEM quality is good enough. Currently, the application of this method to IKONOS stereo DEMs is not completely successful, even if it is done using improved positioning information and a state-of-the-art image matcher(P-Gotcha), which was developed for geometrically distorted images. However, we showed that with 2 metre resolution Lidar data and IKONOS Precision images excellent results can be obtained mainly because of the good delineations of building outlines. Currently we are developing a new stereo DEM construction process, which uses an improved parallelised P-Gotcha (PVM Gotcha) algorithm, an unidirectional positioning error removal method and other matching data sources (colour, features). It is expected that using this scheme it will be feasible to extract GIS data from high resolution satellite stereo data including new high resolution satellite data such as the Quick-Bird satellite.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge the financial assistance of the BNSC-LINK RISKMAP project especially the Project Manager at Infoterra Limited for permission to publish these results. In addition, the authors would like to thank the NASA Earth Science Enterprise project (under contract No. NAS13-98048) for supply of the stereo IKONOS imagery for Barton Bendish and Infoterra Limited for supply of the 2m lidar data over East London. Includes material which is © Space Imaging L.P and is Crown Copyright (Landline® digital vector map data and PANORAMA® DTM).

REFERENCES

Baillard, C., Maite, H., 1999. 3D reconstruction of urban scenes from aerial stereo imagery, A focusing strategy, *Computer Vision and Image understanding*, 76(3), pp.244-258.

Baltsavias, E., Pateraki, M. & Zhang, L., 2001. Radiometric and geometric evaluation of Ikonos Geo images and their use for 3D building modelling. *Proc. Joint ISPRS Workshop High Resolution Mapping from Space 2001*, Hannover, 19-21, CD ROM)

Brenner, C., Haala, N., 1998, Fast production of virtual reality city models, in ISPRS Comm.IV Symposium, *"GIS between visions and applications,"* Stuttgart, ISPRS, 32, part 4, pp.77–84.

Halla, N., Brenner, C., 1999. Virtual city model from laser altimeter and 2d map data, *Photogrammtric Engineering & Remote Sensing*, 65(7) , pp.787-795.

Kilian, J., Halla, N., Englich, M., 1996, Capture and evalaution of airbone laser scanner data, *Int. Archive of Photogrammetry & Remote Sensing*, Vienna, V.XXXI, pt.B3., pp.383-388.

Kim, T., Muller J.P., 1999, Development of a graph-based approach for building detection, *Image and Vision Computing 17*, pp. 3-14.

Kittler, J., Illingworth, J., 1986, Minimum error thresholding *Pattern Recognition*, 19, pp. 41-47.

Kraus, K., Pfeifer N., 1998, Determination of terrain model in wooded areas with airborne laser scanner data, *ISPRS Journal of Photogrammetry & Remote Sensing*, 53, pp.193-203.

Muller J.P., Kim, J., Tong L., 2001, Automated mapping of surface roughness and landuse from simulated and spaceborne 1m data, *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)*, E. Baltsavias, A. Gruen, L. Van Gool (Eds.), Balkema Publishers, Lisse, The Netherlands, pp. 369-380

Lu, Y., Kubik, K., Bennamoun, M., 1998, An accurate approach to localize house area for 3D terrain reconstruction, *Proc. of the 9th Australasian Remote sensing & Photogrammetry Conference, Sydney, Australia,*

Paskoand, M., Gruber, M., 1996, Fusion of 2D GIS data and aerial images for 3d building reconstruction, in XIX congress of ISPRS, Comm.III, *Int. Archives of Photogrammetry & Remote Sensing* ,Vienna, 31, pp.257–260.

Roux, M. , 1997, Cooperative analysis of maps and aerial images for urban scene description ,in SPIE Europto : *Image and Signal Processing for Remote Sensing* III, London, 32 (17), pp.254–267.

Smith, W., Wessel, P., 1990, Gridding with continuous curvature splines in tension, *Geophysics*, 55(3), pp.293-305.

Tao, V., Wang Y., J. Mercer J., 2001, Automatic reconstruction of bald digital terrain models from digital surface models generated from airborne SAR system, The 3rd International Symposium on Mobile Mapping Technology. Cairo, Egypt, *CD-ROM*

INTEGRATION AND FILTERING OF 3D SPATIAL DATA USING A SURFACE COMPARISON APPROACH

H.L. Mitchell*, J.G. Fryer, R. Pâquet

School of Engineering, University of Newcastle, Newcastle, NSW, 2308, Australia
harvey.mitchell@newcastle.edu.au, cejgf@cc.newcastle.edu.au, rpaquet@mail.newcastle.edu.au

Commission IV

KEY WORDS: Database, surfaces, matching, proposal, spatial, DEM, DTM, algorithms

ABSTRACT:

Combining a number of three-dimensional spatial data sets, which have been derived from different sources such as photogrammetry and more contemporary aerial and space-based remote sensing methods, demands that the integrity of the revised database be maintained. It is necessary to maintain or improve the combined data set's "size-to-information ratio". This problem is being tackled by treating both data sets as surface representations, and using a variation of least squares surface matching to permit the deviations of any point from the surface created by the other data set to be examined. The contributing data sets can suffer from systematic errors, which may be interpreted as anomalous regional position errors or height errors in a combined data-base, but surface comparison allows systematic error to be modelled within the matching process. A surface comparison approach also permits the search for objects among regionalised differences. Identifying objects within groups of differences indicates that the differences are real or erroneous, in which case the object can be retained in the combined data set. However, the first task is to ensure the perfect operation of the surface matching process which is subject to weak solutions caused by inhomogeneous point density and inappropriate surface textures. The work is being driven by a real application which arises in a project to form an accurate DTM for flood prediction purposes and subsequent disaster plan management studies. The task is to integrate a massive amount of height data from aerial photogrammetry with all existing ground surveyed points.

1. BACKGROUND: SPATIAL DATA AGGREGATION

The three dimensional spatial data required by interested and relevant agencies, such as the government and semi-government authorities involved in urban and non-urban planning and administration, can be gathered for them from a number of very diverse sources. These include, at the present time (being realistic) GPS ground surveys, aerial photogrammetric surveys (obtained both manually and automatically), aerial laser scanning (Lidar) coverage and traditional ground surveys methods. It is widely recognised that these multiple sources of 3D surface data differ in terms of their coverage, spatial density, accuracy and precision and the nature of the artefacts, objects and the features which they collect. The aerial photogrammetric surveys may be in the form of DEMs derived from automatic correlation in rural areas or as spot heights and contours derived from manual operations, especially in urban areas. Current or historical ground surveys can offer very different classes of output. Lidar and photogrammetry can have different sensitivities to vegetation and motor vehicles, and even buildings if that is in accordance with a photogrammetric operator's instructions. The contrast between photogrammetric collection of data on a regular grid, Lidar TINs, ground-based GPS, and traditional surveys (in which human decision makers are more likely to perceive and locate changes of grade and break-lines and linear features such as road centre-lines) can be vast. The differences in accuracy between ground-based surveys and aerial/space surveys may be in a ratio of 50:1.

When broad area coverage is needed for certain purposes - such as flood control, which is the problem of relevance in this study - agencies can make maximum use of data sources by combining data sets derived from the different sources, into a single entity. The benefits of "multi-sensor integration" for "the fusion of DSMs of equal or different sources with equal or different spatial and temporal properties" (Schiewe, 2000) are well known, as is the fact that the combination of such data to provide an optimum set of ground point information is not as straight-forward as it may at first seem. This is particularly true when the data sources cover large extents but not necessarily identical areas, as they typically do with Lidar and photogrammetry. In the absence of other advice, any authority which disseminates or uses spatial information would normally use data sets from various sources by simply combining the (X,Y,Z) data together. However, such an agglomeration demands that the integrity of the revised database be maintained. The combined data set must be seamless and homogenous. That is, the date must be not have error sources which could cause the combined data to show either boundary effects nor display other differences due to the different errors. Moreover, it is not efficient to keep all points simply because they exist. It is necessary to check whether the new entity suffers from anomalies caused by sources of differences in the data sets. Furthermore, the combination data set's "size-to-information ratio" must be maintained or improved, and points cannot be retained simply because they occurred in one of the contributing data sets. As Ruiz (2000) points out with respect to the creation of a TIN model for Catalonia, "The model must be dynamic because it has to support insertion and deletion, the surface model should be refinable, algorithms have to be robust and data integrity must be preserved."

---

* orresponding author

The problem which we wish to address in this paper therefore is the means of creating a homogenous but filtered data set. That is, firstly sources of differences between the source files need to be examined, and secondly the combined file needs to be culled.

2. SHAPE MATCHING TO COMBINE AND FILTER DATA SETS - A PROPOSAL

The fundamental hypothesis behind the proposal which is being developed in this study, is that, by treating both data sets as surface representations, and using a variation of least squares surface shape matching to fit them together, various objectives of the spatial data merging problem can best be achieved. The aims of the process are:-

1. to detect any broad deviations between the surfaces, because they may indicate systematic error;
2. to detect any localised deviations between the surfaces, because they may indicate objects which have been detected by one system and not the other;
3. to detect any point deviations between the surfaces, because they may indicate individual data points whose retention is unnecessary.

Surface shape matching can be carried out by a number of techniques. However, with 2.5D surfaces as encountered in urban spatial information systems, least squares matching which minimises the surface separations is seen to be quite adequate and other approaches are unnecessarily complex. Mitchell and Chadwick (1999) have previously argued that the method is suited to matching of 2.5D surface comparison applications - when compared in particular with the Iterative Closest Point algorithm first proposed by Besl and McKay, 1992, and developed by various researchers since then. The least squares shape matching method has already seen use in spatial information studies by Jokinen and Haggrén, 1998, Postolov et al., 1999, McIntosh et al., 2000, Schenk et al., 2000, and Habib et al., 2001.

The concept of least squares shape matching for 2.5D surfaces is not complex, being in principle little different from the area-based image matching used to find the correspondence between image patches in digital photogrammetry. The method seeks to minimise the sum of the squares of the surface differences (rather than the noise in the individual observations, although this is of course a related measure). The required parameters are the parameters of the transformation which moves one surface coordinate system into coincidence with the other. While the coordinate systems are theoretically the same, it is the acceptance of the possibility of a small difference that permits interesting surface variations to be sought.

Like many least squares problems, the relevant equations need to be linearised and initial parameter values are generally needed (but see Habib et al., 2001). Fortunately, in the case of 2.5D spatial data comparison, the initial parameters are easy to estimate, being typically zero.

Details of algorithms and numerical implementation are given in a number of papers, which may be found via the references given in this paper. But despite the theoretical feasibility of least squares surface matching, in practice, numerical difficulties and precision problems can be encountered and these are discussed below following explanations of its usage in terms of the aggregation of spatial data.

3. THE ADVANTAGES OF SURFACE SHAPE MATCHING IN SPATIAL DATA MERGING

3.1 Differences across the entire surface: Systematic errors

The first advantage of using surface shape matching is to detect systematic errors. While none of the sources of data, whether from GPS, photogrammetry or Lidar is expected to suffer from any regular systematic error, in reality various errors can infiltrate the output: ground-based GPS surveys can incur a bias from satellite constellations at the time of survey, photogrammetry can suffer from errors in camera parameters including focal length error, Lidar can suffer from errors in the GPS control, calibration errors and problems with on-board sensor errors and gyro-control. Surface comparison by shape matching of can allow for systematic differences.

Systematic error can be detected by seeking patterns within the residuals. However, in a more sophisticated approach, it is also possible that a model of the patterns of systematic error can be integrated within the coordinate transformation which is chosen as the fundamental mathematical model of the least squares problem. Alternatively, mathematical shapes defining twists and other warping distortions can be used.

It is also noted that as an extension of this concept, (Buckley et al., 2002) argue that it is not necessary to control both the photogrammetry and Lidar if they can be fitted to each other using surface shape matching. However, our discussion here accepts, at least initially, that at the current time all commercial photogrammetry and Lidar data are controlled via GPS and/or ground control.

3.2 Differences across regions: Object identification

A second advantage of using surface shape matching across the surface defined by an entire spatial data set is that it permits outliers to be identified. Moreover, we can seek, by "blob" detection for example, from among the outliers, distinct parcels or regions of differences, which may in turn be identified as objects or features or attributes which have appeared in one type of data collection but not in another. These may be certain types of vegetation, moving objects such as motor vehicles, and new objects including buildings. More importantly, however, break-lines can be identifiable as they represent an attribute which is most importantly in the category of a attribute which is typically identified in some data sets (manual photogrammetry, and ground-based GPS survey or angle-and-distance surveys) and not in another. Items identified in this way may then need to be retained in the combined data set.

It is possible of course that the objects can be identified within the isolated "outlier" groups by appropriate analysis of the residuals. Such a capability is not part of this proposal, and is a separate issue, but it is important because an object which exists but cannot be identified may be incorrectly seen as due to errors and the points deleted. The crucial matter now however is that it is possible to group the surface differences, and that they are not always treated individually. The analysis therefore requires that any surface difference be sought as "regions" of differences, which could indicate new objects. Theses outliers can then be analysed for the existence of objects.

3.3 Point Differences: Filtering and blunder elimination

The third advantage of using surface shape matching is to examine the compliance of individual points with the overall surface model. Points which are outliers deviate not only from one surface but from the combination, as surface-based comparison provides information on how well any point fits the entire combined data set. The alternative is to compare nearby points to each other, and not to see the overall picture. If a point deviates significantly - i.e. compared with the typical standard deviation - it may not be worth keeping, even it is not a blunder, as it does not provide useful information. Thus the elimination of single outliers performs a filtering process.

When surface shape matching is used to correlate the surfaces, neither data set needs to be regarded as the correct or the control surface. It is not effective to accept that one point (probably from the "old" data) is correct, and that the new data point is being assessed. This may be true where one set of data is clearly superior, but when combining two sets of data from different remotely sensed sources, for example, one set must not necessarily be regarded as indispensable, superior or fixed. Treating the data sets as surfaces allows each surface to be given equal weight and influence, so both surfaces can be analysed equally and simultaneously.

The filtering of the data set clearly requires that points in either data set which are judged as erroneous must be rejected, and that useful points must not be rejected. More importantly however, some points in either data set which are not judged as erroneous but which add no new information, must nevertheless be rejected. In this process, therefore, it is important that all points be judged as to their contribution to the combined data set.

4. PROPOSAL SUMMARY

If the least squares problem is represented by the need to minimise the sum of the squares of the residuals V in the linearised matrix equation:

$$V = A\,\Delta + L \qquad (1)$$

where Δ = vector of corrections to the required coordinate transformation parameters
A = matrix of coefficients of the corrections to the required parameters
L = vector of constant terms,

then we notice that our three aims are achieved by utilising the following characteristics of least squares estimation solutions:

1. systematic errors can be sought by introducing parameters describing the anticipated error patterns and are therefore sought via development of matrix A;
2. object searching is carried out by analysing the outliers in the vector V;
3. point filtering is carried out by analysing the elements of the vector V.

5. THE MATCHING ALGORITHM

The project development recognises that the first task is to ensure the validity and reliability of basic matching algorithms. The writers' experience with the matching concept means that they are aware of its weaknesses and difficulties; see Mitchell and Chadwick (1999). Shape matching is not robust, being ill-posed - being arguably an offshoot of image matching which is known to be ill-posed, (e.g. Heipke, 1996). Its strength in the determination of the crucial shift parameters (which are most important in relating two spatial data sets positionally) will improve primarily with the existence of texture within the surfaces. Consequently, with the variety of surface textures which are encountered in real surfaces, the match does not always succeed, for the following considerations.

Despite the need for texture, algorithms based on minimising the vertical separation of 2.5D surfaces (or perhaps based on the normal distance) need to recognise that some surfaces have such features such as vertical surfaces (in the case of buildings and so on) and discontinuities (again typically associated with buildings) and the algorithm processing needs to be prepared for these eventualities.

Other considerations arise primarily from numerical issues, such as the need to interpolate corresponding points on one of the surfaces and from the need to estimate gradients. The interpolation error can be quite large because some interpolation distances in sparse TINs can be such that interpolation is simply unreasonable. The crucial point however is that the interpolation error - or the data point sparseness - needs to be estimated within the processing software in order that the dangers be made apparent. If the interpolation error can be estimated, it can at least be used in a weighting of the equations. Interpolation, and estimating the errors associated with it, is seen as particularly important when data is provided at varying spacings, when some regions of data are considerably sparser than others, and when point rejection is being undertaken.

It is also crucial to look at interpolation errors in order to get the difference values correct, and treating the data as surfaces is a means of interpolating among three or even more points. Studying interpolation and its errors can permit weights in the comparison process (especially to lower the weight on sparse areas) in surface matching.

The analysis depends on the classification of the outliers, a question which has been studies in surface matching by Pilgrim (1996) while it also must be recognised that finding outliers may be better pursued using least median of squares, (e.g. Xu and Li , 2000).

A matching program has been tested with both synthetic and real surface data sets, and the limits to the algorithm's robustness have been partly investigated. The program is using

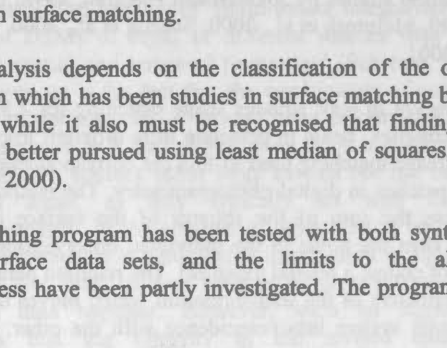


Figure 1. Surface plot of the mathematical function.

a 7-parameter conformal transformation to register the surfaces in the same coordinate system, i.e. match the surfaces. The parameters of the transformation are estimated in a least squares solution, that minimises the normal distances of the points of one surface to the triangulated facets of the other surface. Another program that minimises the remaining differences along the Z axis has been previously created and used by Mitchell & Chadwick (1999). The authors intend to further studies on the comparison of the two algorithms (Schenk, 2000).

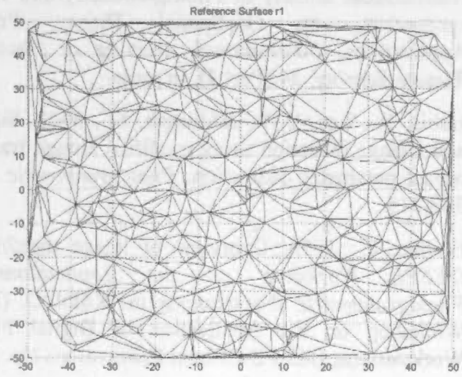
Synthetic data has been generated to test the program. The data consists of a patch of terrain 100m by 100m. The surface is generated by a function with the domain in X and Y from –50 to +50 and heights ranging approximately from –11.5 to +14. A plot of the function is shown in Figure 1.

The X and Y coordinates of the points are generated randomly, with the Z calculated from the X and Y coordinates. Two data sets representing the same surface were created in this manner. The first set was triangulated and used as a reference data set. The other set was transformed, that is the coordinates were shifted, rotated and scaled using a 7-parameter transformation.

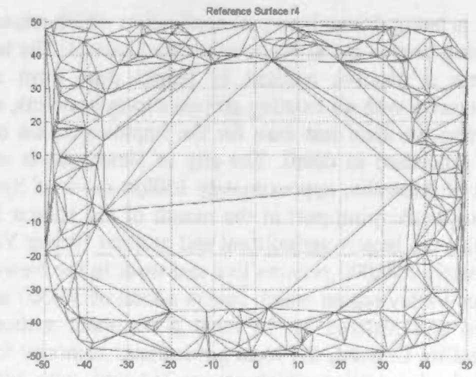
The reference set was altered by eliminating the data in some patches of the set. A sample of the reference data used is shown in Figure 2. The aim was to emulate real situation matching where the reference data might be of high accuracy but patchy as if obtained by surveyors on foot, and to observe and analyse results obtained with such sets. The fact that the triangulation occurred over the bare patches highlighted the necessity to weight the observation to minimise the interpolation errors. At equal density (r4, r5 and r6 contains 50 % of the points contained in r1, and cover 50 % of its area), reference surfaces with spread configuration such as set r4 were found to be the configuration for which the parameters used in the initial transformation were recovered the most closely. A configuration like the surface r5 generated the worst parameter recovery of all, for the reason that the configuration is unstable with relation to rotation with respect of the X and Y axes.

An important part of the research is to predict accuracy of matching process. The least square solution provides the user with statistical information on the matching of a cloud of points to a triangulated surface. It does not provide the accuracy of the matching of the points to the real surface.

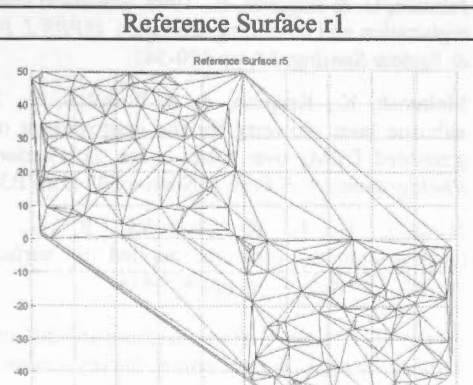
Again an inspection of the results obtained by matching the data cloud to the reference surfaces shown revealed that reference surfaces configured as in r4 had the largest mean normal distance of the three configurations as well as the smallest standard deviation. The distances along the Z axis between the points of the transformed surface and the real surface were calculated in turn to show that the closest fit to the real surface (as opposed to the interpolated surface) when using a patchy reference surface was obtained when using reference data configured as in r4.



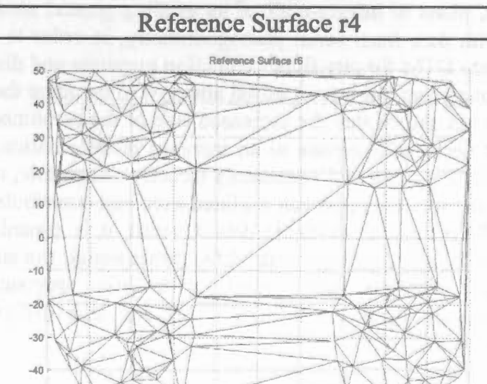
Reference Surface r1



Reference Surface r4



Reference Surface r5



Reference Surface r6

Figure 2. Test surfaces with sparse data.

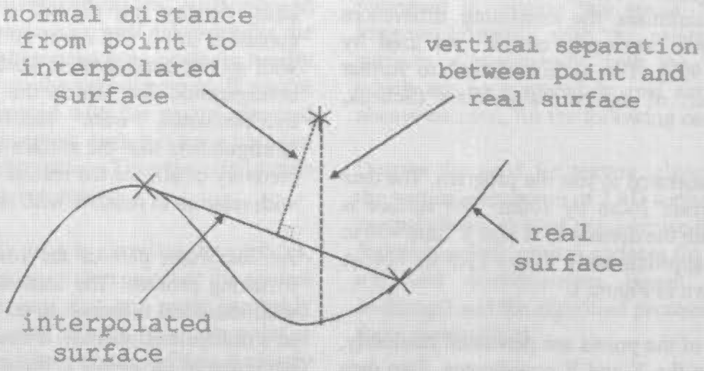


Figure 3. Normal distance to interpolated surface and vertical surface separation.

Prediction to optimise the fit could not in this case be based on the mean of the normal distances. Further research needs to be undertaken with different data distribution, different terrain characteristics (slope and wavelength), different data density and added noise.

6. APPLICATION

The work is being driven by a real application which arises in a project being carried out at Newcastle City Council. The task is to integrate a massive amount of height data from aerial photogrammetry with all existing ground surveyed points, and it thus provides an ideal test case for the implementation of the proposed procedure in detail. The city of Newcastle is on the east coast of Australia, approximately 150km north of Sydney. It has a major shipping port at the mouth of the Hunter River which drains the largely agricultural and pastoral Hunter Valley. Approximately 550,000 persons live and work in the Newcastle and Hunter Valley region which covers a total of 30,000 square km and is administered by 13 local government authorities. Newcastle City Council, the local government authority for the city and suburban region which covers 214 km^2, with 140,000 residents, plans to integrate all of its existing ground surveyed points with data from aerial photogrammetry, in order to form an accurate DTM for city flood prediction purposes and disaster plan management. However, initial attempts to combine the two data sets has shown that the increased size of the combined file does not necessarily equate to an increase in information, and that the data sets are not necessarily directly compatible, and it is proposed that the approach outlined here can contribute to a more efficient use of available data. Overall, it is argued that, when the two data sets are truly three-dimensional, the surface matching technique will permit a holistic approach to simultaneously determining both errors and true differences between the data sets.

7. CONCLUSIONS

It is crucial that the sources of weakness in surface shape matching be eliminated or overcome. However, it is argued that if this can be achieved then the residuals and functional models within the least squares procedure can be used to highlight matters of interest, in terms of single points, groups of points representing objects or systematic error patterns.

REFERENCES

Besl, P. J. &McKay, N. D., 1992. A method for registration of 3-d shapes. *IEEE Transactions, Pattern Analysis & Machine Intelligence,* PAMI-14, pp 239-256.

Buckley, S. J., Mills, J. P., Clarke, P. J., Edwards, S. J., Pethick, J. S. & Mitchell, H. L., 2002. Synergy of GPS, digital photogrammetry and InSAR in coastal environments. International Society of Photogrammetry and Remote Sensing symposium, on GeoSpatial Theory, Processing and Applications, *International Archives of Photogrammetry & Remote Sensing,* 34 (Part 4), in press.

Habib, A., Lee, Y.-R. &Morgan, M., 2001. Surface matching and change detection using modified Hough transformation for robust parameter estimation. *Photogrammetric Record,* 17, pp 303-315.

Heipke, C., 1996. Overview of image matching techniques. *OEEPE Workshop on the Applications of Digital Photogrammetric Workstations* in: Kölbl O. (Ed.), OEEPE - Workshop on the Application of Digital Photogrammetric Workstations, OEEPE Official Publications No. 33, 173-189.

Jokinen, O. & Haggrén, H., 1998. Statistical analysis of two 3D registration and modelling strategies. *ISPRS J. Photogrammetry & Remote Sensing,* 53, pp 320-341.

McIntosh, K., Krupnik, A. & Postolov, Y., 2000. Utilizing airborne laser altimetry for the improvement of automatically generated DEMs over urban areas. *International Archives of Photogrammetry & Remote Sensing,* 32 (Part B3), pp 563-569.

Mitchell, H. L. & Chadwick, R. G., 1999. Digital photogrammetric concepts applied to surface deformation studies. *Geomatica,* 53, pp. 405-414.

Pilgrim, L.J., 1996. Surface matching and difference detection without the aid of control points, Survey Review, 33 (259), pp. 291-304.

Postolov, Y., Krupnik, A. & McIntosh, K. 1999. Registration of airborne laser data to surfaces generated by photogrammetric means. International Archives of Photogrammetry & Remote Sensing ,32 (3-W14), pp. 95 -99.

Ruiz, A., 2000. A countrywide TIN elevations database. International Archives of Photogrammetry and Remote Sensing, 33 (Part B3), pp. 785-791.

Schenk, T., Krupnik, A., Postolov, Y., 2000. Comparative study of surface matching algorithms. *International Archives of Photogrammetry & Remote Sensing,* 32 (Part B4), pp. 518-524.

Schiewe, J., 2000. Improving the integration of digital surface models. International Archives of Photogrammetry and Remote Sensing, 33 (Part B3), pp. 807-814.

Xu, Z. & Li, Z., 2000. Least median squares matching for automated detection of surface deformations. *International Archives of Photogrammetry & Remote Sensing,* 32 (Part B4), pp. 1000-1007.

ACKNOWLEDGEMENTS

The study has been supported by a funding grant from the Australian Research Council, for the period 2001-2003.

Aspects of this study would not have been possible without the co-operation of David Gibbins and the Surveying Group within Newcastle City Council.

AUTO-EXTRACTION OF URBAN FEATURES FROM VEHICLE-BORNE LASER DATA

Dinesh MANANDHAR, Ryosuke SHIBASAKI

Centre for Spatial Information Science
The University of Tokyo, Japan
dinesh@skl.iis.u-tokyo.ac.jp, shiba@skl.iis.u-tokyo.ac.jp

Working Group IV/7

KEY WORDS: Laser Data, Feature Extraction, Mobile Mapping, Texture Mapping, 3-D Modeling

ABSTRACT:

Laser mapping has become quite popular in recent days due to its capability of providing information directly in three dimensions. These days, we observe the emergence of new demands for 'along-street' 3-D data. 'Along-street' 3-D data provide a real or true visualization of the environment that is different from virtual reality. However, there still lacks a reliable and quick method of acquiring 3-D 'along-street' data of the urban area at higher resolution. Most of the vehicle-borne systems developed so far are based on stereo photographs as the main source of data. In order to overcome the problem of acquiring 3-D data in urban area reliably, quickly and at high resolution, we have developed a vehicle-borne laser mapping system (VLMS). The system is equipped with laser scanners and line cameras for data acquisition. The system will assist in building 3-D GIS database of urban areas.

In this paper we focus our discussions on auto-extraction of building faces from the laser data captured by VLMS. In an urban area we encounter many different types of features like buildings, roads, utility facilities, vehicles, trees and others. Geo-referenced raw laser data were used for extraction of urban features. The raw laser data were simply point cloud data in 3-D space without any other information like intensity value. We have classified the laser data into various classes like building faces, road surfaces, tunnels, utility poles, vehicles and trees. Our discussions will be basically limited to the extraction of building faces with some explanation on extraction of road surfaces, trees and tunnels.

1. INTRODUCTION

The vehicle-borne mobile mapping technology has been developed around late 80's. The development of the mobile mapping system became possible due to the availability of GPS signal to the civilian community. As per our knowledge, the vehicle-borne mobile mapping systems are based on image data acquired by CCD cameras. However, image-based device cannot provide automated extraction of 3-D features and in some cases it even fails due to the lack of correct identification of corresponding control points on the image. On the other hand, a laser scanner provides data directly in 3-D. Thus, automated extraction of 3-D features can be achieved by using laser scanner though it lacks texture information. The lack of texture in laser data can be complimented by acquiring image data simultaneously. This led us towards the development of the VLMS, which is capable of acquiring urban data ('along –street' data) quickly and reliably. This system is capable of observing the objects at closer range, thus giving greater details.

For reference of some of the image based systems, refer GPS-VanTM (Bossler et al, 1991), VISAT-Van (Schwarz et al, 1993, Li et al, 1994), TruckMapTM (Pottle, 1995), KiSS (Hock et al, 1995), GPS Vision (He, 1996) and GeoMaster (Tamura et al, 1998). Refer Manandhar and Shibasaki, 2000 for details about VLMS system description and calibration. Refer Manandhar and Shibasaki, 2001 for details about feature extraction from laser data that are not explained in this paper.

Feature extraction from range data is quite a complex task, as the range data contain nothing but only clouds of 3-D points. The extraction is more complex as the observation is from a moving platform (vehicle). The outdoor observation from a moving platform captures all the points reflected from the features like building faces, road surface, trees, poles, cables, guard rails etc. Thus the data consists of a heterogeneous mixture of laser points. The challenging work is to automatically identify and classify these points reflected by each of these features and then finally make a 3-D model to represent the features. In this paper, we present algorithms to classify these points into different groups and then generation of 3-D models (planar faces) to represent building faces.

2. SYSTEM CONFIGURATION

The mobile mapping system is shown in figure 1. The system is called Vehicle-borne Laser Mapping System (VLMS). This system consists of six line cameras, three laser scanners and Hybrid Inertial Survey System (HISS) which consists of DGPS, INS and an electronic odometer for position and orientation data. The laser scanning frequency is 20Hz, which enables driving at a speed of 40km/hr or more (we have acquired data at 80km/hr) without any serious problem. Data can be continuously acquired as long as the long as the INS / GPS errors (especially the drift error) are within the allowable limit. The along track resolution and across track resolution depends on the vehicle speed and the distance from the sensor to the target respectively. The along track resolution varies from 14cm to 110cm for speed from 10km/hr to 80km/hr. The across track resolution varies from 2cm to 52cm for a range distance from 2m to 60m.

3. FEATURE EXTRACTION

We define feature extraction as the process of classifying the laser points reflected by the urban objects into different groups like building surface, road surface, ground surface (e.g. walking path), trees, poles, tunnel, parked vehicles etc.

The range data includes only clouds of points in three dimensions. Figure 2 shows a sample of range data. It has no other information like reflectance value etc. We would like to extract various features usually seen in the urban environment. Basically, we categorize these features into three classes viz: man-made, natural and dynamic (moving) / static (non-moving) features. The natural features are trees, bush etc. The man-made features are buildings, roads, utility facilities (poles etc), pavement, guardrails etc. The dynamic features include moving cars and pedestrians and the static features include parked vehicles (something which can be dynamic but at static state during the observation period). Our primary focus is on extraction of man-made features like building faces, trees and road surfaces.

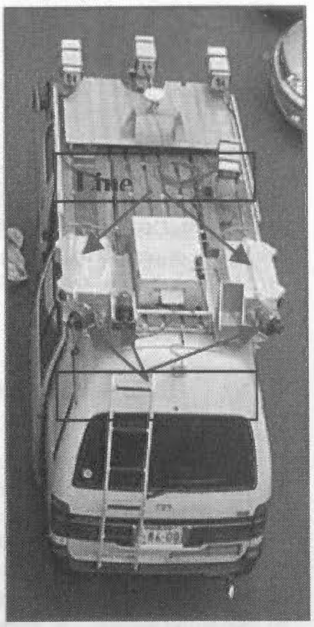


Figure 1: VLMS vehicle

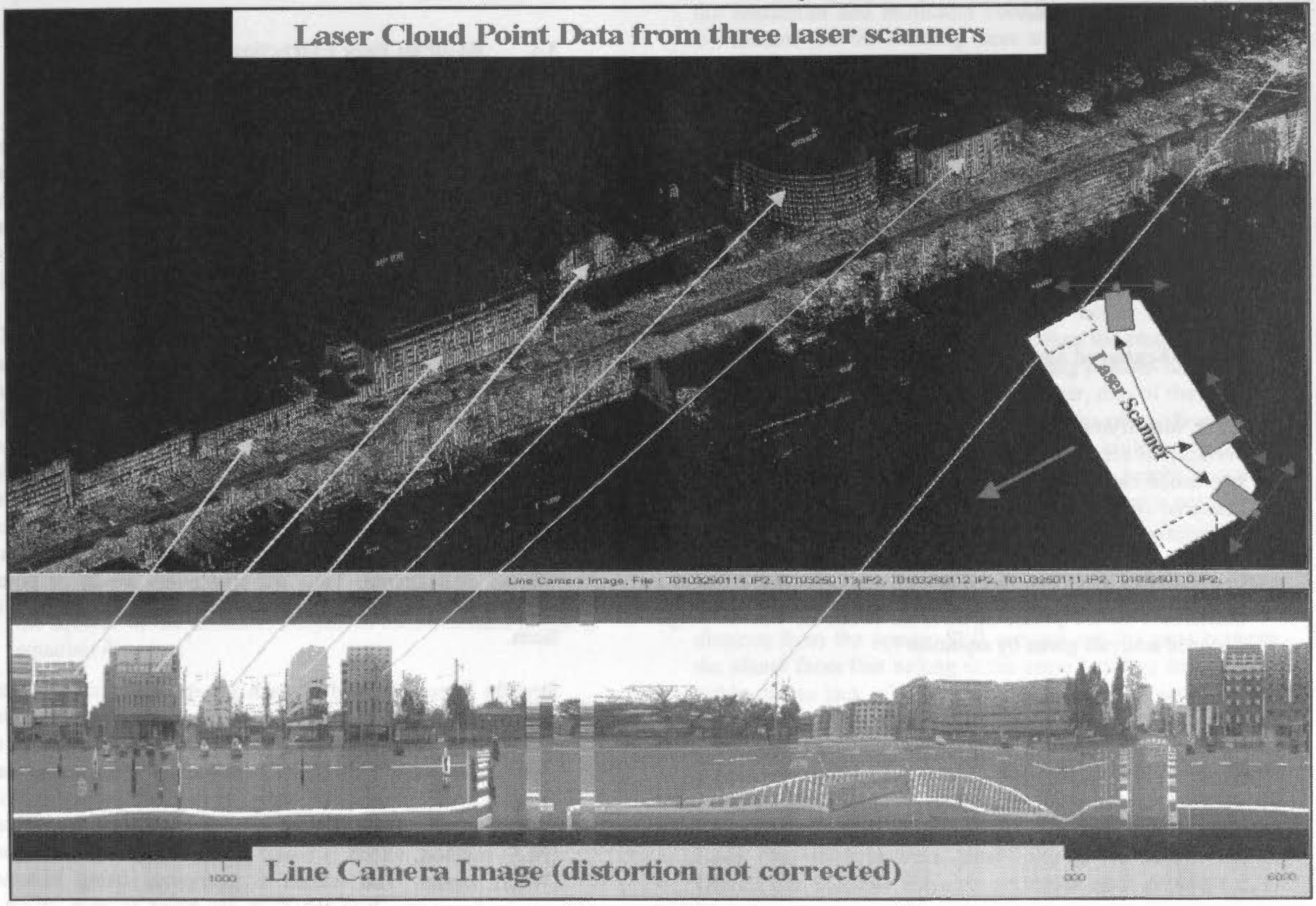

Laser Cloud Point Data from three laser scanners
Laser Scanner
Line Camera Image (distortion not corrected)

Figure 2: Point Laser Data from VLMS. Laser Data from three laser scanners are integrated and shown by red, blue and yellow point

3.1 Basic Concept of Feature Extraction

The laser scanner scans the objects around it's surroundings within 300^0 of angular coverage. As the vehicle moves forward, the scanner scans the objects sequentially at a frequency of about 20Hz. As the system is designed for outdoor mobile environment, we do not have control on viewing the objects as we like. Thus we get quite heterogeneous mixture of range points reflected by all the urban objects within the viewing angle of the scanner. This makes the identification of the objects very complex.

Figure 3 shows some of the laser scan lines that consist of range points reflected by buildings, trees and roads. Generally, we can observe three sets of geometric structure of the range data as shown in figure 3. Building faces reflect points in vertical linear fashion (not necessarily straight line), ground or road surface reflect points in horizontal linear fashion and natural objects like trees reflect points in scattered fashion. Thus a quick look of the data gives us rough indication of buildings, roads and trees. The road surface being the nearest surface to the scanner with respect to other objects, the reflected laser point density is higher compared to other objects. We use this information (geometric structure of points, their scatterness and point density) for feature extraction.

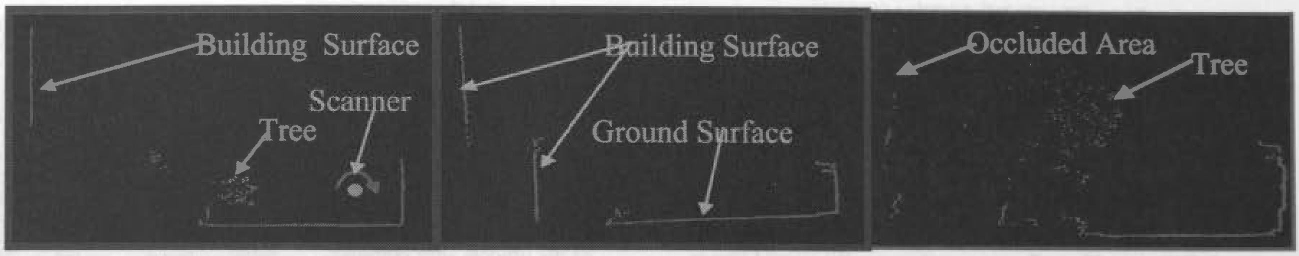


Figure 3: Individual scan lines of laser points. Laser points from natural features are scattered. Laser points from man-made features follow some geometric pattern.

3.2 Road Surface Extraction

Road surface being the nearest object relative to other objects like buildings and trees, the laser point density is also higher due to higher across track scanning resolution. A lot of range points are reflected from the road surface, as there are no obstructions between the scanner and the object. Since, the road surface height varies smoothly with a very small deviation (about 1% slope) along the width of the road, we get reflected range points that have small variation in height value (about 10~20cm).

Histogram analysis of height value is computed for every scan line with a bin width equal to 10cm. The bin with highest value (number of points) is taken and some other parameters to compensate the measurement errors are added to it. Then this value is taken as the threshold value to filter the laser points data are reflected from the road surface. Refer Manandhar & Shibasaki, 2001 for details about it and the results. Figure 4 shows the extracted road surface along with other features.

3.3 Scattered and non-Scattered Points Extraction

After identifying the laser points that belong to the road surface, the remaining points are analysed for classifying into two broad categories of scattered (natural features) and non-scattered (man-made) features. Points from natural features are scattered due to their irregular geometric shape, where as points from man-made features are generally not scattered as they have certain geometric shape. In order to analyse the scatterness of the points in three dimensions, we use range distance analysis given by equation 1.

$$d = \left(d_{n+1} - \left(d_{n+2} - d_n\right)/2\right) \tag{1}$$

where, d_n, d_{n+1} and d_{n+2} are the distance from the sensor position to the nth, (n+1)th and (n+2)th range points on a scan line.

Every transition between the smooth value and scattered value indicates either different features or discontinuity in the same feature (in case of spike like transition) due to location of features at different distance from the sensor or due to partly occlusion.

In order to ease the feature classification, we have analysed the standard deviation (SD) of every cluster of points along the x, y and z coordinates separately. Our hypothesis is that, points from building surface will have much higher SD value in z-coordinates compared to x and y coordinates. Similarly, a ground surface like footpath will have much higher SD value in x or y coordinates compared to z coordinate. Trees exhibit high SD values in all the three directions due to their scattered nature of range points. Point clusters that have smaller SD value on z coordinate (or less than the across track resolution) are classified as vertically oriented points. Point clusters that have higher SD value for either x or y are classified as horizontally oriented points. If the SD value is higher for all the three components then they are classified as scattered points, which generally belong to natural features like trees.

3.4 Building Face Extraction

The laser data as shown in figure 2 are processed scan line by scan line by analysing the geometric orientation of the points on each scan line. As mentioned earlier, the laser data are first processed for the extraction of road surface and then again analysed for identifying the natural and man-made (scattered and non-scattered) features by using the derivative analysis of the range distance. The man-made features exhibit certain geometric properties that are either linear, planar or follow some other geometric shape, but are not scattered. The derivative analysis identifies the points reflected by the man-made features as well as classifying these features into different geometric shape categories like vertically aligned or horizontally aligned points. Vertically aligned points are taken for extraction of the building faces. It is not necessary that the points should align strictly vertical, which is not possible in practice. One scan line may have more than one group of points that are vertically aligned, horizontally aligned or scattered. Thus we take every group of points, which are vertically aligned for the extraction of building faces.

Straight lines are fitted to each group of vertically aligned points on each scan lines. After fitting the line, any points that are beyond the threshold distance from the fitted line are rejected. The line is then fitted once again. This is necessary to remove any possible inclusion of the outliers in the fitting data. The line fitting is done by using a robust least square (RLS) method, which is based on the weighted least square (WLS) model. The weight is computed using bi-square weight function. Once the lines are fitted to each group of scattered points, a line connectivity table is generated. This table is generated by using the line-to-line vertical centre distance and direction vector of a pair of line between the line segments within the scan line as well as between the line segments to the next neighbouring scan line. This helps us to identify the line segments that can be grouped together within the scan line or to the neighbouring scan line to form a surface. This will also help us to separate two different types of surfaces like building face and guard rail or hoarding boards as they are at different range distance as seen from the sensor position.

After creating the line connectivity table, we can simply take a pair of lines that can be connected together and form a patch surface from them. The results of this are shown in figure 6. The laser point data for this result is shown in figure 5. However, due to occlusion or glass windows, there are many empty spaces in the generated surface. When a laser beam pass through the glass window, it is either lost in the space (as the laser scanner can not receive the reflected signal within the prescribed time) or reflected by inner objects like blinds, curtains etc. In the later case, the laser points appear as scattered data and are classified into the natural class (trees) instead of man-made feature class. Since the surface is generated by using only a pair of line, it can generate non-planar surface as well. The problems are however, the empty spaces between the surfaces and irregular shape of the surface. This type of surface generation is not effective for texture mapping as we will have numerous small size surfaces. Due to the fitting errors and some other errors the surfaces that belong to a single building face might not align perfectly. The normal vector of each surface might be pointing at different direction. In order to rectify these problems we have further enhanced the algorithm as explained below.

All the line segments that can be grouped together are selected based on the information from the line connectivity table. Once the lines are identified, the points that are used to fit these lines are identified. Thus we will have a group of scattered points, which are used for fitting the lines. This process leads us a few steps backward again. Of course, we could have taken only the end points of the fitted line for fitting the plane. But, this will induce the fitting error of the straight lines as well. Next, we fit planar face to identify the building surface from these points using the well-known plane equation model given by equation 2.

$$Ax + By + Cz + D = 0 \tag{2}$$

The fitting is done to minimise sum of the squares of the distance from the points to the fitted plane rather than to minimize the sum of the squares of the fitting errors, given by equation 3. Equation 3 is solved to minimize the sum of the squares of the distances subject to the constraint given by equation 4.

$$\|d\| = [Ax + By + Cz + D = 0] / \sqrt{A^2 + B^2 + C^2} \tag{3}$$

$$A^2 + B^2 + C^2 = 1 \tag{4}$$

The Fitting model is solved by using a singular value decomposition method. The points are then projected back to the fitted plane using equation 5.

$$\begin{aligned} Px &= x - A(Ax + By + C + D) \\ Py &= y - A(Ax + By + C + D) \\ Pz &= z - A(Ax + By + C + D) \end{aligned} \tag{5}$$

In order to regenerate the planar surface, the centroid of the projected points is taken as the reference point for the parametric generation of the plane. The surface is generated using a parametric model of the plane equation given by equation 6.

$$X = X_0 + \alpha R + \beta S \tag{6}$$

The vectors R and S in equation 6 are two orthogonal vectors to the normal vector of the fitted plane and they lie on the fitted plane. Hence, the normal vector, vectors R and S are all orthogonal to each other. The parameters alpha and beta are any real numbers and are increased or decreased at a step of 0.5 till the regenerated surface exceeds the boundary defined the maximum and minimum coordinates of the point data. We assume that the building faces are planar in shape though this may not be true for a few buildings. There might be buildings with curved or irregular shape. Irregular shape means a mixture of planar face as well as curved face within a single front face. However, at the moment, for simplicity, we have considered only the planar faces, which can explain most of the building shapes.

Figure 7 shows the results of planar face fitting to laser point data that belong to the building face. The results show that the planar face thus generated can better represent the building faces. The results are better than the results shown in figure 6, which are obtained by drawing patches for a pair of connecting neighbouring lines. However, one of the difficult tasks is to determine the correct size of the planar face so that it can exactly represent each individual building face. For example, identification of correct building height is not so easy. The uppermost rows of laser points may not always represent the top of the building. Another difficulty is to identify the exact boundary between the two similar buildings. The algorithm will produce one single planar face for the buildings that are joined together at the same range distance from the sensor. Still, we can see the gaps between the planar faces that belong to the same building face. This is either due to lack of point data (probably from occlusion) or due to the scattered data (in case of glass windows). The occlusion may be by trees, pedestrians, vehicles and signboards or hoarding boards. Thus the algorithm developed so far though much improved than it's previous version, still lacks the capability of exact representation of the building faces. We will add these capabilities in the future version. One of the possible ways to overcome such problem is to analyse the normal vector of the fitted plane. If the neighbouring normal vectors point to the same direction (analyse the angle between the normal vectors) within a small threshold value and if the distance between the planes (horizontal distance) is very small (a few centimeters, around the across track resolution value), then those planes can be merged together to form a new and bigger planar face. Also, it is necessary to merge all the three data sets taken by three laser scanners. Since, they have different viewing angles, the problems of identifying boundaries can be solved to some extent. Finally, the image data will be used for texture mapping. The texture will also help to determine the boundary between the buildings.

3.5 Extraction of other Features

We have also extracted other urban features like tunnels, utility poles and vehicles using only laser point data. The extraction of tunnels is quite successful. The extraction of vehicles are partly successful and needs further development in the algorithm. The extraction of poles though partly successful is limited by other factors as well. For example, the speed of the vehicle (VLMS) defines the along track resolution. If the vehicle speed is high (80km/hr or so), it is difficult to capture poles as the resolution is about 1.1 metre and normally the pole diameter is less than this. Thus, in order to capture the utility poles, the vehicle speed should not be above 40km/hr, which gives along track resolution of about 50cm. Refer Manandhar and Shibasaki, 2001 for details.

3.6 Conclusion

We have developed algorithms for the extraction of building faces from laser data taken by vehicle-borne laser mapping system. The results of the algorithm are satisfactory. It can successfully classify the laser points that belong to the building faces and generate planar faces to represent these building faces. However, the algorithm can not exactly determine the size and shape (this means to determine the shape of the building accurately within the accuracy of the data, which is around one metre) of the building face and cannot determine the boundary between the two similar buildings. Thus the results from this algorithm may be good enough for visulatisation purpose rather than exact spatial analysis. We will include these capabilities in the future version.

3.7 References

References from Books:

Medioni, G., M.S. Lee, C.K. Tang, 2000, A Computational Framework for Ssegmentation and Grouping, Elsevir Publication, ISBN: 0 444 50353 6

Mikhail, E.M., 1976, Observations and Least Squares

References from Other Literature:

Bossler, J.D., Goad, C.C., Johnson, P.C., and Novak, K., 1991, GPS and GIS map the nation's highways, GeoInfo Systems, March 1991, p.27-37

He, G., 1996, Design of a mobile mapping system for GIS data collection, IAPRS, Vol. XXXI, part B2, pp.154-159

Hock, C., W. Caspary, H. Heister, J. Klemm, and H. Sternberg 1995, Architecture and design of the kinematic survey system KiSS, Proc. of the 3rd Int Workshop on High Precision Navigation, Stuttgart, April, p. 569-576

Hoover, A., Jean-Baptiste, G., Jiang X., J., Flynn, P.J., Bunke H., Goldgof, D., Bowyer K., A Comparison of Range Image Segmentation Algorithm, URL: http://marathon.csee.usf.edu/range/seg-comp/SegComp.html

Li, R. Mobile Mapping – An emerging technology for spatial data acquisition, http://shoreline.eng.ohio-state.edu/ron/teaching/894a/paper1.htm

Manandhar, D., Shibasaki, R., 2000, Geo-referencing of Multi Sensor Range Data for Vehicle-borne Laser Mapping System, 21st Asian Conference on Remote Sensing (ACRS), 4-8 December, Taipei, Vol 2, pp. 932-937

Manandhar, D., Shaibasaki, R., 2001, Proceedings of ACRS 2001–22nd Asian Conference on Remote Sensing, 5-9 November 2001, Singapore, Vol. 2, pp 1113 – 1118Pottle, D., 1995, New mobile mapping system seeds data acquisition, GIM, Vol 9, No. 9, p51-53

Schwarz, K.P., H.E. Martell, N.El. Sheimy, Ron Li, M.A. Chapman, and D. Cosandier 1993, VISAT – A mobile highway survey system of high accuracy. Proc. of the IEEE vehicle navigation and information systems conference, October 12-15, Ottawa, pp. 476-481

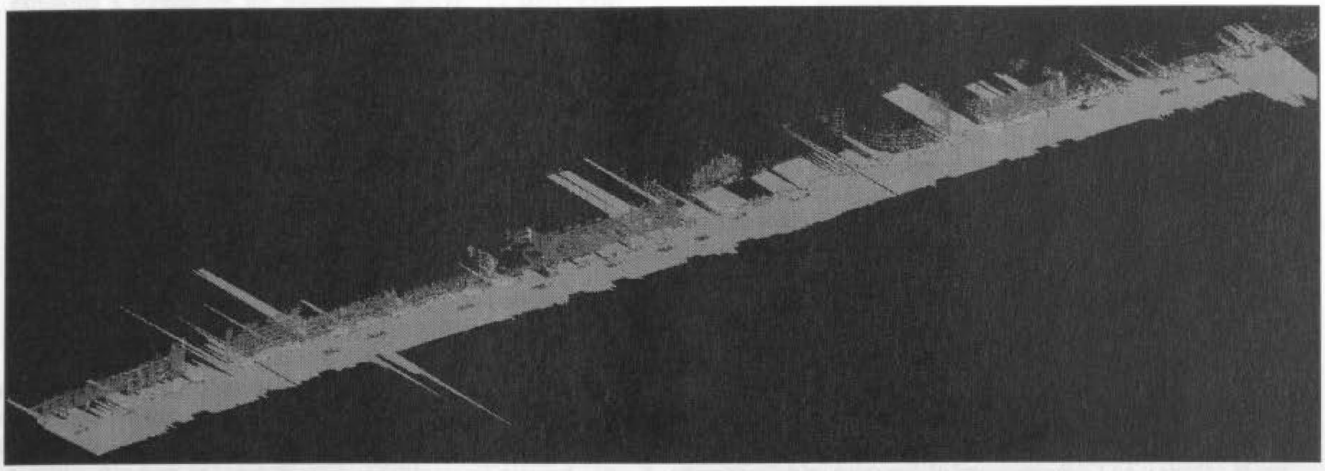


Figure 4: Classified laser point data. Red color shows the patches created from the fitted straight lines that describe building faces, cyan color represents the road surface, green color represents scattered points classified as tress and blue color represents vehicles

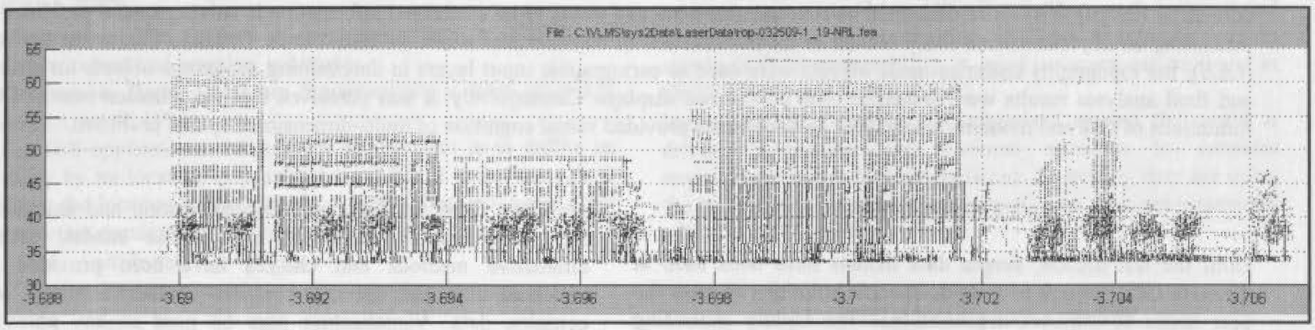


Figure 5: Laser point data as acquired by VLMS.

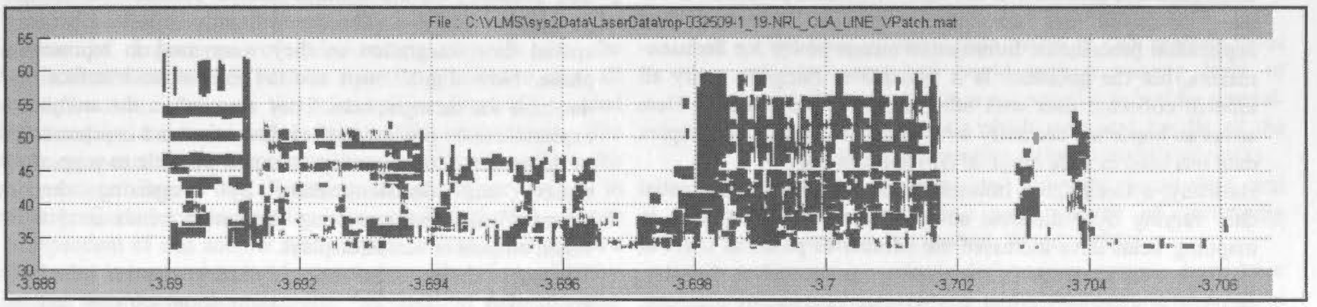


Figure 6: Patches to represent the Building Faces from classified data. The classified data are the points that are marked as vertically aligned

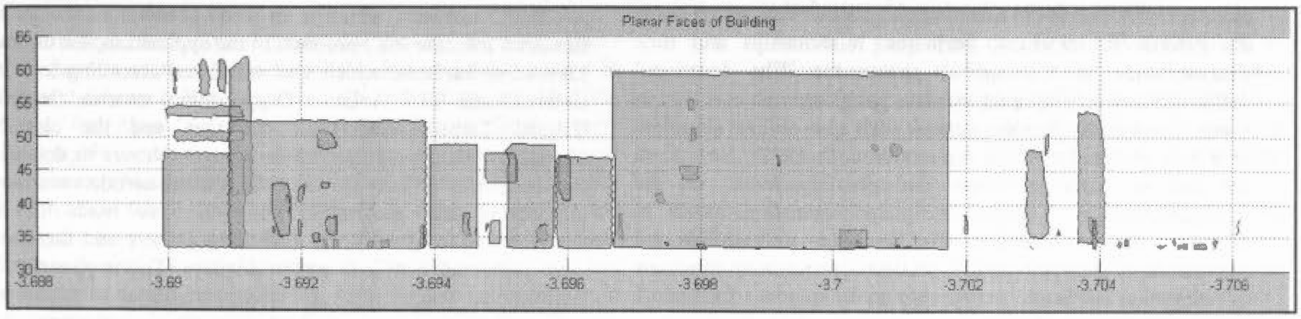


Figure 7: Planar Faces to represent the Building Faces from classified data. The classified data are points that are marked as vertically aligned

A DIFFERENT APPROACH TO THE SPATIAL DATA INTEGRATION

A. Ulubay[a], M.O. Altan[b]

[a] General Command of Mapping, Cartography Department, 06100 – Ankara - Turkey – (aulubay@hgk.mil.tr)
[b] ITU, Civil Engineering Faculty, 80626, Maslak, Istanbul, Turkey – (oaltan@itu.edu.tr)

KEY WORDS: GIS, Obstacle Limitation Surface, Visualization, Spatial Data Integration, Analytical and Visual Integration

ABSTRACT:

The rapid changes in map-based technologies caused a complexity in spatial data and increased user demands and expectations from Geographic Information Systems (GIS). Considering this reality, Spatial Data Integration (SDI) needed to be taken into account by National Mapping Agencies and companies for developing strategies to have better data (complete, standard, robust and flexible) for a wide public use of GIS applications. On the other side, visualization became an essential process in data integration facilities as it was used in every stage in GIS, besides providing useful cartographic representing capabilities in 3-D environment. This paper presents an overview of spatial data integration from different aspects and explores the role of visualization in this context. In the content of this paper it is also examined how to use analytical and visual types of derived information in solving spatial problems by presenting an implementation study related to the construction of obstacle free zones for aerodromes. Obstacle Limitation Surfaces (OLS), the topography and man-made objects were used as cartographic input layers in determining dangerous objects for aircraft and final analyses results were presented with 3-D screen displays. Consequently, it was perceived that visualization removed the limitations of GIS and modern cartographic developments provided visual cognition of multi-dimensional spatial problems.

1. INTRODUCTION

Until the last decade, simple data models have been used in different GIS projects to provide specific solutions to meet the user needs in different organizations. The rapidly expanding ranges of spatial data and increasing application needs caused some gaps and problems in the data integration processes about their combined use associated with data handling and application procedures. Information means power for decision-makers, but the question: Is it possible to integrate easily all kind of collected data with other available information? Here arises an important necessity for a strong integration of complex data sets used in wide range of GIS applications.

Basically, a Geographic Information System deals with spatial data varying from different sources. Recent developments in mapping fields have increased the number of potential users of GIS users in many sectors including transportation, forestry, land registration, agriculture, environment and facility planning. All these disciplines are trying to obtain better data for better solutions. The data structure built upon a data model became more complex due to the nature of geographic data, which holds the records of locations, attributes, relationships and time characteristics of the spatial phenomena. The locational definitions are very complex because geographic objects tend to occur in irregular complex patterns such as a sinuous shoreline or a web of transformation routes (Aronoff, 1991). Now there are just as many problems _and possibly more_ on the management side of implementing an information system as there are on the technical side. In order to highlight the importance of data integration, section 2 presents an overview of integration needs and trends; section 3 explores a framework of an integration model.

In the applied GIS projects decision-based or data-based models were used to obtain the rapid answers. These approaches and traditional techniques used in GIS applications may not be adequate for distinct problems. Current geographic analyses require more than giving attributes or location information. Today new techniques _like cartographic communication and visualization_ are used for better understanding of dimensional problems. With addition of 3-D visualization and animations into the general design of cartographic model, several alternative methods and choices have been provided for understanding and solving complex spatial problems with complex data. Visualization may be used in any phase of geographic processes and in representations to display the shape and geometry of the features and to examine the results of geographic analyses. These new trends became a process of spatial data integration as they were used in representation phase. Now digital maps are the interactive interface to the database for cartographers. They are used in the analyses and explorations by predicting the information and in representation of the results. Consequently, it is also possible to support map-makers and decision-makers by visualizing the data characteristics and extracting the secret trends in data with visual outputs or screen displays.

This paper shows visual and analytical integration in an airport application by deriving information about objects and using combination of geometric and attribute data. In the application study data sets from different sources were integrated in Arc/INFO software, which is an ESRI (Environmental System Research Institute Inc.) product. In the application, the Obstacle Limitation Surfaces, which was calculated according to ICAO (International Civil Aviation Organization) criteria, the DTM (Digital Terrain Model) of the area and the classified planimetric data were used as cartographic layers to determine the dangerous objects for aircrafts around aerodromes. Some predictions and geographic analyses were made in 3-D environment to identify the exceeding objects and the results were presented with 3-D screen displays. The implementation showed that visualization sometimes might be a critical and essential process in geographic analyses requiring third dimension of the related objects. The results also showed that the obtained cartographic 3-D displays help users in the validation of mathematical calculations and in the other processes in geographic information systems.

In this paper, it was not aimed to make a comparison between spatial data integration methodologies; but only contented with exploring a framework for integration of complex geographic datasets.

---

* Ali ULUBAY

2. SDI CONCEPTS AND NEEDS

To understand the requirements of SDI, a firstly description of the basic concepts of spatial phenomena is needed.
The term ***"spatial"*** generally means, having extensions in space or pertaining to space.
Other descriptions are as follows;
Spatial Data: 1) Can be most simply defined as information that describes the distribution of things upon the surface of the earth. In effect any information concerning the location, shape of, and relationships among geographic features (Walker, 1993; De Mers, 1997).
2) Being those that can be individually or collectively referenced to a location (Tomlinson, 1990)
Spatial Feature: A user defined geographic phenomenon that can be modeled or represented using geographic data sets (ESRI, 2002).
Integration: The process of integrating data to obtain and deliver robust and practical solutions.
Geographic Data: Those are fundamentally a form of spatial data (Aronoff, 1991).
For GIS applications the primary requirement is to define an entity by its location in a unique way. But it is not enough to know the location of geographic objects. There are some other important requirements for a geographic entity like;

- attributes,
- relationships with other geographic features,
- time,

An attribute is known as non-spatial information that does not represent location information. The relationships among geographic features may be very numerous and complex. The degree of importance of relationships depends on the application used in GIS, because it is not possible to store all the information about all the relationships. Generally needed relationships are stored and defined in GIS applications. The time component is very important because it directly affects the appropriate usage of data. Knowing the time is very critical in GIS projects because location data attributes often change independent of one another with respect to time. This causes a difficulty in management of spatial data. Therefore, effective spatial data management requires that location data and non-location data be variable independent of one another. Attributes can change character but retain the same spatial location, or vice versa (Dangermond, 1990).
As it is seen, the nature of geographic data is very complex and it is hard to store and handle all the data and relationships in large databases. But for years GIS technology provided the basis for many types of decisions ranging from simple geographic analyses to management of complex spatial problems. The scientists have developed some useful techniques and tools in the computer environment to store, manipulate and analyze geographic data. But integration methodologies have not improved as developments in the data collection and solution of geographic problems. Now it arises as a problem to integrate spatial data sets, which they may be collected:

- in different time periods,
- for different purposes,
- from different sources,
- in different scales,
- in different accuracies,
- in different Datum,
- in different projections,

The articles listed above are giving an idea about SDI processes. A user must have a complete understanding of the content of digital spatial data sets and their qualities in order to make maximum use of its information potential. Scientists and general public want to be able to conflate multiple sets of spatial data i.e. integrate spatial data from different sources (Saalfeld, 1998).
Today spatial data integration using multiple geo-based data has been regarded as one of the primary GIS application issues, because improved geographic data should lead to better conclusions and better decisions. In various GIS applications integrated data provide many advantages to the users and they facilitate interoperability of spatial data, because they include greater positional accuracy and logical consistency and completeness. But each new data set and item can only be utilized only if it can be placed correctly into the context of other available data and information (UCGIS, 1998). Many organizations and data users have developed and promoted standards for spatial data collection and representation. Besides national agencies, Open GIS Consortium and International Organization for Standardization (ISO) are working to develop standards for interoperability. A good summary can be found in GETF (1996) and NAPA (1998).
Increasing environmental consciousness around the world is driving companies and national agencies to consider environment issues in their decisions. Therefore, they are using the international standards to better manage their environmental affairs (UCGIS, 1998).

3. VISUALIZATION AND SDI

If traditional GIS applications are criticized from different aspects, following considerations may be comprehended:

- The level of object data model in GIS projects tends to represent real world with a varying level of completeness. But it is not possible to build a general-purpose data model, which store and handle all the relationships representing different aspects of reality.
- Geographical data are generally incomplete and almost always have some degree of uncertainty (Goodchild, Buttenfield, Wood, 1994).
- Current geographic information systems have been designed for querying and maintaining static database representing static phenomena and give little support to those users who wish to represent dynamic information (Hornsby, Egenhofer, 1997).
- The use of geographic data from different sources, in different scale is totally inappropriate and may cause to significant errors in geographic data processing.
- Current GIS have some limitations in solving dynamic volumetric problems and visualizing characteristics of complex data sets.

However, it was required to obtain the maximum use from available data sets stored in large databases.
In the past ten years, the role of visualization has increased because of the developments in computer _hardware and software_ technology. Improvements in scientific visualization stimulated cartographers and GIS specialists to use its visual capabilities in various stages of GIS applications. Today, visualization provides visual thinking, cognition and interactive communication capabilities in solving spatial problems and gives some easy extraction methods for error detection in spatial data. Although visualization has been the cornerstone of

scientific progress throughout history, the term scientific visualization (ViSC) has recently come to refer to the exploration of data and information in such a way as to gain understanding and insight into these data. Generally this is conducted in powerful three dimensional graphical computer environments (Hearnshaw, Unwin 1994). It is clear that seeing is a good way towards understanding of spatial phenomena.

If made a description: traditionally, the term visualization has been used to describe the process of graphically conveying and presenting end results (Dibiase and others 1992; MacEachren and Kraak 1997).

Recent concern over spatial data integration comprising data quality has increased the interest to the visualization tools to comprehend the reliability of data. The current trends in development of information systems are to move from specific applications to complex environments. This requires integration of large amount of data from different sources and capability of integrated processing. Data integration processes should provide robust and reliable information to serve a wide range of GIS users. Integrating spatial data generally comprises projection, precision, accuracy and scale. Spatial Data Integration demands may be considered in two parts (Kobben, 1999). Visual Integration demands, which comprises the processes of,

- Matching coordinate system and projection
- Compatible scale (re-scaling should be possible only within certain limits)
- Compatible file formats (originally equal or converted)
- Visualization made by especially for the combination

Analytical Integration demands, which comprises,

- Visual integration demands
- Attribute matching (accuracy, classification)

These two steps including some processes are considered in digital cartographic model at different levels.

Visualization is used in every stage of GIS processes, traditionally in the;

- collection of data,
- error detection and quality control,
- checking attribute values,
- geographic analyses,
- representing end results.

and now it must be also used in SDI to have robust, complete, standard, classified and flexible geo-databases. The new trends in modern cartography and visualization provide following advantages:

- strong interaction,
- easy interpretation
- better understanding
- easy communication

But they all need is integrated _re-engineered_ spatial data. Although there is not any standardization in this context, it may be thought that they all fulfill the requirements of new trends in dynamic cartography and GIS.

After the integration of data and establishment of GIS, the degree of accomplishment can be measured by rating the following questions:

- Ability to accept (?%)
- Ability to process (?%)
- Ability to analyse (?%)
- Ability to present (?%)

If the total percent is high, then it means that the amount of reliable information and solutions to be extracted will be high.

Visualization encompasses the display of quantities or qualities of visible or invisible phenomena through the combined use of points, lines, a coordinate system, number, symbols, words, colour and animation (Beard K., William M., 1993).

Generally it can be said that visualization has become one of the most important process in geographic information systems and in data integration procedures for making data comprehensible, accessible and manageable. There is a strong need to integrate all the data before establishing large databases for nations.

4. IMPLEMENTATION: DETERMINATION OF OBSTACLES NEAR AIRPORTS

Problem:

Developments in air transportation are presenting many problems besides its advantages. The most important problem is that the obstacles near airports which may cause accidents. These obstacles may be natural (trees, hills, etc.) or man-made constructions like towers, antenna, power lines and high buildings. There are Obstacle Limitation Surfaces that should be implemented around airports to provide safe zones (obstacle free zone) for aircrafts to fly safely near airports. These surfaces are:

- approach surface
- transitional surface
- conical surface
- inner horizontal surface
- outer horizontal surface

which having slopes and distances from runway accent. International Civil Aviation Organization - ICAO, has described the criteria and parameters of these surfaces to be implemented for all airports. These criteria may change from one another with respect to their characteristics. Detailed information about the surfaces and parameters can be found in Annex-14 of International Standards and Recommended Practises (ICAO, 1999). The problem is to determine automatically the obstacles (natural or man-made), which exceed the limitation surfaces and become dangerous for aircrafts. With the implementation study a user interface was presented to find and query all features, which are dangerous. Visual displays, which can be prepared interactively, provided better understanding of volumetric spatial problems relevant to OLS and other features with their attributes.

Solution and Results:

This part presents a quick review of data, calculations and processes used in GIS application and concentrates the visual results with highlighting the importance of 3-D cartographic analyses in understanding and solving dimensional problems.

The information about data sources used in the implementation and their structures are as follows:

- ICAO Criteria for OLS,
  ASCII file, used in the calculation of OLS like DTM,
- 1:25 000 scale digital contours of the area, in DGN format,
- Vegetation
  Digitized from existing maps in 1:25 000 scale
  Attributes were collected by surveys
- Buildings and other planimetric features, digitized from 1:1000 and 1:5000 and 1:25 000 scale maps
  Height attributes are taken from municipality
- Geological Data, digitized from 1:25 000 scale maps, attributes are obtained by geological surveys

- Hydrological data, digitized from 1:25 000 scale maps

All the digital data transformed to Universal Transvers Mercator projection system in European Datum-1950 and format conversions were made in order to use them in Arc/INFO software. The height values of Obstacle Limitation Surfaces were calculated with Fortran program and the output ASCII file was processed again in Arc/INFO software in order to obtain it in GRID format. Figure.1 presents a visual display of calculated OLSs.

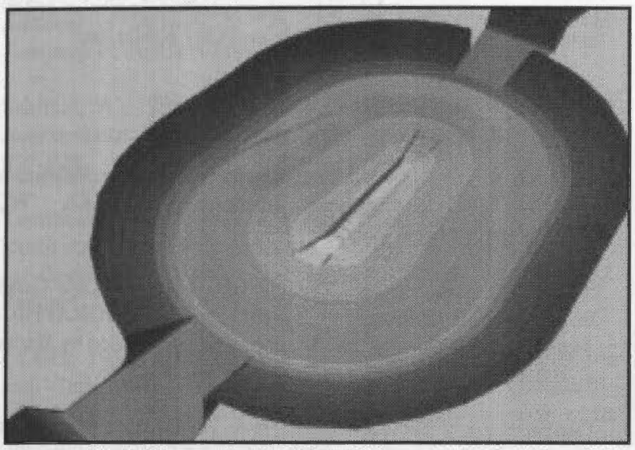


Figure 1: Obstacle Limitation Surfaces

The DTM of the area was obtained in TIN (Triangulated Irregular Network) module of Arc/INFO from contour information that is available. These two layers were processed cell by cell in GRID module to calculate the exceeding areas of the terrain. These areas can be calculated with the equation (Ulubay, 1999):

$$\text{Dangerous Area}^{\text{Matrix}} = \text{DTM}^{\text{Matrix}} - \text{OLS}^{\text{Matrix}} \quad (1)$$

The positive values obtained by above equation constitute dangerous area for aircrafts. The combined display of these two layers was given in Figure.2. In this figure it is obviously seen the dangerous parts of the terrain that form an obstacle for airplanes.

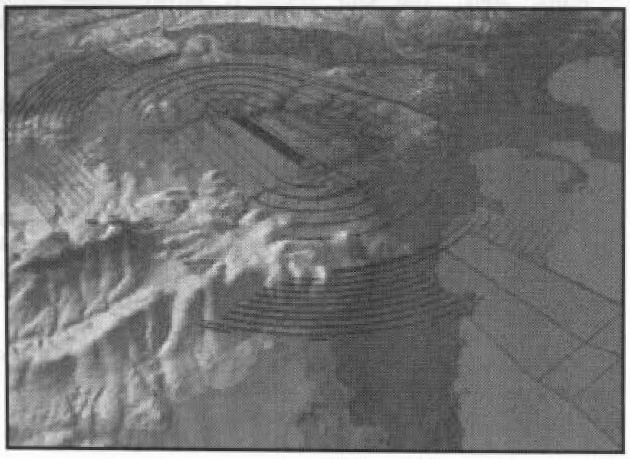


Figure 2: Combined view of OLS and DTM

The geological structure of the area shown in Figure.3 and buildings animated with height attributes shown in Figure.4 are other perspective displays, which provide a general view of the area. Figure.3 also shows that under ground can also be visualized with attributes that are very important for excavation of dangerous hilly areas.

In the final analyses, the spatial relations between buildings, OLSs and terrain in 3rd dimension was examined by visualizing the characteristics of the situation. Any perspective or zoomed views from any direction can also be obtained for specific investigations. Figure.5 and Figure.6 show generally the buildings and natural objects, which form obstacles for aircrafts.

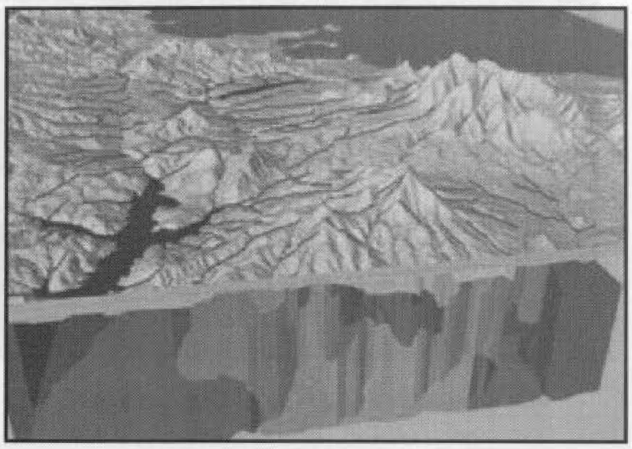


Figure 3: Geological structure

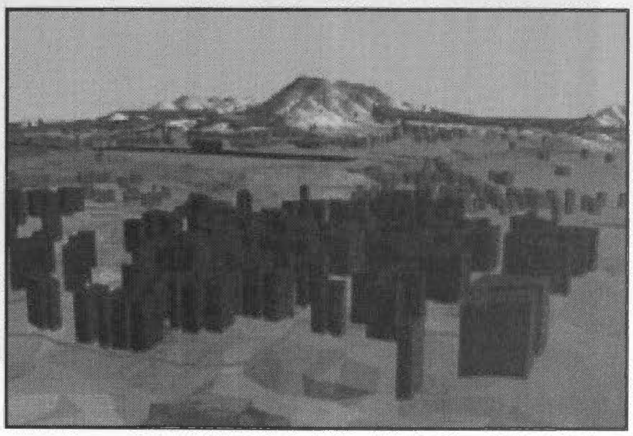


Figure 4: Buildings near airport

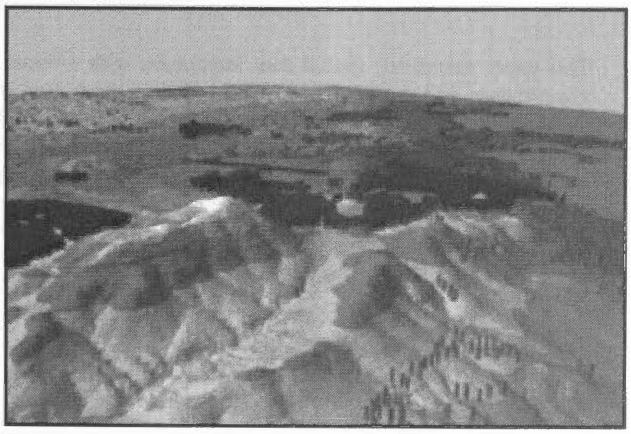


Figure 5: Perspective view of airport and buildings

A detailed cartographic analysis in 3rd dimension can be seen in Figure.7, which explains shielding method for new buildings to be constructed for determination of if they exceed OLSs or not. Different applications are used for shielding methods in different countries.

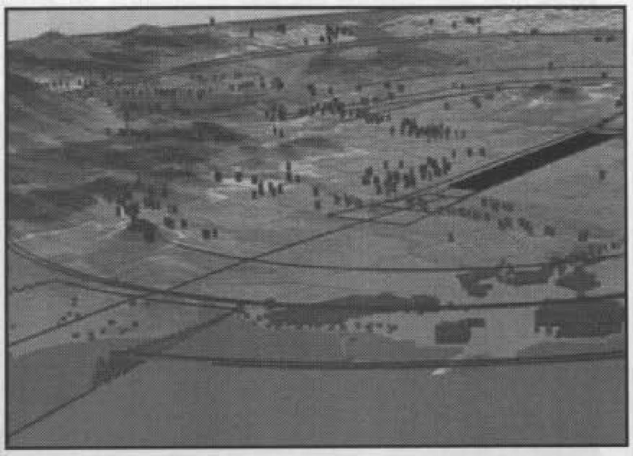


Figure 6: General view of dangerous areas

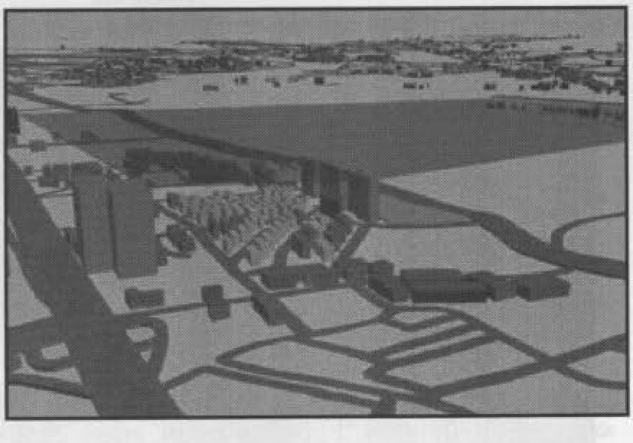


Figure 7: General view of dangerous areas

5. RESULTS AND CONCLUSION

This paper examined; spatial data integration with a focus on visualization and outlined that it is needed in every step of a GIS application and in solving dynamic spatial problems in 3-D environment. Investigations showed that spatial data integration and visualization are very essential to be able having rapid solutions and better understanding of complex dynamic problems. It was explored that completeness, robustness, versatility, efficiency and ease of generation in geographic data sets can be obtained by spatial data integration procedures. On the other hand visualization tools provide abilities to make interactive analyses, real-time modelling, future analyses and dynamic representations with these data in GIS applications.

The implementation study showed that integrated spatial data, from different sources and visual types of derived information, could be combined in solving 3-D problems with visualization tools. The results also showed that carefully constructed views and 3-D cartographic displays help users in showing data quality and analyses and they also may be used to validate mathematical analyses.

It was perceived that visualization removed the limitations of Geographic Information Systems and modern cartographic trends; visual thinking and interactive map communication provide visual cognition of complex dimensional problems.

Finally, can be said that future success of Geographic Information Systems depends on the progress in visualization and developments in data integration methodologies in order to establish global database for wide and easy public use.

REFERENCES

Aronoff, S. 1991. Data Management, *Geographic Information Systems: A Management Perspective*, pp.162, WDL Publications Ottawa, Canada.

Beard K.; Mackaness W., 1993. Visual Access to Data Quality in Geographic Information System: in CARTOGRAPHICA: *Mapping Data Quality,* pp. 37, compiled and edited by Barbara P. Buttenfield, University of Toronto Press, Canada

Dangermond, J. 1990 A Classification of Software Components Commonly Used in Geographic Information Systems, pp.30, in *Introductory Readings in Geographic Information Systems*, edited by Donna J. Peuquet, Duane F. Marble, Taylor & Francis.

De Mers, M.N., 1997 Fundamentals of Geographic Information Systems, New York, John Willey & Sons.

Di Biase, et.al., 1992. Animation and the Role of Map Design in Scientific Visualization in *Cartography and Geographic Information Systems*, 19(4).

Donna J, Peuquet., 1990. A Conceptual Framework and Comparison of Spatial Data Models, in: *Introductory Readings in Geographic Information Systems*, pp. 250, edited by Donna J. Peuquet, Duane F. Marble, Taylor & Francis.

ESRI, 2002. The GIS Glossary, Environmental System Research Institute, Inc. Redland, California. http://esri.com/library/glossary/glossary.html (accessed 18.Jan.2002).

GETF, 1996. EARTHMAP: *Design Study and Implementation Plan*, Annandale, Global Environment & Technology Foundation, p. 57.

Goodchild, M.; Buttenfiel, B.; Wood, J., 1994. Introduction to Visualizing Data Validity, in: *Visualization in Geographical Information Systems*, pp. 141, edited by: Hilary M. Hearnshaw & D. J. Unwin, The Association for Geographic Information.

Hearnshow, H., M.; Unwin, D. J., 1994. The Process, *Visualization in Geographic Information Systems*, AGI, pp xii, John Willey & Sons Ltd., England.

Hornsby, K. ; Egenhofer, M., 1997. Qualitative Representation Chpice, In Spatial Information Theory, edited by S. Hirtle and A. Frank.

ICAO, 1999. International Civil Aviation Organization, Annex-14, *Aerodromes, International Standards and Recommended Practices*, Third Edition.

Kobben, B. 1999 *Cartographic Visualization and Spatial Data Integration* (lecture notes), ITC.

MacEachern, M.; Kraak, M.J. 1997. Geographic Visualization: Advancing Agenda in *Computer & Geosciences* 23(4).

NAPA 1998 *Geographic Information for the 21st Century: Building a Strategy for the Nation*, Washington National Academy of Public Administration, p. 358.

Saalfeld, A., 1998. Conflation: Automated Map Compilation, *International Journal of Geographic Information Systems*, 2(3), 217-228.

Tomlinson R. F., 1990. Current and Potential Uses of GIS: The North American Experience, pp. 142, in *Introductory Readings in Geographic Information Systems*, edited by Donna J. Peuquet, Duane F. Marble, Taylor & Francis.

UCGIS, 1998. Research paper on the Spatial Data Acquisition and Integration, University Consortium for Geographic Information Science, http://www.ucgis.org/research_white/data.html (accessed 20.Feb.2002).

Ulubay, A., 1999 Surface Analyses and Simulation Model to Determine Dangerous Objects For Aircraft's Landing and Take-off, *PhD. Thesis*, Istanbul Technical University, Istanbul.

Walker, R., 1993. AGI Standards Committee GIS Dictionary, Association for Geographic Information.

FUSING STEREO LINEAR CCD IMAGE AND LASER RANGE DATA FOR BUILDING 3D URBAN MODEL

Masafumi NAKAGAWA*, Ryosuke SHIBASAKI**, Yoshiaki KAGAWA*

Graduate School of Frontier Sciences, Institute of Environmental Studies*
and
Center for Spatial Information Science**
University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505
- mnaka@iis.u-tokyo.ac.jp

Commission IV, WG IV/7

KEY WORDS: 3D urban model construction, Data fusion, Line sensor image, Laser range image, Stereo matching Image segmentation, High-resolution image

ABSTRACT:

Photogrammetry is a current method of GIS data acquisition. However, as a matter of fact, a large manpower and expenditure for making detailed 3D spatial information is required especially in urban areas where various buildings exist. There are no photogrammetric systems which can automate a process of spatial information acquisition completely. On the other hand, laser range finder has high potential of automating 3D spatial data acquisition because it can directly measure 3D coordinates of objects, though the resolution is so limited at this moment. With this background, the authors believe that it is very advantageous to integrate laser scanner data and stereo CCD images for more efficient and automated acquisition of the 3D spatial data with higher resolution. Laser scanner can measure coordinate values of an object directly, but it is rather difficult to recognize the object with only laser data, for its low resolution. On the other hand, CCD image represent detailed texture information on the object's surface and has potential of detailed 3D data generation as long as careful manual operation is applied. In this research, by using laser scanner data as an initial 3D data to determine the search range and to detect possibility of occlusions, a method to enable more reliable and detailed stereo matching method for CCD images is examined.

1. INTRODUCTION

1.1 General introduction

Recently, in GIS field, vigorous demand for detailed simulations and analyses, especially, in urban areas are observed. Examples are a propagation analysis of electric wave for wireless communication, a flood analysis, an analysis of wind caused with high-rise building and landscape simulation. 3D spatial information faithful to the real world is needed for such simulations. Photogrammetry is a current method of the data acquisition. However, as a matter of fact, a large manpower and expenditure for making detailed 3D spatial information is required especially in urban areas where various buildings exist. There are no photogrammetric systems which can automate a process of spatial information acquisition completely. On the other hand, laser range finder has high potential of automating 3D spatial data acquisition because it can directly measure 3D coordinate of objects, though the resolution is so limited at this moment.

With this background, the authors believe that it is very advantageous to integrate laser scanner data and stereo CCD images for more efficient and automated acquisition of the 3D spatial data with higher resolution. Laser scanner can measure coordinate values of an object directly, but it is rather difficult to recognize the object with only laser data, due to its low resolution. On the other hand, CCD image represent detailed texture information on the object's surface and has potential of detailed 3D data generation as long as careful manual operation is applied. In this research, by using laser scanner data as an initial 3D data to determine the search range and to detect possibility of occlusions, a method to enable more reliable and detailed stereo matching method for CCD images is examined.

At the first step of a data processing flow, straight edges extraction and region segmentation have been done with the TLS image, and they are individualized. Next, removing non-building areas have been done for reliable extracting buildings. Then, 3D coordinates of features like edges and regions are computed. As a consequence, identifying individual buildings is done. Thus, finally, 3D building models are generated.

1.2 Three Line Scanner TLS

Three Line Scanner (TLS) is an optical sensor for aerial survey. TLS is composed of three linear CCD arranged in parallel, and it can acquire three images of each direction (forward, nadir and backward) at the same time. Orienting it on an aircraft perpendicularly to flight direction, and scanning a ground plane, a treble stereo image of a ground object can be acquired (See Figure 1). As a result, occlusion area can be extremely reduced. Using two images of the three, it is also possible to get 3D coordinates by stereo matching.

As one of advantages of a linear CCD sensor, more pixels can be arranged in a single scene compared with an area CCD sensor. This means that a linear CCD sensor can achieve a resolution comparable with that o f an air photo. Though a linear CCD sensor can acquire data only by one line at a time (The ground resolution of TLS data in this research is 3 cm approximately). However, time of acquiring each line image is different. Since position and direction of each line when acquiring image is also different, orientation cannot be done by an existing method of photogrammetry. Moreover, the image is greatly influenced by fluctuation of an airplane position and attitude because TLS is air-borne. But setting up a stabilizer between an airplane and TLS, the fluctuation's influences can be reduced.

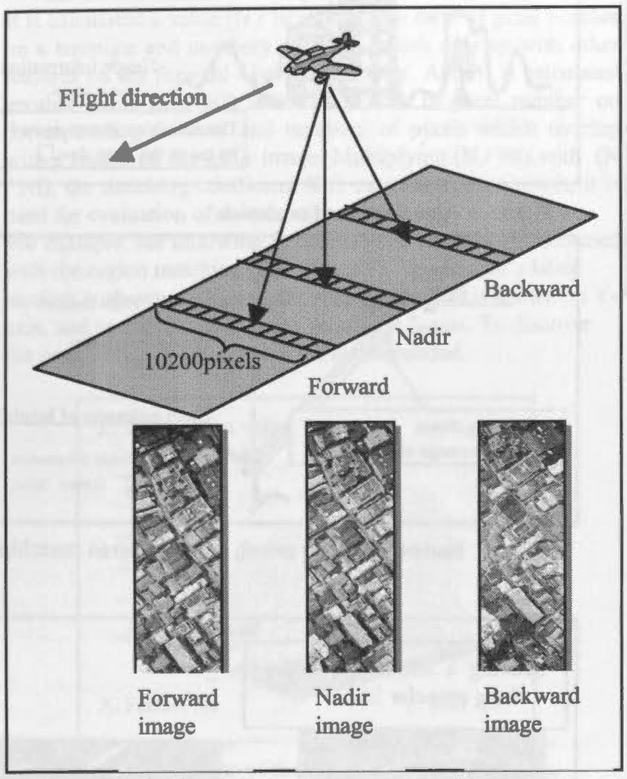

Flight direction
Backward
Nadir
10200pixels
Forward
Forward image
Nadir image
Backward image

Figure 1. Method of TLS data acquisition

1.3 Laser Scanner Data

Laser scanner is a sensor, which irradiates laser to an object and measure a distance to the object by measuring returning-time as the laser reflects. If the position of sensor and the angle of irradiating laser beam are known, 3D coordinates of a place, where laser hits, can be acquired easily. And, an automation of measurement with laser scanner is simply processed too, because 3D coordinates of a measurement point can be acquired directly.

However, it is difficult to know what the object is, because laser data provides only information about object's coordinates and reflection strength. And, in general, by the restriction of laser output, a ground resolution of airborne laser is 2 m. It is not high resolution to make use of the laser data as 3D city data. To extract a building in a city area from laser data, auxiliary data is required.

Laser scanner data using in this study is acquired synchronizing with TLS. The resolution is 2 m approximately. The laser scanner data is depicted (Figure 2).

2. METHODOLOGY OF INTEGRATING LASER RANGE IMAGES AND TLS IMAGE

2.1 The outline of our approach

The outline of our approach for data fusion has following steps.

☐ Remove non-building areas,

☐ Compute 3D coordinates of features like edges and regions,

☐ Identify individual buildings.

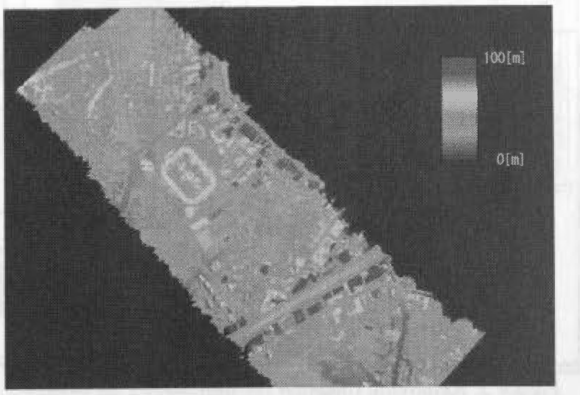


Figure 2. Data of laser scanner

Finally, 3D model can be generated. In this procedure, laser data is used for constructing rough 3D data as an initial data, and TLS images are used for refining the rough 3D data with very high-resolution images. In a sense, TLS images are used as a main data and laser data is used as an auxiliary data (See, Figure 3).

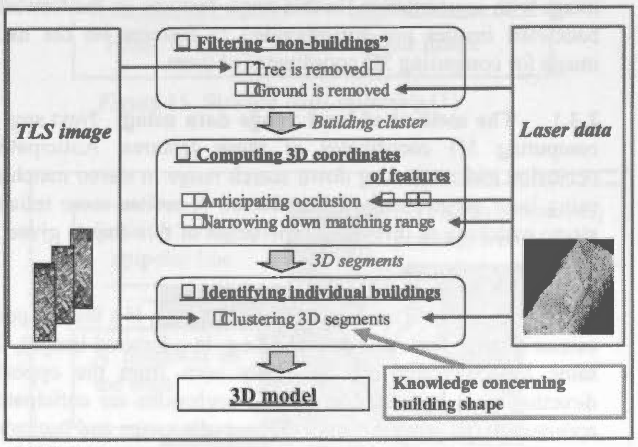

☐ Filtering "non-buildings"
☐Tree is removed☐
☐Ground is removed
Building cluster
TLS image
Laser data
☐ Computing 3D coordinates of features
☐Anticipating occlusion
☐Narrowing down matching range
3D segments
☐ Identifying individual buildings
☐Clustering 3D segments
3D model
Knowledge concerning building shape

Figure 3. The outline of our approach

2.2 Filtering "non-buildings"

At the first step of the data processing flow, non-building areas such as trees or roads are removed to extract buildings. To remove tree, at first, edges are extracted from TLS image. Furthermore, straight lines are extracted from the edge image. By removing very short straight lines and curved edges, trees are remove (See, Figure 4).

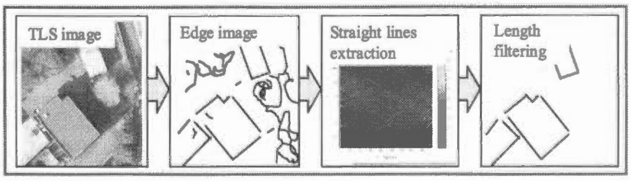

TLS image
Edge image
Straight lines extraction
Length filtering

Figure 4. Removing trees

Additionally, relatively flat ground such as roads is removed approximately using the laser range image. At first, using height information of laser data, building area and ground area are divided. In the next stage, a mask image is generated with this ground area. After masking TLS image with this mask image, region segmentation is applied (See, Figure 5).

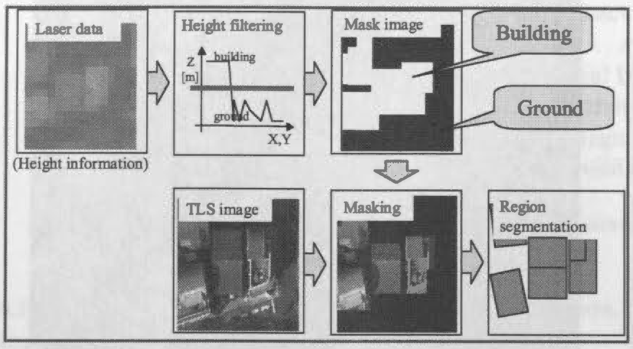

Laser data
(Height information)
Height filtering
Z
[m]
building
ground
X,Y
Mask image
Building
Ground
TLS image
Masking
Region segmentation

Figure 5. Removing roads

As a result, building region segments can be extracted. By doing these two-stage operations, almost non-building areas can be removed from TLS image, that means, features which may belong to buildings are extracted.

2.3 Computing 3D coordinates of features

Previously, regions and straight edges are extracted from TLS image with segmentation. In this stage, features on the forward / backward images are corresponded to features on the nadir image for computing 3D coordinates of them.

2.3.1 The method of laser range data using: Next step is computing 3D coordinates of those features. Anticipating occlusion and narrowing down search range in stereo matching using laser range images is conducted to realize more reliable stereo matching in this stage. The detail of this step is given as follows.

1) Anticipation of occlusion: In TLS image, if a feature point cannot be seen from one direction (e.g. in a forward image), the same feature point may be likely seen from the opposite direction (in a backward image). If occlusions are anticipated, appropriate image pairs (in this case, nadir image and backward image) can be selected. For anticipating the occlusion, the laser data are used as the height information (See, Figure 6).

2) Narrowing down search range in stereo matching: In stereo matching, the range of the search for feature point is relatively very wide, because there is no information on approximate height of features such as buildings. But, by using laser data, there are good estimates of height, which allow us to narrow down the range of the search (See, Figure 7).

With these methods, and it is possible to decrease mismatching and processing-time drastically.

2.3.2 Region matching: In this stage, region matching have been done for extracting outlines of building objects. Selecting a region on the nadir image, the region is cut out as a template with region information. Projecting the template on the forward / backward image, a corresponding region is searched on the epipolar line (See, Figure 8).

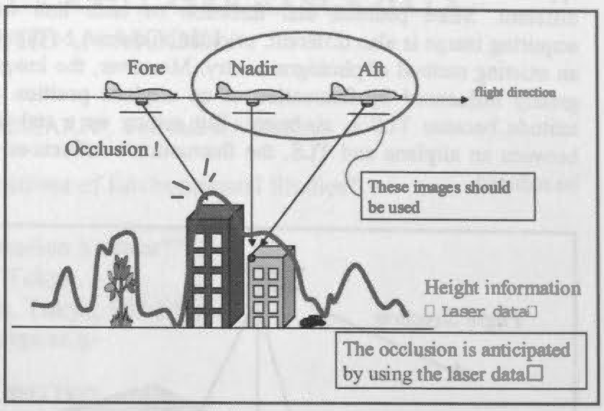

Fore
Nadir
Aft
flight direction
Occlusion !
These images should be used
Height information
□ Laser data□
The occlusion is anticipated by using the laser data□

Figure 6. Anticipating of occlusion

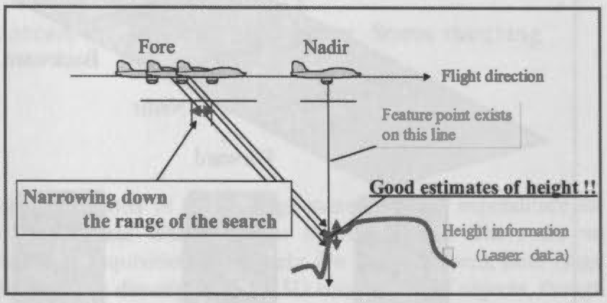

Fore
Nadir
Flight direction
Feature point exists on this line
Narrowing down the range of the search
Good estimates of height !!
Height information
□ (Laser data)

Figure 7. Narrowing down search range in stereo matching

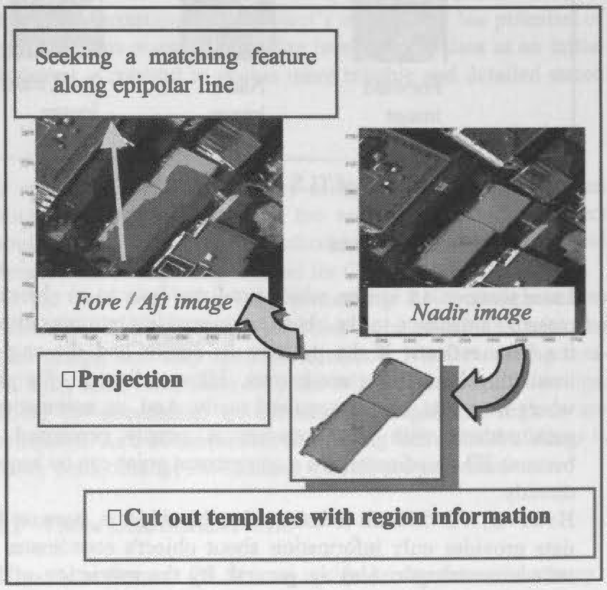

Seeking a matching feature along epipolar line
Fore / Aft image
Nadir image
□Projection
□Cut out templates with region information

Figure 8. Outline of the region matching (1)

When region matching is done, it is necessary to define corresponding features. In this study, it is defined as follows.

Definition 1: Shapes and sizes of features are resembled.
Definition 2: Colors of features is resembled.

In this stage, regions are corresponded each other with two steps.
At first, region matching, based on feature's shapes and sizes, is applied as the first step. And, obtaining a matching coefficient, corresponded features are selected roughly. The matching

coefficient is a reference value that shows a corresponding level. In this study, a matching coefficient is calculated as follows (1).

$$R = (N / Nt)\,(N / Nf) \qquad (1)$$

where N = pixel number which is overlapped
Nt = pixel number of a template
Nf = pixel number of a corresponding feature

It is calculated a value (N / Nt) which is a ratio of pixel number on a template and numbers of pixels which overlap with other features on the forward / backward image. And, it is calculated another value (N / Nf) which is a ratio of pixel number on corresponding features and numbers of pixels which overlap with a region on the nadir image. Multiplying (N / Nt) with (N / Nf), the matching coefficient R is calculated. As a result, it is used for evaluation of the corresponding level.
For example, the following figure shows a result that is obtained with the region matching (See, figure 9). The feature's label number is shown on X-axis, the seeking number is shown on Y-axis, and correlation values are shown on Z-axis. To discover the peak in this result, regions are corresponded.

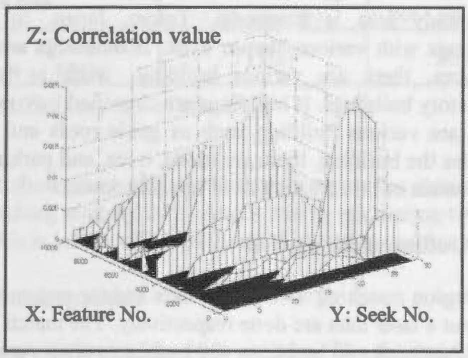


Figure 9. Outline of the region matching (2)

As the next step, region matching, based on feature's colors, is applied. In this study, the area correlation method is applied. At this time, using the result that is obtained in the first step as an initial value for the searching a matching point, the searching range can be narrowed. As a consequence, corresponded features are selected exactly.
Finally, 3D region is generated with stereo matching (See, Figure 10). The height information obtained here becomes an initial value in the next stage (edge matching).
Moreover, using the laser data, rough corresponding features can be done instead of the first step, because the laser data can be used as an initial height value for the region matching. Therefore, substantially, region matching becomes one step only of the use of the area correlation method as the region matching.

2.3.3 Straight edge matching: In this stage, straight edge matching is done for extracting details of building objects. To select a straight edge around a region in the nadir image, it is projected in the forward / backward image. Also, a corresponding straight edge is searched on the epipolar line (See, Figure 11). At this time, the method in the first step of the region matching is applied.

Moreover, edges are corresponded with similar informations of edges as a clue for the matching in this stage. The clues are as follows, a direction of the edge, a length of the edge, the average density on both sides of edge, a relation concerning position with region, and so on (See, Figure 12).

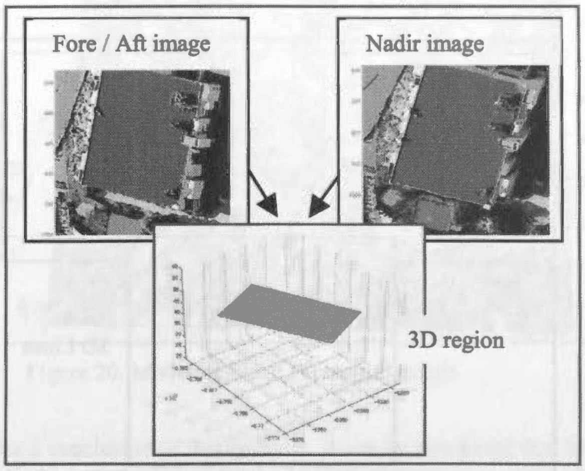


Figure 10. Generating 3D region

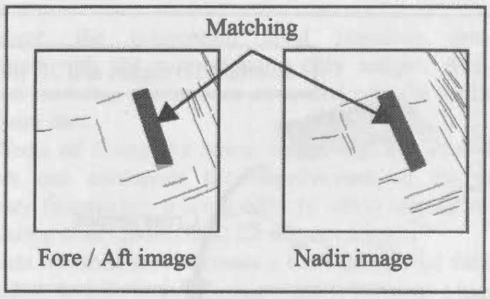


Figure 11. Straight edge matching (1)

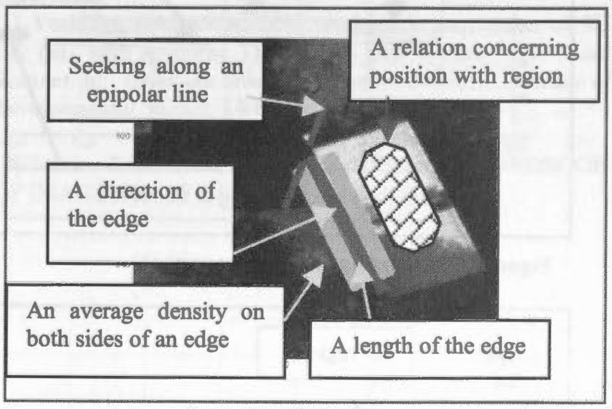

Seeking along an epipolar line
A relation concerning position with region
A direction of the edge
An average density on both sides of an edge
A length of the edge

Figure 12. Straight edge matching (2)

Finally, calculating 3D coordinates with forward intersection, 3D lines are generated with stereo matching (See, Figure 13).

2.4 Identifying individual building

Generating 3D region and 3D lines with the region matching and straight edge matching, they are overlaid (See, Figure14).

Figure 14 shows a cross section of the segments. Red arrow is indicating a height level that many 3D lines gather. At the next step, the 3D region is updated with 3D lines, because generated

3D lines with edge matching have more correct values than 3D regions which are generated with region matching (See, Figure15).

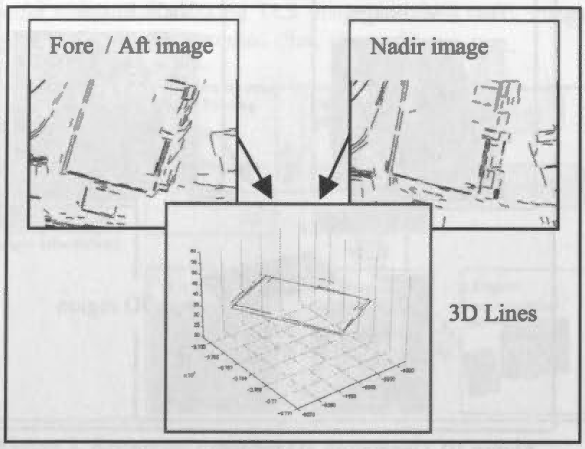


Figure 13. Generating 3D lines

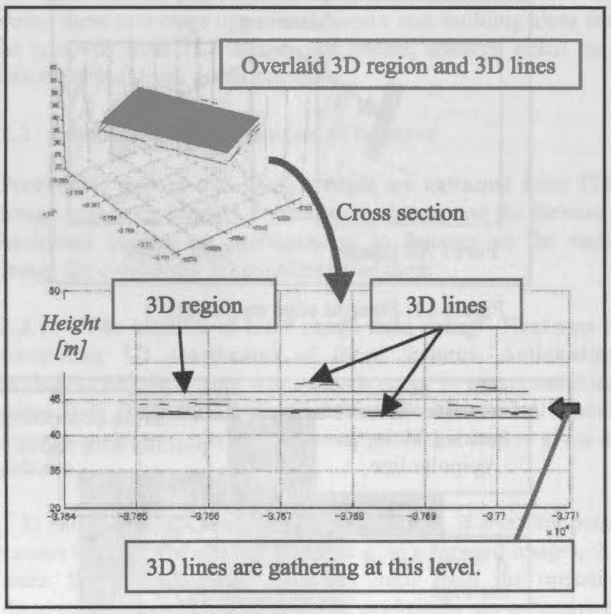


Figure 14. Generating 3D building parts (1)

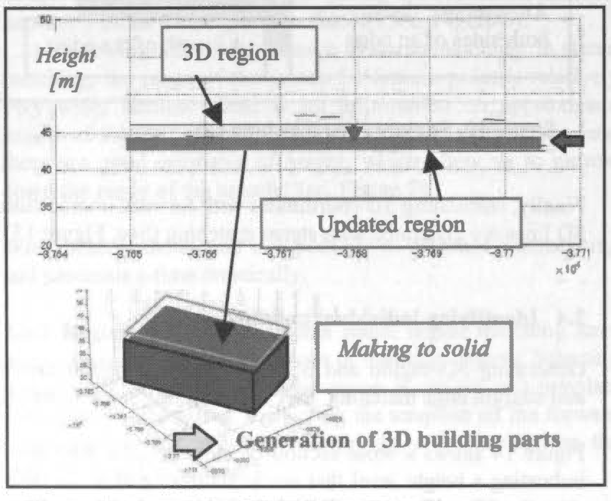


Figure 15. Generating 3D building parts (2)

Then, a solid model is generated with lofting the 3D region. This model is called 'building parts'.
However, a real building cannot be expressed enough only by this model. Then, a building object is generated with merging adjoining 3D building parts (See, Figure16). In this way, 3D segments are clustered.

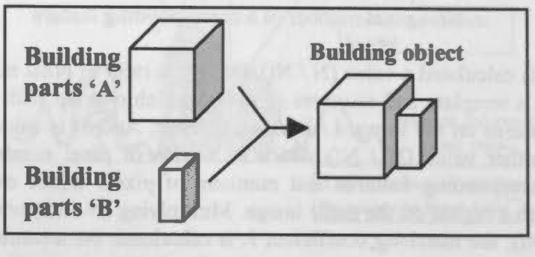


Figure 16. Merging 3D building parts

3. EXPERIMENT

3.1 Study area

The study area is Roppongi, Tokyo, Japan. In this area, buildings with various shapes exist. If buildings are classified by sizes, there are various building such as houses and multistory buildings. If buildings are classified by roof's shapes, there are various building such as gable roofs and flat roofs. Besides the building, there are roads, trees, and parking lots, etc. The terrain of this area is full of ups and downs.

3.2 Outline of the method

The region matching with a laser data and the region matching without a laser data are done respectively. The match accuracy is calculated for 40 buildings (86 building parts). As a result, how using the laser data is effective is considered.

4. RESULT

4.1 Matching result

The matching result is evaluated by measuring the gap of the direction of the line (vertical direction in the image) of regions (See, Figure17). And, the result (region matching) is shown Figure18. Blue points show matching results without laser data, and red points show matching results with laser data. Figure 18 shows that fusing TLS images and laser data makes result better.

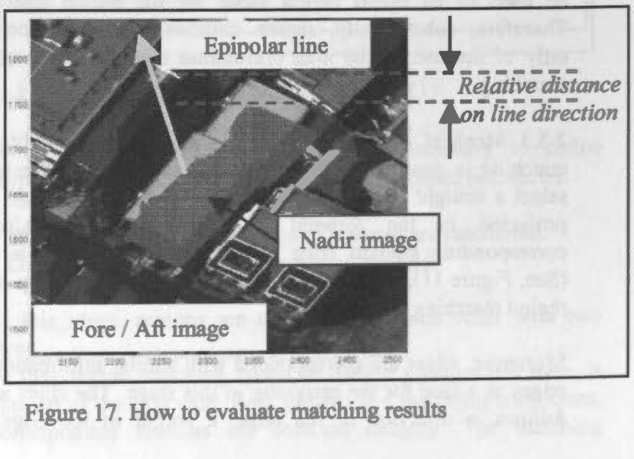


Figure 17. How to evaluate matching results

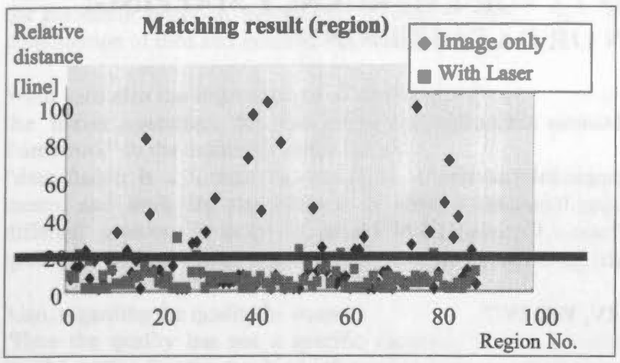


Figure 18. Matching results (Region matching)

By the way, defining 20 line as a threshold of success / failure, the results can be given in the following table (See, Table 1).

	Image only	With Laser data
Success rate	68% (59 / 86 objects)	98% (84 / 86 objects)
CPU time (86 objects) (pentium4 1.8GHz)	711[s]	363[s]

Table 1. Success rate of matching

Table 1 shows that using the laser data makes the success rate in the matching almost 100%. And it makes calculating-time about half. This is just an effect of fusing TLS image and laser data.

4.2 Modeling result

Modeling results are shown as follows. The figure 19 shows a solid model, and the objects are classified by buildings.

Moreover, using TLS nadir image, a textured 3D model is shown figure 20.

5. SUMMARY

In this research, combining TLS images with laser scanner data, a more efficient method of making 3D spatial information with higher resolutions is examined.

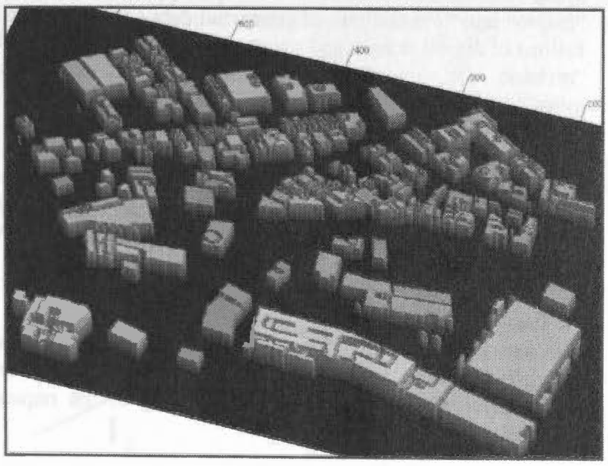


Figure 19. Modeling result (Solid model)

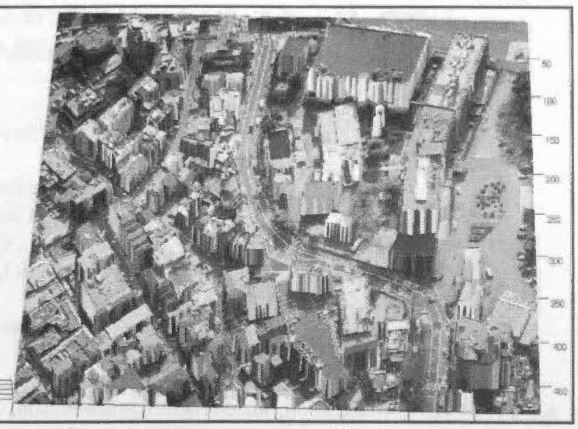


Figure 20. Modeling result (Textured model)

As a conclusion of this research, it can be concluded that basic shapes of buildings can be possible to generate full-automatically by fusing TLS images and the laser data. However, region segmentation has to be done under a good condition.

Moreover, the automation level improves dramatically compared with the method using only images. And a more detailed modeling can be done compared with the method using only laser data.

As effects of fusing the stereo image and the laser data, the authors can enumerate the improvement of the matching accuracy (improving it from 68% to 98%) and shortening of calculating-time (about 50%) for the matching.

The data obtained here becomes a basic 3D spatial data for city modelling, and the authors believe that it becomes a big support to digital mapping.

References

[1] Yoshiaki, KAGAWA, 2001. Automatic acquisition of 3D city data with air-borne TLS (Three Line Scanner) and Laser Scanner. In: *Graduate School of Frontier Sciences, Institute of Environmental Studies, University of Tokyo*, Japan.

[2] Mikio, TAKAGI, & Haruhisa, SHIMODA. HANDBOOK OF IMAGE ANALYSIS.

MERGING OF HETEROGENEOUS DATA FOR EMERGENCY MAPPING: DATA INTEGRATION OR DATA FUSION?

Florin Savopol and Costas Armenakis

Center for Topographic Information
Geomatics Canada, Natural Resources Canada
615 Booth Str., Ottawa, Ontario, Canada K1A 0E9
{fsavopol, armenaki}@NRCan.gc.ca

Commission IV, WG IV/7

KEY WORDS: Integration, fusion, heterogeneous data, emergency mapping

ABSTRACT:

Many terms are used to name and define these data operations: "fusion" and "integration" of geospatial data or "integration (or fusion) of digital images and geospatial information", as well as "revision (or updating) of geospatial (or topographic) information (or data bases). The present paper will try first to delimitate the use of these terms in the context of the research work done for the CIT-O (Centre for Topographic information – Ottawa, Natural Resources Canada).

In an emergency situation the authorities in charge of mapping support will face two major challenges:

1) to deliver 'immediately' up-to-date existing topographical information showing the situation before the emergency occurs (position of existing roads, bridges, community facilities, strategic buildings, etc.);

2) to get as quick as possible digital images from the disaster area in order to understand and monitor the situation, to evaluate the damages and the risk for injuries or more damages and to support the rescue operations.

To meet these challenges there is a need to deal with a range of heterogeneous geodata consisting for example of various sources, geometries, scales, resolutions, types, accuracies and dates. In an emergency mapping situation, the choice of data sources to be integrated / fused could be limited and the user can be forced to use data and images with a resolution outside the normal limits. The present work evaluates the fusion of images with a significant difference in spatial resolution in the typical framework of an emergency mapping project. It also investigates the fusion possibilities of the various data with respect to their enhancement of feature interpretation and extraction as well as the integration of imagery with existing topographic data. Relations and criteria are established for the evaluation of the fusion processes, while certain relations can be established between the resolution of the imagery and the scale of the original cartographic product.

1. INTRODUCTION

In an emergency situation resulting from a major natural disaster or from another event (a major ecological or industrial accident, an act of war, etc) the authorities in charge of mapping support will face two major challenges:

1) to deliver 'immediately' up-to-date existing topographical information on both digital and classical (paper) support showing the situation before the emergency occurs (position of existing roads, bridges, community facilities, strategic buildings, etc.);

2) to get as quick as possible digital (eventually metric) images from the disaster area in order to understand and monitor the situation, to evaluate the damages and the risk for injuries or more damages and to support the rescue operations.

To meet these challenges there is a need to deal with a range of heterogeneous geodata consisting for example of various sources, geometries, scales, resolutions, types, accuracies and dates.

Starting from the experience accumulated during the effort for updating the National Topographic Database of Canada (NTDB) using satellite imagery, the present research investigates the possibilities integration and fusion of different kind of digital imagery with exiting topographic data in the context of emergency mapping.

The availability of cartographic data in digital form had encouraged users to manipulate, to merge and combine geospatial data from different origins, scales and content in order to solve theirs needs.

2. DEFINITIONS

Many terms are used to name and define these data operations: "fusion" and "integration" of geospatial data or "integration (or fusion) of digital images and geospatial information", as well as "revision (or updating) of geospatial (or topographic) information (or data bases).

Some definitions are:

On images fusion:

.."image fusion is the combination of two or more different images to form a new image by using a certain algorithm" Van Genderen, Pohl (1994).

On fusion of information in general:

The data fusion is ..."set of methods, tools, and means using data coming from various sources of different nature, in order to increase the quality (in a broad sense) of the requested information" Mangolini (1994).

US DoD (Wald, 1998) uses more general definition without reference to the quality:

"data fusion is a multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation and combination of data and information from multiple sources"

Wald uses also the improvement of quality al primary scope of the fusion operation. He has introduced the "conceptual framework" in the definition of the fusion:
"data fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources, in order to obtain information of greater quality", Wald (1998).

Also, regarding the quality he states:
"Here the quality has not a specific meaning. It is a generic word denoting that the resulting information is more satisfactory for the <customers> than before performing the fusion process. For example, a better quality may be an increase in accuracy, or in the production of more relevant data." Wald (1998).

Some authors state that data fusion is the result of input from at least two different sensors. Some accept also the result of two different data sets, event if these are generated from the same sensor.

In general, integration / fusion aim to achieve the following:

1.- To enhance and improve the overall quality and reliability of the final output to support better geospatial operations.

2.- To enhance the information content of the final output to support better decision-making operation.

3.- To augment insufficient data and information to support better geospatial representation and to achieve better solutions.

To satisfy the requirements of the above objectives we see that the operations to bring diverse, distinct or separate elements into a unified whole can be applied to either methods or to data or to tools or to any combination of these. The blending mode of operations uses terms such as integration, merging, combination and fusion, which often are interchangeable in practice. We will try to make an effort to clarify these terms in the realm of geomatics and put them into some perspective.

Starting from the input space with at least two different inputs we mathematically map them to the output space. Depending on the mapping operation and mapping output we propose the following use of terms:

If the mapping operations or the output of them result in a separable type of output, where the individual characteristics of each of the input elements are preserved then the terms "combination" or "integration" may be the more appropriate ones (Fig. 1).

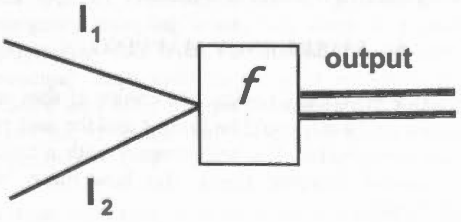


Figure 1. Combination / integration

If the mapping operations or the output of them result in a unified blended type of output, where the individual characteristics of each of the input elements are not preserved, then the terms "merging" and "fusion "may be the more appropriate ones (Fig. 2).

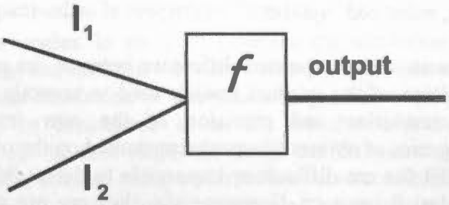


Figure 2. Merging / fusion

3. INTEGRATION OR FUSION?

The increased use of Geospatial Information Systems (GIS) for new thematic mapping application is pushing the need for quality, up-to-date data ready for end-user consumption. Historically, in mapping applications we have been working with heterogeneous data. Map compilation from aerial photographs is a combination of using control points determined by geodetic means and the analogue recorded images on film. In this case we see that the image data is augmented by the introduction of the control points to provide the spatial reference. Today we see the use of GPS and INS measurements to provide the information needed for the exterior orientation. This is an example of data integration, which provides the positional and orientation elements for each image. For cartographic purposes we see examples of using one set of updated features as a reference to correct the geometry of the rest of the features.

If we look now at the operation of image classification, the input sources are the various multispectral channels and the output is a layer of thematic classes. This merging of input sources to an entire new entity is considered as data fusion.

Another example of data fusion is the pan sharpening operation, which takes advantage of both the high resolution panchromatic image and lower resolution of the multispectral image. The result of this operation is an entire new enhanced multispectral or synthetic image and is considered as a typical data fusion process.

The National Topographic Data Base (NTDB) was created by Natural Resources Canada from the National Topographic Series (NTS) maps at 1:50 000 and 1:250 000 scales. The 1:50000 can be considered as the general, base scale in Canada. The original reprographic material was scanned, vectorized and the resulting data was structured in order to produce the NTDB. An important number of these NTS maps were already aged at the moment of conversion and the need of updating the NTDB became soon evident. Different methods were used in the past decade for the updating of NTS maps and NTDB files in the frame of research projects, experimental production and partially in normal cartographic production. Many data source were used as aerial photography, orthophotos and ortho-images from SPOT, IRS and Landsat 7 imagery. Let us examine the case of one of the currents method in use: Landsat 7 images are converted into ortho-images using DEM derived from existing NTDB files and control points extracted from the original aerotriangulation. A fusion operation is performed using the panchromatic Landsat 7 image (15 m pixels resolution) and the multispectral channels of Landsat 7 (30 m pixels resolution). The resulting multispectral (synthetic) image is used partially for classifications (water, forest) and partially for interactive change detection and updating process perform by superimposing the old (NTDB) vector information over the new, enhanced "synthetic" multispectral ortho-image.

Because of the important difference between the resolution and precision of the original images used to compile the map and the resolution and precision of the new imagery, some categories of objects (themes) represented on the old map and in NTDB file are difficult or impossible to detect /identify on the Landsat 7 imagery. Consequently, they are not subject of the updating process.

The resulting, updated NTDB file and the derived printed map have an enhanced quality as a result of the updating process. The updated NTDB file and the derived printed map offer a better representation of the terrain reality than the old one. Some of the represented objects have been extracted from the new Landsat 7 imagery and others from the original aerial photography. Is the resulting (updated) NTDB file the result of a fusion process between the old digital topographic data (or data base) and the new Landsat7 imagery?

As a new improved cartographic product this can been seen as the result of fusion operation. On the other hand, from the database point of view the characteristics of the elements are preserved, therefore it can be seen as an integration process.

In other cases only updated layers are replaced in a database, leaving the rest of the information as it was. This is the case when only the roads from a NTDB file are replaced ("updated") by the data obtained from the Road Network data (RN). This DB is obtained by direct measurements using GPS methods. This is clearly a case of data integration.

One of the most popular approaches or products is the superimposition of vector data over image data (Figure 3).

As both inputs are put together by the process of "addition" without any change in their properties, this is consider the result of data combination or integration and not as data fusion.

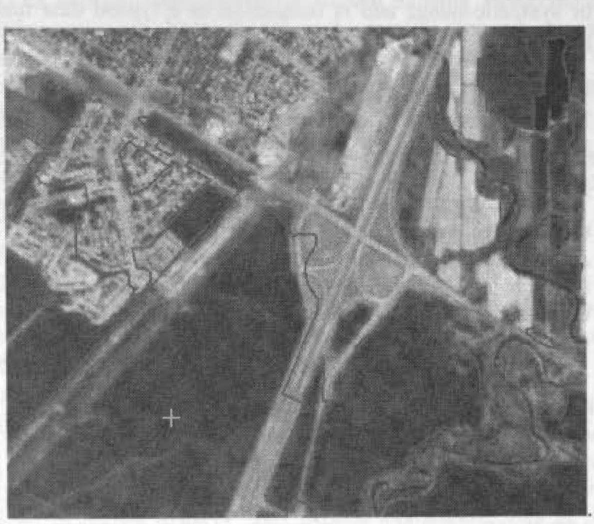


Figure 3. Example of vector / image data integration by superimposition

4. CONSIDERATIONS

When integration or fusion occurs at the data level a number of factors need to be considered. These aspects include the type of datasets (vector-vector, vector-raster, raster-raster), scale and resolution of datasets, co-registration and the temporal decorrelation.

Merging vector data of different scales both the database (e.g., content, attributes) and the cartographic generalization aspects must be considered. If the scale of the output is the same with the scale of one of the input datasets, then we may have a case of data assimilation. If not, then the output is the result of data merging. In both cases the result of this data conflation is a new fused product.

In vector-raster case we usually deal with the superimposition of old vector datasets over newly acquired ortho-imagery. The relationship (ratio) between the scale of the vector data versus the image resolution needs to be considered. For example, what is the impact in spatial sense of superimposing small scale vector data over high resolution image data.

In the raster-raster cases the ratio of pixel size between the high resolution panchromatic image and the low resolution multispectral image is important. It may result in having erroneous spectral data in the resulting high resolution synthetic image, leading to classification errors.

In every of the above case, proper co-registration of the two data sets is a prerequisite. Examples of fusion between a high resolution panchromatic image and a low resolution multispectral image are shown in figure 4 and figure 5.

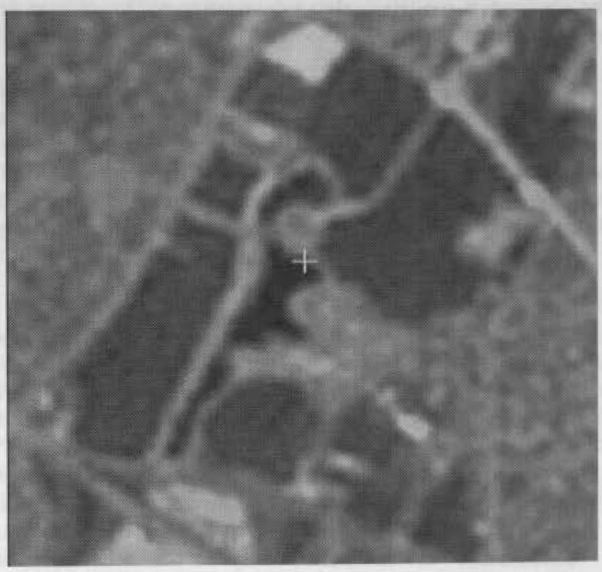


Figure 4. Original IRS multispectral image

5. EMERGENCY MAPPING

In an emergency-mapping situation, the choice of data sources to be integrated (or fused) could be limited and the user may be forced to use cartographic data and imagery with a resolution outside the normal accepted limits. But how these 'normal limits' are defined?

In this case, it is useful to distinguish between cartographic data as compiled, edited and symbolize data and imagery.

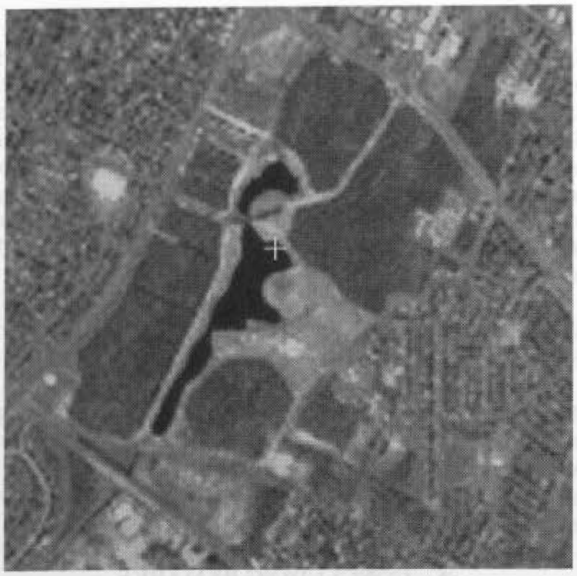


Figure 5: Fused image

In the first case, the 'normal' limit in resolution (or scale) difference could be considered the moment when the magnitude changes produced by the cartographic editing process and the symbolization of one of the data sets reaches a critical level at the scale of the data set with the higher resolution. If the mismatching is significant, this integration operation is to heavy, impossible ore will produce distorted, unacceptable results. Figure 6 shows the case of the integration of a more recent 1:50 000 scale NTDB file with an old 1:250 000 scale NTDB file. The discrepancies produced by the cartographic processes are evident.

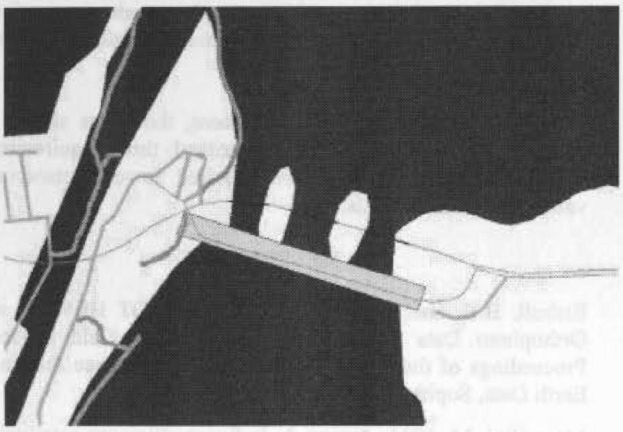


Figure 6: Example of problems produced by scale differences between two vectors maps (courtesy Jean Brodeur, NRCan).

This type of problems could be present in a situation of emergency mapping when, i.e., there is a need to integrate topographic data from the federal government with local / municipal large-scale vector data in order to offer recent, detailed and more complete information to the user. In both situations, it seems that we cannot talk about a data fusion situation.

In other situations, it is required to integrate a high-resolution digital image (raster image data) with an existing, eventually aged map (vector data). Typically, the high-resolution image will offer recent, up-to-date general topographic information or, in the case of an emergency situation, it will show the results of an accident, a natural disaster or an act of war. This will help to understand the situation, to evaluate the damages, to detect the potential hazards and to plan rescue and protection operations.

We can distinguish between several situations:

I - Similar scales. In an ideal situation, the resolution of the digital image should be in the same order of magnitude with the accuracy and resolution of the vector map (data consistency). These can be considered as a rule for the limit in scale difference. At the first stage. the image can be oriented using control points selected using a correlation between details points easy to identify on the vector map as road intersections and the corresponding point on the image. By superimposing the vectors over the image, the vectors representing each object should coincide with the image of that corresponding points in the image. More images could be oriented and assembled together as a mosaic. Is this 'mosaic' the result of a fusion or a merging process?

A more elaborated photogrammetric product, as a rigorous ortho-photo could be produced under normal circumstances. This high resolution, new ortho-photo could be augmented be superimposing selected features from one or from several existing (older) vector maps. We have in this case an <image-map> type product. In some situations, it could be important to have more information about a extended surrounding area. If the existing vector maps are not updated, it should be necessary to update these maps (what is described in the first section of this paper as a 'type 1 tasks' for emergency mappers). The specific emergency situation and the critical time limitations will not permit to use a normal cartographic updating process. A solution is to use any available recent imagery, as medium or high resolution satellite images. The conditions are that these images are significant more recent than the existing vector maps and their resolution meets the specific requirements. Finally, we my have a complex cartographic product that includes: 3-D data visualisation, a multi-source image-map with images from two or more sources with significant differences in spatial and spectral resolution, and one or more selected vector maps. From the point of view of the user, it seems to be a fused product. Figure 7 shows an example of fusion for 3-D visualisation using Landsat 7 and IRS (panchromatic) imagery, CanMatrix (scanned NTS maps) 250K CDED data (DEM derived from NTS 1:250 000 scale maps).

II- Large scale difference. It is possible that the new digital image has a much higher resolution than the old vector map. This situation is favourable for the point of view of the team in charge with the emergency situation, such images offering an increased potential for objects identification and helping for other tasks (situation understanding, response planning, etc.). From the point of view of the cartographic team, some additional problems will emerge. The volume of image data will increase. The spatial representation of features will be different.

It will be more difficult to correlate control points between the vector maps and the images. If the higher resolution images are controlled using the low resolution map, there will be probably some misalignments between the digital images. This situation is expected and we can accept that these misalignments are inside the error "budget" of the vectors map. Assuming that there are no problems at the cartographic information level, this is more a question of "visual comfort" and aesthetic aspect.

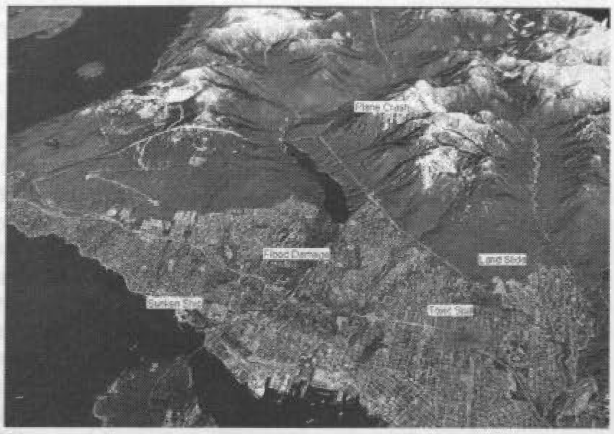


Figure 7. Example of fusion for 3-D visualisation (courtesy John A. Ells, Natural Resources Canada).

We can further distinguish between 2 situations regarding the image sensor and the working environment:

II-a : If there is not enough information about the geometry of the sensor or no appropriate photogrammetric software is available, the result of mixing and geo-referencing the image date will have the precision of the cartographic product used for the control (as geo reference) or less due to relief displacements. That is much lower as one can be expect from images of this type. This is a case of data integration.

II-b : If there is enough information about the geometry of the sensor and modern photogrammetric tools are available, the initial precision / accuracy of the original vector map(or topographic DB) could be improved significantly. The control obtained from the vectors map (or topographic DB) will be used only for the absolute orientation of the new images but the internal, relative precision of the element visible on the images will have a higher precision, in the same order of magnitude than the resolution of the digital imagery.
From the users point of view in an emergency mapping environment, this is considered as a data fusion situation.

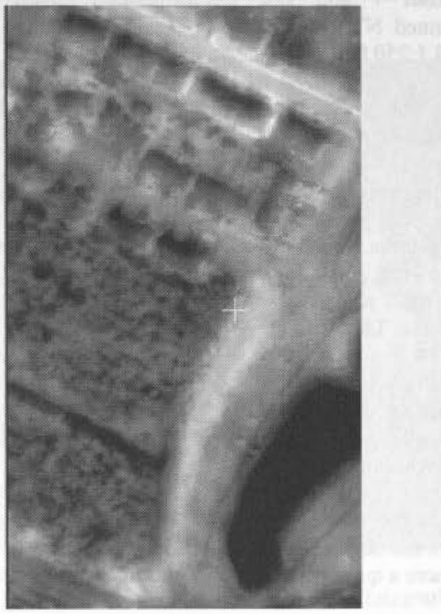


Figure 8: IKONOS (1m) with IRS LISS (23m) fusion

III - Image-to-image fusion. In cases where we deal with images of various resolutions and wavelength bands, the objective is to end up with output pixels consisting of the highest possible resolution and the richest multispectral information. Considering that usually emergency situations have a local coverage we may face the case to blend images with significant differences in resolution. For example, IKONOS with 1m resolution with multispectral IRS LISS 23m resolution (1:23 ratio factor) (Figure 8).

We can find examples, reporting ratio factor of up to 1:32 (Ersboll et al., 1998).

The large difference in resolution between the fused images creates a fuzzy effect. However, some additional thematic information can be extracted. For example the presence of vegetation (grass) can be seen on the right part of the image near the water body.

6. CONCLUDING REMARKS

In a complex, multi-source cartographic product it is difficult to establish very precise, absolute limits between fusion and integration type processes. If we consider the most accepted definitions (as Wald, 1998) and the improvement of quality from the aspect of a more relevant, more 'satisfactory' data for the user, most of the described products are considered as fusion. But on the other hand, from the database point of view, if the characteristics of the original elements are preserved, most of them can be seen as the result of an integration process.

Considering the differences in scale and resolution of the original geospatial data no precise limit can be imposed. It depends very much on the properties and type of the final product. It is seen as a general use mapping product as a printed map, more restrictions should be imposed in order to preserve the homogeneity of the product.

In an emergency mapping environment, the limits should be more flexible in order to meet critical time requirements. Further theoretical studies are required to understand and validate empirical results.

7. REFERENCES

Ersboll. B.K. and all, 1998: Fusion of SPOT HRV XS and Orthophoto Data Using a Markov Random Field Model - Proceedings of the Second International Conference Fusion of Earth Data, Sophia Antipolis, 28-30 January 1998.

Mangolini, M. 1994: Apport de la fusion d'images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation . – Thèse de Doctorat, Université de Nice – Sophia Antipolis.

Ranchin, T., 2001: Data Fusion in Remote Sensing: Examples. – Proceedings of 4th Annual Conference on Information Fusion, Montréal, 7 – 10 August 2001.

Van Genderen, Pohl, 1994: Image Fusion, issues, techniques, and applications. Proceedings of the EARSeL workshop on intelligent image fusion, Genderen J. van and Capppellini Eds.

Wald, L., 1998: Data Fusion: A Conceptual Approach for an efficient Exploitation of Remote Sensing Images - Proceedings of the Second International Conference Fusion of Earth Data, Sophia Antipolis, 28-30 January 1998.

WG IV / 8:

GLOBAL ENVIRONMENTAL DATABASES

TOWARDS A GLOBAL ELEVATION PRODUCT: COMBINATION OF MULTI-SOURCE DIGITAL ELEVATION MODELS

Achim Roth*, Walter Knöpfle*, Günter Strunz*, Manfred Lehner**, Peter Reinartz**

* DLR, German Remote Sensing Data Center,
** DLR, Remote Sensing Technology Institute
Oberpfaffenhofen, 82234 Wessling, Germany

KEY WORDS: Digital Elevation Model, SRTM, MOMS-2P, SAR-Interferometry, Stereo, Data Fusion

ABSTRACT

Mapping and monitoring the Earth's surface is subject to various application fields. This three-dimensional problem is usually split into a two-dimensional planar description of the Earth's surface supplemented with the height information provided as separate digital elevation models. However, from the global point of view there is still a tremendous need for suitable height information at a resolution level of about 1 to 3 arc seconds. The Shuttle Radar Topography Mission (SRTM) will help to fill this gap by providing high quality digital elevation model (DEM). In order to provide terrain information for areas where either SRTM data will not be available or the corresponding resolution is not sufficient a combination with other sources will be required.

The German Remote Sensing Data Center (DFD) of DLR intends to implement such a global DEM service. All SRTM/X-SAR data will be processed to elevation data and will serve as the backbone as it provides a global net of homogenous elevation information. This net can be used for the absolute orientation of other DEMs as geometric reference, but also for the improvement of the height quality by integrating elevation data from a variety of other sources by DEM fusion and mosaicking techniques.

The paper describes the principles and corresponding accuracy of space borne missions for the derivation of DEMs. The main focus is on the DEM products of SRTM/X-SAR. Furthermore, the ERS-1/2 tandem configuration and the MOMS-2P mission are described. The technique to combine multi-source DEMs is outlined, which is based on the concept of height error maps. The method is illustrated by practical examples. Finally, an outlook is given on further investigations.

1 INTRODUCTION

Within the last years a growing need for high resolution elevation data with global coverage became clearly evident. Although in the meantime global elevation data, e.g. the 1 km resolution GLOBE data set (Hastings & Dunbar, 1998) became available, high quality elevation data at a resolution level of about 1 to 3 arc seconds are still lacking in many regions of the Earth and therefore are of high interest for a large number of applications worldwide.

However, several dedicated missions, that have been launched in the last years can contribute to a significant improvement of this situation in the near future. This is mainly due to the data acquired during the Shuttle Radar Topography Mission (SRTM). SRTM mapped the land surface within 56° southern and 60° northern latitudes covering 119 million km^2 (Pessagno, 2000) with NASA's C-band and approximately 58 million km^2 with DLR's X-band system. The data of both systems are currently processed using SAR interferometry.

The European Space Agency ESA built up an excellent global data basis operating it's ERS-1 and ERS-2 satellites in the tandem mode for interferometric DEM production, even providing multiple coverage. Moreover, e.g. the RADARSAT system provides elevation data based on SAR interferometric processing as well as SAR stereo processing.

Satellite systems based on optical stereo sensors as the SPOT system acquire data in different viewing angles for stereo image processing. The German stereo camera MOMS-2P was mounted on the PRIRODA module of the Russian space station MIR. It delivered approximately 65 million km^2 of high quality and high resolution imagery (Schroeder et al, 2000).

Each of those techniques and systems is able to provide high quality elevation information. However they are also restricted to their inherent limitations. The imaging geometry, the look direction and the look angle as well as illumination conditions are important parameters. Repeat pass acquisitions are affected by changes on ground and in the atmosphere reducing the precision of the final DEM product. Simultaneous operating single pass systems like SRTM and MOMS-2P provide optimized imaging conditions. The relative position of the imaging apertures is directly observed. However geometric limitations still exist and a total coverage of the earth's surface could still not be achieved.

In order to further improve this situation a technique for DEM fusion and mosaicking is required allowing to overcome the limitations and to provide a global DEM.

2 DEM SOURCES

In principle, a variety of space borne missions and sensor systems exist, which provide valuable data sources for the generation of DEMs. The techniques, that are currently used, can be grouped according to the sensor technology into SAR (interferometry and stereo) and optical stereo methods.

SAR interferometry utilizes the phase of SAR signals to measure differences between the distances from two radar antennas to the ground in an accuracy of a fraction of the wavelength. The acquisitions are either performed simultaneously (single pass interferometry) or at two different dates (repeat pass interferometry). The shuttle mission SRTM provides data for single pass interferometry, the satellite systems ERS-1/2, Radarsat and ENVISAT-ASAR provide data suitable for repeat pass interferometry. A detailed description can be found e.g. in (Bamler & Hartl, 1999).

SAR stereo (or radargrammetry) exploits the intensity of the back-scattered signals. Within SAR images the terrain slope mainly influences the brightness. Similar to the optical case parallaxes are measured within the image pair. The capability of using multiple incidence angles is realized in Radarsat and ENVISAT-ASAR. (Leberl, 1990) and (Toutin, 1999) describe in more detail the SAR stereo processing.

Optical stereo techniques from space are in general based on line scanner images, which are acquired either at two different satellite passes (dual pass stereo) or in a single pass e.g. by a three-line scanner system (single pass or along track stereo). For the dual pass configuration it is necessary that the sensor can be tilted to cover the same area at two different acquisition dates providing a suitable base/height ratio for DEM generation. For instance, the panchromatic sensors of SPOT satellites as well as of the IRS-1C/1D enable such capabilities. The first along track stereo scanner in space was the German MOMS-02 camera, which was flown as technological experiment on the second German shuttle mission (MOMS-02/D2) as well as on the Priroda module of MIR space station (semi-operational MOMS-2P, see Kornus et al, 1999). Within the next years several satellite missions will be launched, which will provide along-track stereo capabilities. These systems will offer stereo data at high resolution, e.g. SPOT-5 (10 m resolution), IRS Cartosat-1 (2.5m) and ALOS PRISM (2.5 m).

The following chapters will describe in more detail the three missions SRTM, ERS-tandem and MOMS-2P, which were used as data source for the practical investigations described in chapter 4.

2.1 SRTM / X-SAR

SRTM mapped the Earth's surface between February 11th–22nd, 2000. Within only eleven days approximately 80% of the land mass was covered being the home of 95% of the world's population. Two antenna pairs were operated during the mission working in two different frequencies - NASA's C-band-system and the German-Italian X-SAR. Due to the 230 km swath-width of the C-band a full coverage will be achieved. The X-band swath is only 50 km wide leading to a coverage net showing the biggest distance of approximately 150 km at the equator. The distance gets smaller with increasing latitudes until the scenes overlap at about 55° north and south (see figure 1). A benefit of the X-SAR system is the high frequency, which leads to a high elevation sensitivity.

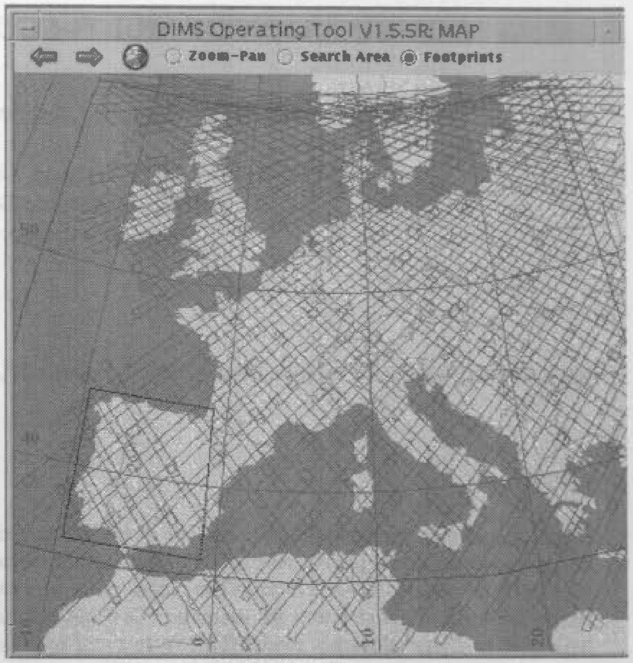


Figure 1. SRTM/X-SAR coverage Europe

SRTM provided the capability of single pass SAR interferometry for the first time in space. The active antennas sitting in the shuttle's cargo bay were supplemented by a second pair of antennas mounted on a boom that was deployed in space. Temporal de-correlation being the major limitation of repeat pass interferometry doesn't effect the DEM quality as the observations are performed simultaneously. Another benefit is the direct measurement of the baseline. Throughout the entire mission the shuttle-orbit and attitude, as well as the position of the outboard relative to the inboard antenna were observed from the Attitude and Orbit Determination Avionics (AODA) instruments.

Not only the land surface but also the ocean heights were measured by SRTM and therefore it is able to "see" the elevation reference itself. Utilizing this effect the processing of the image strips - the so called "data takes" - always starts at the coastal areas.

The direct observation of the baseline as well as the fixing of the elevation processing on the sea surface significantly improve the processing throughput as the time consuming and mainly interactive determination of ground control points is not required. Additionally this procedure enables a homogeneous product quality.

SRTM/X-SAR provides a backbone for global DEM production. This net can be used for the absolute orientation of other DEMs as geometric reference but also for the improvement of the height quality by integrating elevation data from a variety of other sources by DEM fusion and mosaicking techniques.

2.2 ERS Tandem

Interferometric DEM generation is also possible using ERS tandem data. The processing differs to SRTM in some details but the main steps are identical. A big advantage of ERS is the multiple coverage. Beside the radar frequency the baseline length mainly determines the height sensitivity. Optimal configurations can be selected and combined in order to improve the DEM precision. In particular for Europe the best image pairs can be selected from a huge archive. E.g. figure 2 shows the ERS-tandem coverage of Germany.

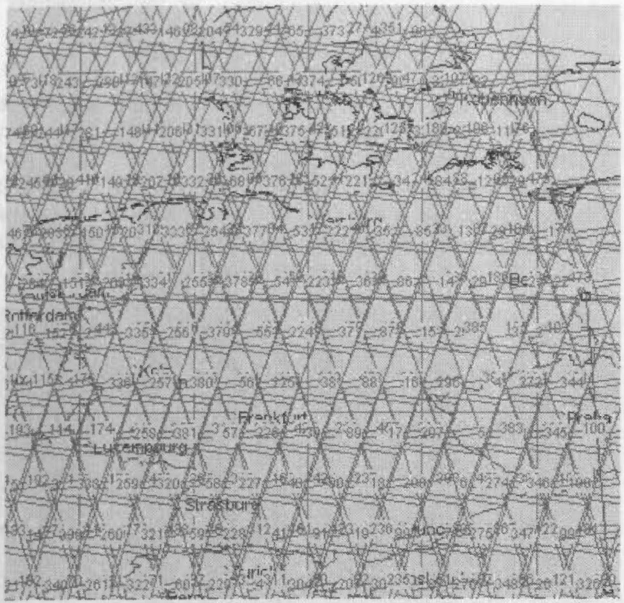


Figure 2. ERS tandem coverage of Germany

Limiting factors of ERS-tandem are the repeat pass conditions and the steep incidence angle. Temporal de-correlation is a significant limitation. Water bodies de-correlate within fractals of a Second. Dense vegetation and precipitation cause a loss of coherence, too. Whenever de-correlation appears the noise level in the interferogram increases. Atmospheric distortions additionally corrupt the phase information.

Common to all SAR systems is their side-looking geometry. SAR specific terrain distortions are foreshortening, layover and shadow. The derivation of extreme foreshortening, layover and shadow areas is not possible. ERS uses a relatively steep incidence angle of 23° that causes foreshortening and layover, but less shadow. Contrary SRTM / X-SAR has an incidence angle of 40°, which mainly causes shadow and less layover.

2.3 MOMS-2P

The MOMS-02 camera was flown on board the Priroda module of the Russian space station MIR during the mission MOMS-2P from 1996 to 1999. Approximately 65 million km^2 were imaged at an orbit inclination of 51.6° and from an altitude of 400 km, which resulted at a ground pixel size of 18 m and 6 m, respectively. The majority of all data takes was acquired in so-called mode D, that is a combination of two inclined stereo and two nadir looking multi-spectral channels with a swath width of approximately 100 km and a ground resolution of 18m. The stereo angle of 21.4° results in a base/height ratio of approximately 0.8, which results in a mean height accuracy of better than 10 m (Müller et al., 2001). Figure 3 shows a coverage map of MOMS-2P over Europe. Beside the along-track stereo capability a highly accurate navigation package (dual-antenna GPS receiver and gyros) was used for the first time in space with respect to civilian use.

The main limiting factor for DEM generation from MOMS-2P is, and this is of course true for all optical systems, the availability of cloud free images or images with low cloud coverage. Therefore, the resulting DEMs derived from these optical stereo data in most cases do not provide full coverage of large areas, e.g. whole countries or continents.

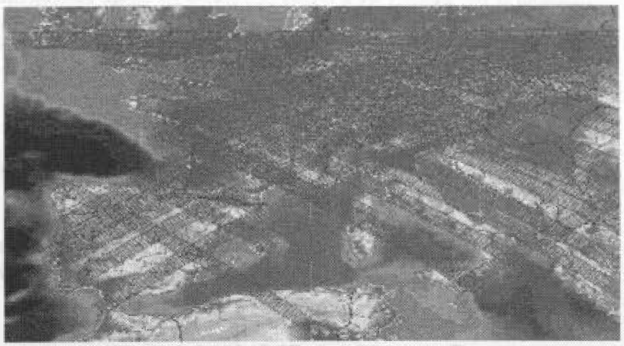


Figure 3. MOMS-2P coverage over Europe

3 DEM FUSION AND MOSAICKING

A line of processors was developed at DLR for interferometric DEM generation. It consists of four individual processing systems:

- The screener
- the SAR processor
- the interferometric processor (Genesis)
- the geocoding and mosaicking system (GeMoS).

Within the first two steps – the screening and the SAR processing – the corresponding images of the two antennas are individually prepared for the interferometric processing. This step comprises quality control, image formation and focussing of the SAR data. The output is the complex product which provides the amplitude/intensity and phase information.

The interferometric processor determines the interferogram of the differential phase, which is direct proportional to the difference of the distance between a ground position and the two antennas. The processing comprises spectral shift filtering, optional slope adaptive filtering, co-registration, multi-looking, coherence estimation, flat earth phase removal and the phase unwrapping (Eineder & Adam, 1997).

Finally the geocoding and mosaicking processor converts the phase into elevation information, geocodes the height values and compiles the individual DEMs to a large area data set (Roth et al, 1999).

For the paper at hand the mosaicking procedure is the most relevant part of the InSAR system. It compiles the individual DEMs to a large area data set. Overlapping areas from both crossing and adjacent scenes are utilized for cross-checking and an accuracy improvement. Multiple coverage reduces the height errors as several independent measurements can be combined. This procedure was developed for the generation of DEM mosaics from different sources, primarily based on interferometric SAR (Knöpfle et al, 1998). But it can also be used to incorporate DEMs provided by other sources and techniques, especially from optical stereo systems. The pre-condition is the availability of a height error description. The mosaicking and fusion algorithm takes into account the different prior accuracy of the DEMs and derives the quality of the resulting DEM.

3.1 Error sources

In general the quality of a DEM produced by SAR interferometry is determined by three different error sources (Bamler, 1997):

- the phase noise representing the measurement accuracy of the interferometric phase,
- the imaging geometry,
- possible atmospheric distortions in case of repeat pass interferometry.

A detailed description of the error sources and their functional dependency with the resulting height error is given in Knöpfle et al. (1998).

The accuracy of a DEM derived from optical stereo data is determined by the following groups of error sources:

- the measurement accuracy and density of tie points in the image space,
- the accuracy of control points in image and object space,
- the error budget of orbit and attitude determination and
- the overall imaging geometry (base to height ratio etc.).

The measurement accuracy of tie-points in the image space can generally be estimated from the matching procedure. For instance, if least squares matching is used, then the standard deviation of point determination can be derived within the estimation process for each point. This accuracy is mainly depending on the gray value gradients and the degree of correspondence of the homologous points. The overall accuracy including the influence of the imaging geometry can be determined from the inverse of the normal equation matrix for the determination of the exterior orientation parameters and the forward intersection for the computation of all ground points. Moreover, to take into account the subsequent interpolation process for the derivation of a regular grid DEM the density of points is also considered in the calculation of the resulting height error.

3.2 Height Error Map

The height error is calculated for each grid point in the raster DEM and is stored in a height error map. For each DEM a corresponding height error map with the same dimensions is generated, which is used as input for the following mosaicking procedure. The imaging geometry defines the scene specific level of accuracy, while the coherence (InSAR) or the correlation precision (optical stereo) are considered as local measure for each raster cell.

3.3 DEM Fusion and mosaicking

The mosaicking procedure creates large area DEMs from various input DEMs stored in the same coordinate system and pixel spacing. They need to be provided together with the corresponding co-registered height error map. This step also comprises the fusion of DEM data sets covering the same area. This is done by a weighted averaging of the individual height values within the overlap areas. Additionally a statistical outlier test is performed in order to identify and correct errors that cannot be detected in the individual models without a proper reference.

The height error map contains for each pixel an error estimate utilized as local weight during the averaging process. Whenever the elevation of the ground position is measured several times the result is statistically improved. The procedure allows the elimination of identified and labeled erroneous measurements in one of the single DEMs caused e.g. by disadvantageous imaging conditions like layover or shadow. In this case the low weight prevents from the consideration of the erroneous value.

4 PRACTICAL EXAMPLE

As test site an area in the south west of Germany was selected (figure 4). It covers approximately 110 km x 140 km of the Black Forest and the southern Rhine Valley. The area shows different types of terrain, the flat Rhine Valley up to the mountainous Black Forest and the Vogues in France. The heights are ranging from 140 m to 1500 m above mean sea level. The test site covers parts of Germany, France and Switzerland.

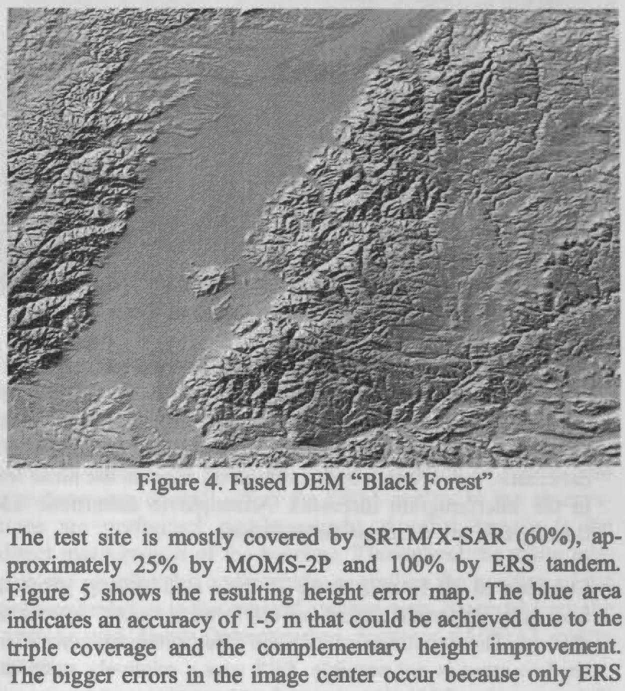


Figure 4. Fused DEM "Black Forest"

The test site is mostly covered by SRTM/X-SAR (60%), approximately 25% by MOMS-2P and 100% by ERS tandem. Figure 5 shows the resulting height error map. The blue area indicates an accuracy of 1-5 m that could be achieved due to the triple coverage and the complementary height improvement. The bigger errors in the image center occur because only ERS tandem is available. The densely forested area causes decorrelation. South of this area a sharp horizontal line appears where additionally MOMS-2P covers the test site.

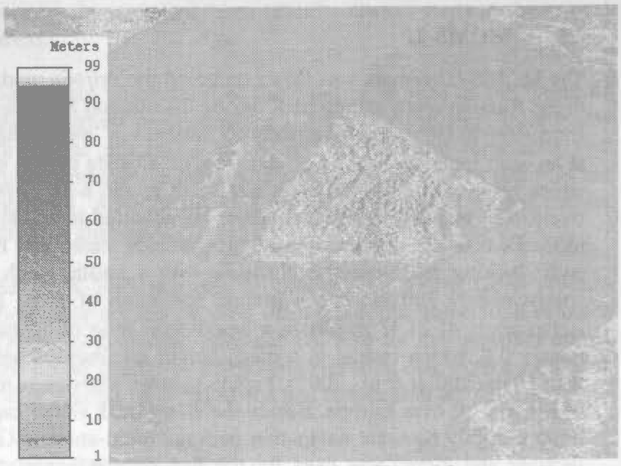


Figure 5. Fused Height Error Map

Comparing the color shaded DEM and the height error map it is obvious that the different DEMs could be seamlessly mosaicked. MOMS-2P and ERS tandem were adjusted to SRTM/X-SAR which refers to WGS84 as horizontal and vertical datum.

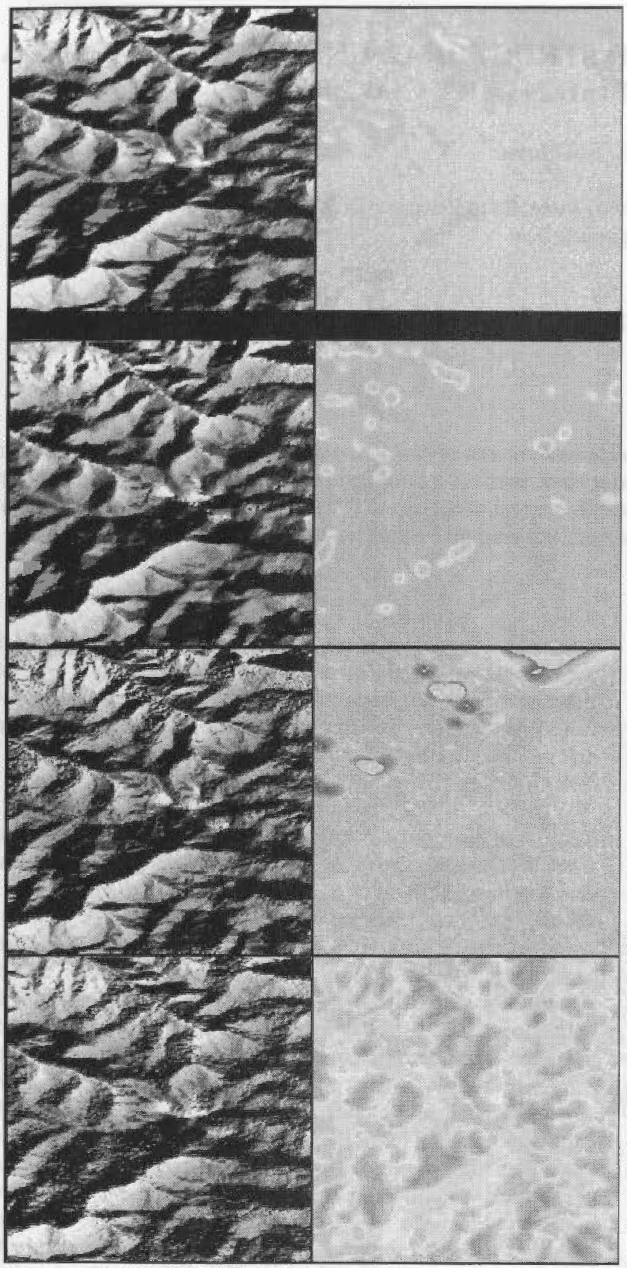


Figure 6. Detail of DEM and height error map fused data set (top), MOMS-2P, SRTM / X-SAR, ERS-Tandem (bottom)

The DEM-detail in figure 6 demonstrates the quality improvement of the complementary fusion. It covers an area of approximately 10 km x 10 km around the "Feldberg", the highest mountain of the Black Forest. It is surrounded by steep valleys. The two columns show the elevation model and the corresponding height error map (identical color bar as figure 5) of the fused data set, MOMS-2P, SRTM/X-SAR and ERS tandem.

The height error map of SRTM/X-SAR contains areas affected by radar shadow (brown to white) causing an increase of phase noise and a noisy appearance of the resulting DEM. However those areas were precisely mapped by MOMS-2P and the fusion DEM provides the best quality out of both.

5 CONCLUSIONS

Recent developments at DFD concerning the fusion and mosaicking of digital elevation models resulting from different sources have been described. The results show, that a seamless high quality elevation model is generated and the fusion considerably improves the quality of the DEM product. The tool is operated in the SRTM/X-SAR ground segment for the mosaicking of individual scenes to a large area DEM.

Further developments shall be performed regarding the utilization of the SRTM/X-SAR products as reference for the stereo processing of optical data. It shall be investigated whether tie- and control points can be derived from the X-SAR image.

REFERENCES

Bamler, R. (1997): Digital Terrain Models from Radar Interferometry. In: Fritsch, D., Hobbie, D. (Eds.): Photogrammetric Week '97, Wichmann, Heidelberg, 93-105.

Bamler, R., Hartl, P. (1998): Synthetic Aperture Radar Interferometry. Inverse Problems, 14 (4), 1-54

Eineder M., Adam N. (1997) A flexible system for the generation of interferometric SAR products. Proceedings IGARSS'97, Singapore

Hastings, D.A., Dunbar, P.K. (1998): Development and Assessment of the Global Land One-km Base Elevation Digital Elevation Model (GLOBE). International Archives of Photogrammetry and Remote Sensing, 32(4), 218-221.

Knöpfle W., Strunz G., Roth A. (1998): Mosaicking of Digital Elevation Models derived from SAR Interferometry. International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 4, 306-313.

Kornus W., Lehner M., Ebner H., Fröba H., Ohlhof T. (1999) Photogrammetrci point determination and DEM generation using MOMS-2P/PRIRODA 3-line imagery. ISPRS Workshop on "Sensors and Mapping from Space", Hannover

Leberl F. (1990) Radargrammetric Image Processing. Artech House, Norwood, MA, 595 pages

Müller R., Lehner M., Müller R. (2001) Verification of digital elevation models from MOMS-2P data. Proceedings ISPRS-Workshop, High resolution mapping from space 2001, Hannover, Germany

Pessagno C. (2000) SRTM status report. In: Dowman I. (ed) Validation of DEMs from satellite data. Reports and proceedings of 7th Meeting of Terrain Mapping Sub Group, CEOS WG Calibration and Validation, Gaithersburg, USA

Roth A., Knöpfle W., Rabus B., Gebhardt S., Scales D. (1999) GeMoS – A System for the Geocoding and Mosaicking of Interferometric Digital Elevation Models. Proceedings of IGARSS'99, Hamburg, Germany

Schroeder M., Kornus W., Lehner M., Müller R., Reinartz P., Seige P. (2000): MOMS – the first along track stereo camera in space: a mission review. Proceedings of the ASPRS Congress, Washington May 2000, CD-ROM, 8p.

Toutin T. (1999) Error Tracking of Radargrammetric DEM from RADARSAT Images. IEEE Transactions on Geoscience & Remote Sensing ,1999 , Vol. 37 , No 5 , pp. 2227-2238

SPECIAL-PURPOSE MAPPING OF BURIED STRUCTURAL UNCONFORMITIES IN A BASEMENT AND SEDIMENTARY COVER

Valentina B. Sokolova

Sevzapgeologia State Geological Enterprise, 24/1, Odoevski street, Saint Petersburg 199155, Russian Federation,
e-mail: szggp@eltex.ru

KEY WORDS: Deep Mapping, Remote Sensing Data, Photointerpretation.

ABSTRACT:

This paper deals with results of interpretation which, covering northwestern area of the East-European plain, has been based on a nontraditional technique. As a result of thus handling remote sensing data, particular information has been acquired as to structural pattern of geological formations in a Preriphean basement with a depth interval of 0.1 km to 9.0 km, as well as about the basement's tectonics and those geodynamic processes which the sedimentary cover overlying the Preriphean basement has been subject to.

1. INTRODUCTION

Within a period since 1966 up to 2001 Sevzapgeologia State Geological Enterprise has been working on a quite new and nontraditional technique of remote sensing data interpretation, just in order to meet additional requirements in respect to covered territories. Using such interpretation technique one can construct geological, structural and geodynamic sketch maps of structural stages' planes (scouring) within a depth interval of 0.1 km to 10 km, and even deeper. Definition of rock composition is based, for the most part, upon textural and structural patterns of buried bodies which, as seen on photos, show themselves through the cover in the form of symbols that reproduce these patterns.

2. CASE STUDIES

Major goal of interpretation of remote sensing data covering northwestern area of the East-European plain (EEP) within the confines of EEP-Baltic Shield conjuction zone consisted in producing sketch maps showing geological formations of the Preriphean basement, its tectonic framework, and geodynamic processes in a sedimentary rock complex. As it follows from the information thus obtained, three components, differing in terms of composition of rock complexes, constitute the area in question, namely (1) metamorphic formations, in particular - three varieties of gneiss (according to succession in which they have formed and their textural-structural features), and crystalline schist (2) intrusive and volcanogenic formations: ultrabasic rocks in tension zones, gabbroids, dikes, endogenic dikes including, volcanites that vary in composition and make up depression slopes, and (3) dynamometamorphism zones made up of cataclastic rocks, tectonites with slide marks, filling extended fault zones, as well as fusion and torsion tectonites. A conventionally accepted model reproducing succession in which they had formed has been constructed.

As long as tectonics is concerned (Figure 1), established was the fact that in the said area the major portion of the Preriphean basement's plane betrays that large structural forms striking NE, NW, in meridional and latitudinal directions. Meridional section (lake Vozhe projection) of the territory over a larger portion of its plan view is cut into one western and one eastern megablock, these two contrasting in intensity of tectonic processes. In the western active megablock an arcuate structural zone was outlined as the one that had undergone incessant process of activation. What is more, Sevzapgeologia SGE was the first to detect and plot here six flat troughs ranging in size from 55 x 45 km to 115 x 150 km.

These troughs could be compared with volcano-tectonic depressions regarded as promising in terms of mineral deposits. Further southwards off them, in the adjacent Tver province, nearby the town of Tver, mapped was a classical specimen of a large spiral structure with a downslope movement. Its total area amounts to 10,000 sq. km, and it seems most likely that the structure originated in upper layers of the basement under the influence of one-sided mechanical movements. The eastern megablock has become consolidated and steady in regard to the western megablock, and a limited amount of intrusive formations has injected its body. Along its meridional sections the megablock is divided into three small-scale blocks, and along the boundaries thereof displacements, varying in amplitude, took place while these blocks had been forming. Such movements are reckoned to be as large as 100 km. Once the eastern megablock had become dissected along meridionally oriented faults, it never underwent more or less intense tectonic movements since then. Latitudinal block structure is 15-60 km wide and it extends from the western periphery of the investigated area to the eastern one for more than 1,100 km. It is known to run, either definitely or less surely, across all structural zones, and it features unsteady tectonic setting.

Evidences of tectonic activity are shown on a sketch map of geodynamic processes in sedimentary rocks (Figure 2). Within the confines of the whole area under research faults were detected and classified according to injection fillers.

The sketch map of geodynamic processes illustrates those movements that affected all sedimentary strata varying in overall thickness from 0.1 km to 9.0 km. The following conclusion was drawn from comparing (a) tectonic sketch map of Preriphean basement, (b) additional structural and tectonic sketch maps of two levels of scouring surfaces in sedimentary strata, and (c) sketch map of Cenozoic processes. The latitudinal block structure and arcuate structural zone proved to be active objects. Other geodynamic events occurred, and indeed are

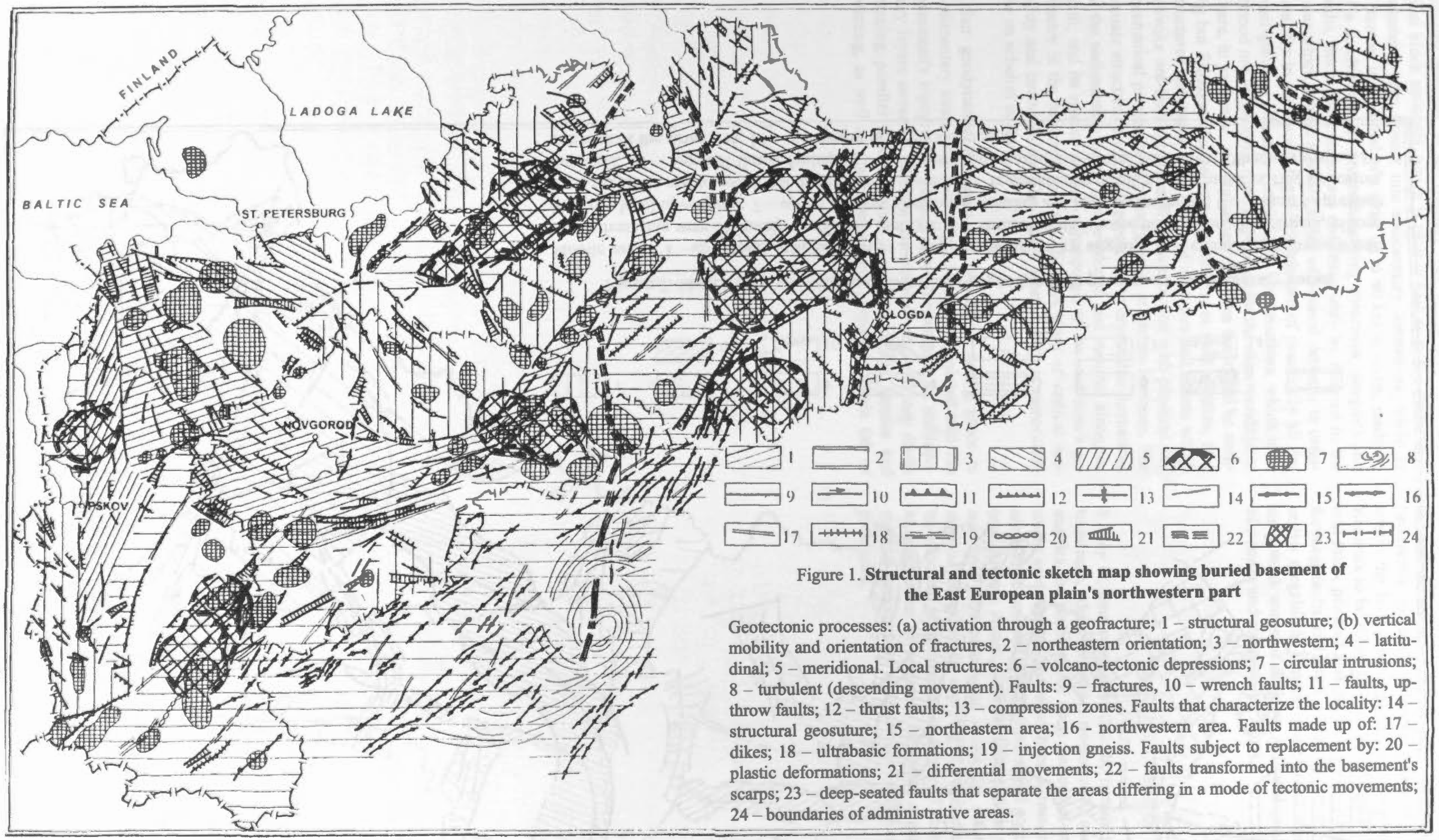


Figure 1. **Structural and tectonic sketch map showing buried basement of the East European plain's northwestern part**

Geotectonic processes: (a) activation through a geofracture; 1 – structural geosuture; (b) vertical mobility and orientation of fractures, 2 – northeastern orientation; 3 – northwestern; 4 – latitudinal; 5 – meridional. Local structures: 6 – volcano-tectonic depressions; 7 – circular intrusions; 8 – turbulent (descending movement). Faults: 9 – fractures, 10 – wrench faults; 11 – faults, upthrow faults; 12 – thrust faults; 13 – compression zones. Faults that characterize the locality: 14 – structural geosuture; 15 – northeastern area; 16 - northwestern area. Faults made up of: 17 – dikes; 18 – ultrabasic formations; 19 – injection gneiss. Faults subject to replacement by: 20 – plastic deformations; 21 – differential movements; 22 – faults transformed into the basement's scarps; 23 – deep-seated faults that separate the areas differing in a mode of tectonic movements; 24 – boundaries of administrative areas.

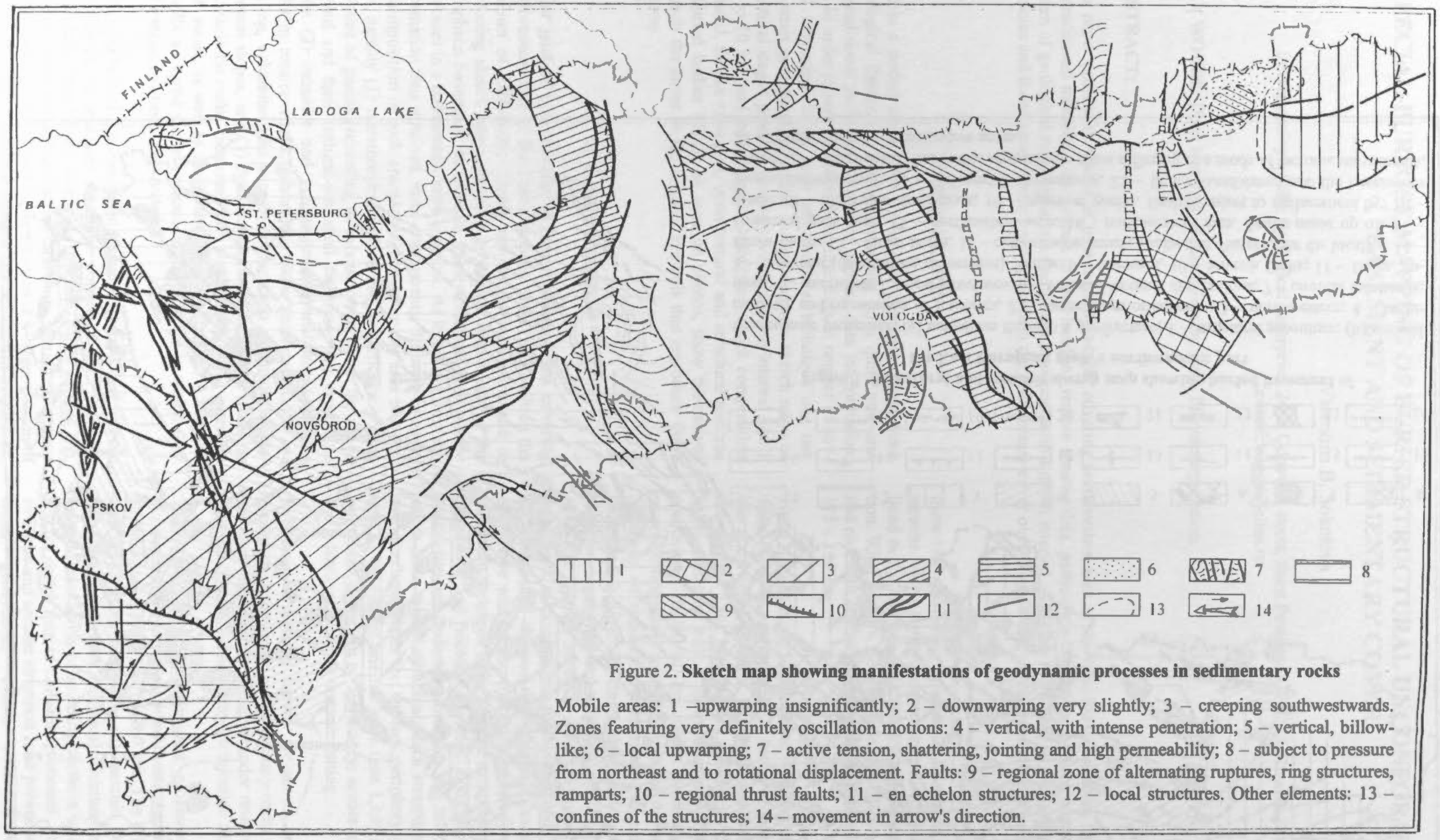

FINLAND
LADOGA LAKE
BALTIC SEA
ST. PETERSBURG
NOVGOROD
PSKOV
VOLOGDA
1
2
3
4
5
6
7
8
9
10
11
12
13
14

Figure 2. **Sketch map showing manifestations of geodynamic processes in sedimentary rocks**

Mobile areas: 1 –upwarping insignificantly; 2 – downwarping very slightly; 3 – creeping southwestwards. Zones featuring very definitely oscillation motions: 4 – vertical, with intense penetration; 5 – vertical, billow-like; 6 – local upwarping; 7 – active tension, shattering, jointing and high permeability; 8 – subject to pressure from northeast and to rotational displacement. Faults: 9 – regional zone of alternating ruptures, ring structures, ramparts; 10 – regional thrust faults; 11 – en echelon structures; 12 – local structures. Other elements: 13 – confines of the structures; 14 – movement in arrow's direction.

actually recorded, in sedimentary rock series only. The latitudinal block structure, which is plainly mature in the Preriphean basement, runs further into sedimentary formations, while in Cenozoic it is plotted as a strip, up to 40 km in width, made up of a combination of proportional ring structures, ramparts and faults. Such block structures can reasonably be referred to the areas which caused a whole series of mineral deposits to originate, for in such areas there is a mixture of practically all deposit-growing initiative elements, or constructions, which have formed not at the same time and not under the same conditions. There, in particular, are high-angle fractures that could be nothing but filtration channelways or feeders, as well as dikes, ring structures and many other forms of elements and events which provoke ore-forming processes in connection with physical and mechanical properties of country rocks at certain horizons. The arcuate structural zone, well preserved in crystalline formations of the sedimentary strata, has developed and matured along the fault, and its eastern portion corresponds to the said zone's geometry in the basement and still displays signs of vertical mobility and permeability. The zone's western part is remarkable for en echelon faults.

Other geodynamic processes are plotted as taking place in sedimentary strata, and their origin is probably related with the basement's topography. Thus, billow-like warping of sedimentary layers occurred within a gap between regional step faults running parallel to each other. Zones of intense tension and jointing, as well as shatter zones and permeable zones (all of them being so common in Cenozoic rock complexes) are forming and developing over the tilted segments of the Preriphean basement's surface. Additional maps of two scouring levels within the formations and in the cover enabled the following elements to be reconstructed therein: fault-line uplifts and flexural folds, plicated and ring structures, grabens, regional block-forming faults and fault zones. As an exception, the western active megablock features widespread shatter zones, tectonically weak zones of different rank and genesis, as well as mobile zones and compression zones.

3. CONCLUSIONS

It may be thus concluded that the nontraditional interpretation technique, while it requires minimum of costs, labour and time, enables the problems of deep geological mapping to be effectively sold, and on the other hand it reasonably cuts down the areas to be covered by geophysical and geochemical exploration and drilling. Again, such technique gives power to check and maintain geodynamic and ecological safety when developing mineral resources. By applying this technique, the territories can be allocated and classified in terms of safeness, zones exposed to danger can be outlined, industrial averages and emergency state of environment with ruinous ecological aftereffects can be precluded.

ELABORATION OF PRODUCTS DERIVED FROM THE EO DATA AND GIS INFO-LAYERS FOR FLOODING RISK ANALYSIS

Gheorghe Stancalie*, Corina Alecu, Elena Savin

National Institute of Meteorology and Hydrology (INMH), 97, Soseaua Bucuresti-Ploiesti,
Sector 1, 71552 Bucharest, Romania – (george, corinaal, esavin)@meteo.inmh.ro

KEY WORDS: Natural Hazards, Flooding, Earth Observation (EO), Risk Indices, Geographic Information System (GIS), Sensitive Area, Land Cover/Land Use, Image Processing

ABSTRACT:

The risk of flooding due to runoff is a major concern in many areas around the globe and especially in Romania. The modern management of spatial data related of river flooding largely relies on the functional facilities supplied by the Geographic Information System (GIS) info-layers, combined with Earth Observation (EO) data-derived information, Digital Elevation Models (DEM) and hydrological modeling. The paper assumes this multidisciplinary approach, in view to establish a methodology, which should further allow the elaboration of products useful for flooding risk analysis, such as: more accurate updated maps of land cover/land use, comprehensive thematic maps at various spatial scales with the extent of the flooded areas and the affected zones, maps of the hazard prone areas, etc. The supplying of the products, extracted directly from EO data or obtained after regrouping digitizing and integrating them in a GIS environment, has highly contributed to the elaboration of precise thematic maps at different scales, with corresponding generalizing levels. There are stressed the facilities supplied by the geo-referenced information and the satellite images to manage floodings during their characteristic phases: before, during and after floodings; accent is laid on the pre and post-crisis phases. The application was developed for the Arges hydrographic basin in Romania, a critical area, keeping in mind that it withholds many localities, including the capital and also important economic centers.

1. INTRODUCTION

Ever since the first satellite observations of the Earth, numerous studies have been achieved using information from the satellite systems and the facilities of GIS technology in the topic of managing flooding connected phenomena. Satellite images are objective information sources, available in time and relatively cheap, for the determination of parameters necessary to monitor and assess floods and their consequences (Wang et al., 1995).

The main contribution of EO derived information could be envisaged at the level of mapping aspects. EO satellites can provide information necessary to elaborate cartographic documents, important for hazard and vulnerability mapping. EO derived data are useful to locate, measure and spatially represent up-to-date information such as: flooded areas, parameters on vegetation, human presence and terrain properties, which are directly used in the decision - making process.

In order to obtain high-level thematic products the EO useful extracted data must be integrated with other non-space ancillary data (topographical, pedological, meteorological data) and hydrologic/hydraulic models outputs. So, the management of data helping to monitor the generated high floods largely relies on an integrated approach using the functional facilities supplied by the GIS (Harms et al., 1996).

This approach may be used in different phases of establishing the sensitive areas such as: the management of the database - built up from the ensemble of the spatially geo-referenced information; the elaboration of the risk indices from morpho-hydrographical, meteorological and hydrological data; the interfacing with the models in order to improve their compatibility with input data; recovery of results and the possibility to work out scenarios; presentation of results as synthesis maps easy to access and interpret, additionally adequate to be combined with other information layouts resulted from the GIS database.

2. STUDY AREA – THE ARGES BASIN

The study area represented by the Arges hydrographic basin is affected by flooding as ascertained by the frequency of these phenomena: 1972 – 1973, 1975, 1979, 1983, 1991, 1997, 1998. This area is so much the more critical one, considering that it comprises the capital, as well as several other important economic centres. As regards the results damages, the high flood of March 29 - April 10 1997, alone affected nearly 600 households and annexes, over 4000 ha of arable land, 4 bridges, more than 40 footbridges and 40 Km of roads, along with important hydrotechnical constructions.

Situated in the south of the country, the Arges river, its tributaries included, constitutes one of the most important hydrographic basin in the country with respect to its hydro-energetic potential and the water supplies of the industrial centres and crop lands. Through the position it occupies in a physico-geographic ensemble with specified characters, with a moderate continental climate, sufficiently humid in the mountain areas and drier in the plain, is playing a large variety of the relief and lithology, the Arges river hydrographic basin has a complex hydrological regime.

3. STRUCTURE, ATTRIBUTES AND FUNCTIONS OF GIS DATABASE

The database was structured as an ensemble of file-distributed quantitative and qualitative data.

Organizing the files was focused on a relational structure as information thematic layouts. With this approach, spatial objects may be described through three main property classes: those determining the objects position on the Earth surface, those establishing spatial relations with other objects and those referring to the spatial objects attributes.

Coding the existing relationships between spatial entities was performed with the help of facilities supplied by the complete topology of the PC ARC/INFO software. The achievement was viewed of a database apt to allow elaborating indices of hydrological risk to flooding, made up of information associated to three main criteria:

- morphology and topography: hydrographic network, dams and channels network, sub-basins, confluence order; hydrographic and/or drainage network density; mean slopes of the sub-basins; mean slopes in different river sections;
- land cove/land use: land impermeability, in-soil water resilience to infiltration;
- vulnerability to flooding: communication ways network (highways, railroads), localities, meteorological, hydrometric stations network and rain measurement points network.

GIS database info-layers have been divided in three large categories:

- those relative static (land topography and altimetry, basin and sub-basin boundaries, the communication ways network, the localities, the network of meteorological and hydrometric stations);
- those with specific dynamic characteristics which required periodical updates, with frequencies of the order of year (land cover/use, hydrographical network, the network of dams and canals);
- those determined by particular events.

The database was connected so as to allow obtaining synthetic representations of the hydrological risk for the Arges basin, through using the risk parameters separately or combined. For instance, there can be considered the mean slopes of the mountainsides, the in-soil water absorption, the water resilience to infiltration and the vulnerability degree. The advantage consists in that the system allows greater parameters combination diversity, allotted with various weights, associated to the mentioned three main criteria.

4. ELABORATION OF PRODUCTS FOR FLOODING RISK ANALYSIS

Within the framework of flood surveying, optical and radar satellite images can provide up-to-date geographical information. Integrated within the GIS, flood derived and landscape descriptive information are helpful during their characteristic phases of the flood (Tholey et al., 1997) :

- before flooding, the image enables the description of the land cover of the studied area under normal hydrological conditions;
- during flooding the image data set provide information on the inundated zones, flood map extent, flood's evolution;
- after flooding, the satellite image point out the flood's effects, showing the affected areas, flood deposits and debris, with no information about the initial land cover description unless a comparison is performed with a normal land cover description map or with pre-flood data.

The products useful for flooding risk analysis, referred to: accurate updated maps of land cover/land use, comprehensive thematic maps at various spatial scales with the extent of the flooded areas and the affected zones, maps of the hazard prone areas.

4.1 EO data processing and analysis

Optical (LANDSAT-TM, IRS-PAN/LISS, SPOT-AN/XS) and radar satellite data (RADARSAT) have been used to perform the analysis for inventory purposes and different kind of flood related thematic information.

A series of specific processing operations for the images were performed, using the ERDAS Imagine software; geometric correction and geo-referencing in the stereographic 70 map projection system (using control points on the 1/50 000 scale map), image improvement (contrast enhancing, slicking, selective contrast, combinations between spectral bands, re-sampling operation), statistic analyses (for the characterization of classes, the selection of the instructing samples, conceiving classifications).

Optical high resolution data have been used to perform the analysis for the inventory purposes under normal hydrological conditions as well as for determining the hydrographic network. The radiometric information contained in these images allows the derivation of both biophysical criteria and those from human activity, through supervised standard classification methods or advanced segmentation of specific thematic indices. Once extracted, these geographical information coverages were integrated within the GIS for further water crisis analysis and management.

The interpretation and analysis of remotely sensed data in order to identify, delineate and characterize flooded areas was based on relationships between physical parameters such as reflectance and emittance from feature located on the surface: reflectance and/emittance decreases when a water layer covers the ground or when the soil is humid; also reflectance and/emittance increases in the red band because of the vegetation stress cause by moisture; reflectance and/emittance changes noticeably when different temperatures, due to thick water layer are recorded.

In the microwave region the water presence could be appreciated by estimating the surface roughness, where water layers smooth surfaces dielectric constant is then heavily correlated to soil water content. In case of radar images the multitemporal techniques was considered to identify and highlight the flooded areas. This technique uses black and white radar images of the same area taken on different dates and assigns them to the red, green and blue colour channels in a false colour image. The resulting multi-temporal image is able to reveals change in the ground surface by the presence of colour in the image ; the hue of a colour indicating the date of change and the intensity of the colour the degree of change. The proposed technique requires the use of a reference image from the archive, showing the « normal » situation.

4.1.1. The land cover/land use mapping: The methodology for the achievement of the land cover/land use from medium and high-resolution images developed within the Remote Sensing & GIS Laboratory of NIMH (Stancalie et al., 2000), is based on the observation of the following requirements:

- the structure of this type of information must be at the same time cartographic and statistic;
- it must be suited to be produced at various scales, so as to supply answers adapted to the different decision making levels;
- up-dating of this piece of information must be performed fast and easily.

The used methodology implies following the main stages below:

- preliminary activities for data organizing and selection;
- computer-assisted photo-interpretation and quality control of the obtained results;
- digitisation of the obtained maps (optional);
- database validation at the level of the studied geographic area;
- obtaining the final documents, in cartographic, statistic and tabular form.

Preliminary activities comprise collection and inventorying of the available cartographic documents and statistic data connected to the land cover: topographic, land survey, forestry, and other thematic maps at various scales.

To obtain the land cover map, images used must have a fine geometrical resolution and rich multispectral information. That is why the data preparation stage consisted in merging data obtained from the panchromatic channel, which supply the geometric fineness (spatial resolution of 5 m for the IRS, 10 m for the SPOT), with the multispectral data (LISS for IRS, XS for SPOT), which contain the multispectral richness.

The computer-assisted photo-interpretation finalizes in the delimitation of homogeneous areas from images, in their identification and framing within a class of interest. Discriminating and identifying the different land occupation classes rely on the classical procedures of image processing and leads to a detailed management of the land cover/land use, followed by a generalizing process, which includes:

- identification of each type of land occupation, function of the exogenous data, of the "true-land" data establishing a catalogue;
- delimitation of areas suspected to represent a certain unity of the land;
- expanding this delimitation over the ensemble of the image areas, which display resembling features.

Validation of results from photo-interpretation, mapping (by checking through on land sampling at local and regional level) and building up the database aims at knowing the reliability level and the precision obtained for the delimitation of the units and their association to the classes in the catalogue.

The satellite based cartography of the land cover/land use is important because it makes possible periodical updating and comparisons, and thus contribute to characterize the human presence and to provide elements on the vulnerability aspects, as well as the evaluation of the impact of the flooding.

The figure 1 presents the updated land cover/land use maps (7 main classes) for a test area situated in the middle section of the Arges basin, characterized by a large variety of the land cover/use classes. The product was obtained using IRS/PAN and IRS/LISS images from 9 April and 23 September 1999.

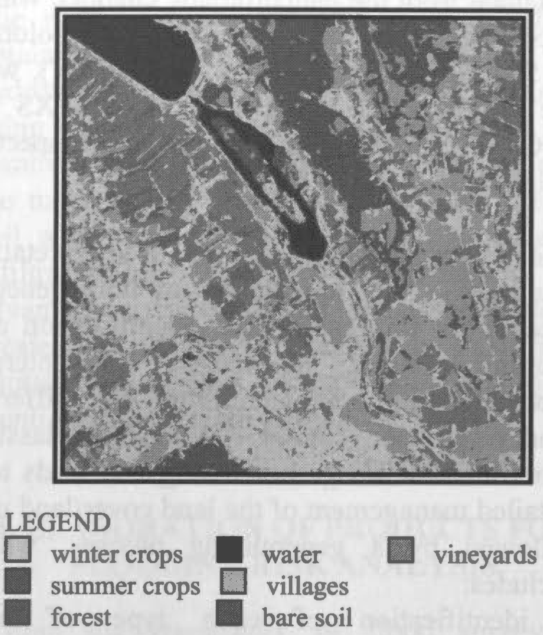


Figure 1. The land cover maps for the test-zone in the Arges basin, obtained from IRS images.

4.2 Methodology for the flooding risk maps preparation

The assessment of the flooding risk hazard requires a multidisciplinary approach; coupled with the hydraulic modelling, the contribution of geomorphology can play an exhaustive and determining role using the GIS tools.

In order to obtain the flood risk map for the test-area, a important step refers to construction of the DTM and its integration, together with the land cover/land use maps in the GIS using a common cartographic reference system.

The DEM was realized after the following steps:

- scanning the maps. 12 plans at 1:5000 scale and 3 topographic maps at 1:25000 were scanned at 300 dpi resolution;
- geo-referencing the maps. The scanned images were geo-referenced in STEREO 70 projection;
- colour separation. Each raster information layer is extracted as a linear image (black & white without gray tones;
- vectorization of raster images.
- merging the maps;
- generating a triangulated irregular network (TIN) model.

In a TIN the point density on any part of the surface is proportional to the variation in terrain. A surface is a continuous distribution of an attribute over a two-dimensional region. TIN represents the surface as contiguous non-overlapping triangular faces. The surface value is estimated in this way for any location by simple (or polynomial) interpolation of elevation in a triangle. As elevations are irregularly sampled in a TIN, it is possible to apply a variable point density to areas where the terrain changes sharply, yielding an efficient and accurate surface model. A TIN preserves the precise location and shape of surface features.

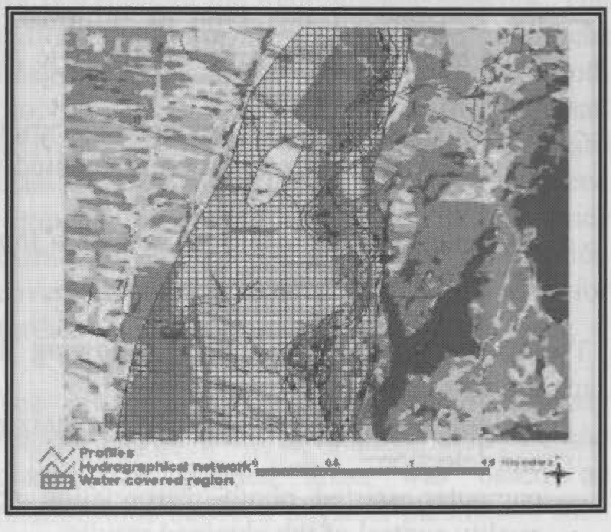


Figure 2. The risk map for the test-zone in the Arges basin

Various morphological criteria may be extracted from a DEM, designed using topographical maps and geodetic measurement criteria, such as altitudes, slopes, exposures, transversal profiles and thalweg locations. All these parameters are useful to evaluate the local or cumulated potential flowing into a zone of the basin. This approach is also useful to get a realistic simulation of floods, taking into account the terrain topography, the hydrological network and the water levels in different transversal profiles on the river, obtained from the hydraulical modelling.

Using the GIS database that include the DTM, the land cover/land use maps and the vector info-layers (hydrographical, dams and canals networks, the communication and localities network etc) several outputs of hydraulic simulations (with different flood probability) could be superimposed in order to elaborate the risk maps to flooding.

In the figure 2 an example of a risk map for a test area of the Arges basin is presented.

5. CONCLUSIONS

Considering the necessity to improve the means and methods to assess and monitor flooding, the paper presents the capabilities offered by remotely sensed data and GIS techniques to manage flooding and the related risk. The applications were developed on the Arges basin in Romania.

Although satellite sensors cannot measure the hydrological parameters directly, remote sensing can supply information and adequate parameters to contribute to identify and map the hydrological risk at the basin level (Puyou-Lascassies et al., 1996).

The basic concepts of the creation of products useful for flooding risk analysis (maps of land cover/land use, thematic maps of the flooded areas and the affected zones, flooding risk maps) can be summarized in the following stages:

- preparation of the EO data: radiometric, geometric correction and other basic image processing;
- combination of the interpretation techniques: automated and photo-interpretation tools, as well as radiometric techniques;
- multi-sensor approach: make the analysis process easier and provide the opportunity to improve the quantity and quality of the information obtained;
- multi-temporal approach: gives the possibility to monitor low frequency (land modification) and high frequency phenomena (evolution of the floods boundaries);
- integrated approach: the high level products are based on the EO derived information combined with other ancillary data, hydrologic/hydraulic models outputs using the GIS facilities.

REFERENCES

Harms J., Massue J.P., Tarel G., Pirre G., Petit J.C., Meyer-Roux J. and Cinotti C., 1996. Use of space technologies for major risks management. In: *Proc. of the 16-th Symposium: Integrated Applications for Risk Assessment and Disaster Prevention for Mediterranean,* EARSel Publications, pp84.

Puyou-Lascassies Ph., Harms J., Lemonnier H. and Cauzac J.P., 1996. Evaluation de contribution des moyens spatiaux a la gestion du risque d'inondation. *Projet de demonstration,* SCOT Conseil, pp. 54.

Stancalie G., Alecu C., Catana, S., Simota M. 2000. Estimation of flooding risk indices using the Geographic Information System and remotely sensed data. In: Proc. of the *XX-th Conference of the Danubian Countries on hydrological forecasting and hydrological bases of water management,* Bratislava (on CD).

Tholey N., Clandillon S, Fraipont P., 1997. Flood surveying using Earth Observation Data. In: *Proc. of the Eurisy Colloquim "Earth Observation and the Environment: benefits for Central and Eastern European Countries",* Budapest, pp. 77 – 88.

Wang Y., Koopmans B. and Pohl C., 1995. The 1995 flood in The Netherlands monitored from space – a multisensor approach. International *Journal of Remote Sensing,* vol. 16, pp. 2735 – 2739.

INTEGRATION OF REMOTE SENSING DATA AND FIELD MODELS OF IN-SITU DATA IN A GIS FOR ENVIRONMENTAL SENSITIVITY INDEX MAPPING; A NIGERIAN EXAMPLE

Fabiyi O.Oluseyi

Department of Geography
University of Ibadan
Ibadan, Nigeria.
E-mail seyifabiyi@yahoo.com
Tel 234-02-8108090,or 234–02- 8102900
Mobile: +234 08023271512

Keywords: Environmental Sensitivity index, GIS,

ABSTRACT

Environmental Sensitivity Index (ESI) is used to measure the degree of impact of a hazard-causing event within a given environment, in order to be able to put in place an adequate contingency plan. Until recently ESI had mostly been calculated based on geomorphological concept, recently however the focus is on a combination of many factors that are directly or indirectly influenced by the focused activity, such as physical, biological and anthropological and socio economic factors.

The need to integrate data from different sources in the GIS platform brings up a question of efficiency of data integration. The varying data sources utilize different data models and accuracy level; it raises a methodology question on the proper integration of data from different platforms without introducing high error level of estimation.

The paper attempts preparation of ESI of an inland oil field based on a combination of remotes sensing data (SPOT XS) and sampled data collected from the field. The approach utilized map algebra in ArcView Spatial-analyst environment to combine unique indexes from various sources to arrive at a single ESI atlas. The paper also identifies contribution of such method in developing countries, where extensive part of the country does not have adequate information for preparation of such maps

INTRODUCTION

Environmental Sensitivity Index (ESI) is used to measure the degree of impact of a hazard-causing event within a given environment. The focus of most ESI is to measure the level of sensitivity of the environment to oil spill. ESI thus serves as a valuable approach to Oil spill risk assessment, as the result would depict the degree of risk of each environmental component to oil spill. Until recently ESI had mostly been calculated based on geomorphological concept. But the international standards stipulate that ESI should rather be calculated based on a combination of various parameters, such as physical, biological and anthropological and socio economic factors.

Prior to the advancement of GIS some form of environmental index mapping have been produced using a refined method of conventional map making. With the advent of GIS the research interest has been on the appropriate methods to integrate this new technology to the preparation of ESI mapping. ESI mapping is meant to be a unique surface interpolation of all the environmental components in the environment captured with a single index that represents the sensitivity level of different locations to particular hazard causing event such as oil spill.

Using GIS for ESI require some basic concepts and assumption especially as it relates to issues of index assignment and map generalization. The paper advances a procedure to map the resources in a given environment for the purpose of GIS based mapping of environmental components. It also ranks the resulting classification into different sensitivity index level.

Three issues were addressed in this paper.

1. Spatial interpolation of field data for mapping of environmental components.

2. Integration of data from different sources and level of occurrence to map environmental components
3. Environmental sensitivity classifications based on the Combined integration of different data and different weighting factors.

Theoretical Underpinning.

One of the goals of ESI is to establish in quantitative terms the value of the surrounding environment, generally referred to as Environmental Sensitivity Index (ESI). The definition of ESI of a particular zone provides an environmental state of reference that can be readily used for contingency planning and regularly updated as new elements or changes come into play.

Remote sensing has been acknowledged as invaluable in both preparation of ESI and surveillance of changes caused by hazard causing events. Remote sensing is particularly useful in areas with poor mapping materials like Nigeria.

However despite the valuable contribution of remote sensing to the preparation of environmental sensitivity index mapping; remote sensing on its own is not sufficient to estimate the sensitivity level of the environment. The in situ data such as soil survey, bacteria count, conductivity level of soil, socio economic characteristics of the inhabitant, animal habitat and count among others. Though some level of estimate by deduction can be done from the remote sensing products but ESI require such level of accuracy of estimate than can be deduced from the remote sensing data.

The need to integrate data from different platforms in the GIS environment brings up a question of efficiency of data integration. The variant data sources utilize different data models; it raises a methodology question on the proper integration of data from different platforms without magnifying the existing errors in constituent data. Data obtained from the field can be captured in GIS platform through different model approach. . In this study however the field model is adopted.

Goodchild (1993) describe field model as an attempt to represent the spatial variation of a single variable using a collection of discrete objects. A spatial database may contain many such fields or layers, each able in principle to return the value of one variable at any location (x,.y) in response to a query, and fields may be associated with variables measured on either continuous or discrete scales.

He identified six common field models in GIS, which are:

1. Irregular point sampling: the database contains a set of turples (x,y) representing sampled values of the variable at a finite set of irregular spaced location.
2. Regular point sampling: as above but with points regularly arrayed, normally on a square or rectangular grid (e.g Digital elevation model)
3. Contours: The database contains a set of lines., each consisting of an ordered set of (x,y) pairs each line having an associated z value; the points in each set are assumed connected by straight lines.
4. Polygons: the area is partitioned into a set of polygons such that every location falls into exactly one polygon (quardrant);
5. Cell grid: The area is partitioned into irregular grid cells the value attached to every cell is assumed to be the value of the variable for all locations within the cell.
6. Triangular net: the area is partitioned into irregular triangles; the variable is specified at each triangle vertex

The paper attempts preparation of ESI of a part of Nigerian Niger Delta with specific focus on oil related substances using a combination of remotes sensing data (SPOT XS) and sampled data collected from the field. The approach utilized map algebra in ArcviewSpatial analyst environment to combine unique indexes from various sources to arrive at a single ESI atlas.

One of the major problems faced in dealing with field data in a GIS is the adequate sample size to generate surface interpolation. What level of data can be measured to generalize the occurrence of environmental parameters. Generally speaking ESI indexing is an attempt to arrive at a singe quantitative value for a given space which represents the level of sensitivity to a particular human activity or natural phenomenon.

In other word it is essentially reclassifying the space into discrete groups of sensitivity levels. Classification in geography has been observed as the systematic grouping of objects or event into classes on the basis of properties or relationships that they have in common.

This can be done via logically, subdividing a population or agglomerating like- individuals (i.e. (1) subdivision methods (2) agglomeration methods.)

Subdivision methods: this is done through carefully defined criteria, usually which can be the purpose or absence of one or more attributes of classification purpose of focus often referred to as classification scheme.

Agglomeration methods: takes a number of individuals and assembles item into classes according to some grouping procedure.

One major attribute of classification is that it must be exhaustive. the simplest classification is the one that involve one criteria in which case the element involved are categorized into two in or out, yes or no, true or false. GIS packages readily classify issue, into this limited and simple scheme.

In any research therefore that involves classification the researcher must develop or adopt a classification scheme, which must be exhaustive on the basis of the project at hand.

The classification scheme can be project/research based or universal.

GIS in environmental researches brought about a basic question of integrating field dependent models, and classification scheme in GIS. It is impractical for a specific GIS to provide all analytical capabilities for all kinds of users. On the other hand, to produce a specific software for every field or project would not only be unreasonable but would further aggravate the problems of rigid boundaries between disciplines which GIS seek to reduce. Chud and Ding 1992 suggested that spatial analysts develop their own models and integrate those models with a GIS for their different purposes e.g. Willer(1990).

The various suggestions implies that each GIS user must be able to combine knowledge of spatial relationships with programming capabilities which will inevitably makes GIS to be an alienating approach. The focus of GIS is to reduce the rigid boundaries between geographers and other professionals that have interest in spatial attributes of the fundamental parameters of their subject.

In the process of classifying the space into groups based on the data obtained about objects of phenomenon in space it has been rightly established that some form of prediction is involved. Demers (2000). According to him spatial data taken from sampling yields three major types of manipulation: data at non-sample location can be predicted from sampled location in order words data from one set of spatial units can be converted to others with different spatial configuration.(Muehrcke and Muehrcke1992) The non sampled locations are predicted from the sampled areas. It is noteworthy however that the prediction of non-sampled spaces involves some form of assumption, which may miss or meet the target of spatial interpolation or extrapolation.

For a given ESI maps it is assumed that data of given index value are essentially uniform and homogeneous, but it is clear that this is not possibly true in any given context therefore some form of assumption and error limit must be set to establish a unique single index for each spatial units in the area of concentration. However the smaller the unit of aggregation the better for the GIS based index mapping. Sometimes grid or raster cell serves as units but it is clear that the cell values are essentially aggregated.

Since it is often impracticable to collect data from every spatial unit in the study area. of the
This paper seeks to devise taxonomy of spatial classification of environmental parameters in a GIS context and examine the level of accuracy of such classification.

We shall not attempt programming a new extension but would apply different methods of classification of field data to arrive at a given index. The paper represents a methodological contribution to spatial modeling in the GIS environment.

3.0 Methods.

The data ere obtained in-situ from the site with the use of GPS and a large image format to identify the area with the remote sensing images. The sample points were chosen based on random selection of point and the chosen point form site for data collection. The samples collected include soil, water and socioeconomic characteristics of the study area. Though remote sensing material used is adequate enough to examine the physiographic (above surface) component of the study area, the subsurface elements require in-situ data collected from the surface. The baseline data are equally important to the accurate estimate of the sensitivity index mapping.

The baseline data collected were analyzed in a commercial laboratory and a part summary of the result of the analyses was presented on Appendix 1. After different types of analysis were performed on the data components by soil and water experts. The result obtained was therefore tabulated in a spreadsheet format.

A principal component analysis was performed on the gamut of data obtained and this was done to extract components of variations across the surface. Those components that have eigen value above one were chosen as causing the greatest variation across the surface of the study area hence could be used as classifying components across the surface.

Principal component analysis is a widely used technique for collapsing a set of interrelated variables into a smaller or same number of uncorrelated dimensions or variates. These dimensions are also described as being orthogonal because they represent perpendicular variates in the domain of the transformed set of variables.

The purpose of using principal components analysis is to collapse the parameters into smaller components.

The Varimax method of rotation was used and the components with eigen value of one and above were selected from the data. Four components were selected and their receptive eigen values are presented below:

components	eigen value	% contribution
Ph value of soil	3.0706	38.4
Sulphate	1.74574	21.8
Conductivity level	1.08813	13.6
Phosphate	1.01451	12.7

The total percentage contribution of the components is 86.5 %. The vegetation and land use quality were derived from the remote sensing data which was classified based on major land use characteristics of the study area. A supervised classification of the SPOTXS image of the study area was carried out to identify the major land use classes. See Fig 1.

Consequently different types of sensitivity index maps were produced based on these five components/ dimensions viz. pH value, sulphate, conductivity phosphate and land-use types.

Through surface interpolation and kriging method of the data sets; a total of five Initial index maps were produced based on the five parameters viz, land use (human habitat and vegetation) pH values, Sulphate level, conductivity level and phosphate level. The five index maps were integrated in a single map through map algebra using the resulting components. This is adapted approach from Ginsburg's Atlass of economic development (Harman, 1960) .The resulting index maps from the five variables are presented in figure 3(a-e).

Environmental sensitivity index is heavily based on the effect of human and natural phenomena and activities on human habitat. Though consideration is place on the effect on animal and plant kingdom the ultimate measurement is on the resultant effect on human settlement.

To this end weighting factor was introduced to rank the land use index high in the study area before map algebra was used to find the index value for the study area. The final index map is presented in Figure 4.

It should be noted however that the indexing was not based on any project in view therefore the susceptibility to disturbance was generally considered

Conclusion

The paper demonstrates the immense contribution of in-situ data captured in a GIS environment integrated with remote sensing data in a GIS environment for environmental sensitivity indexing of areas of low mapping materials or areas that have not been properly surveyed or surveyed with some form of error.

Apart form the data models been useful for ESI estimate is also a tremendous eye opener to the mapping of earth phenomena as well as map generalisation.

Implication for researches in environmental sciences

The paper shows the following especially in environmental researches.

a. The data collected from the field can be translated across the surface for generalization ad deduction.

b. Spatial relationship of data can be integrated with minimum level of error

REFERENCES.

DeMers, M.N., (2000). "Classification and Purpose in Automated Vegetation Maps." *Geographical Review,* 81(3):267-280.

Goodchild. F. M (1993)."The state of GIS Environmental Problem solving", in *Environmental Modeling with GIS* edited by M.F. Godchild. Oxford University Press. New York

Harman, H. (1960):Modern Factor analysis. Chicago. (9.4.3)

Muehcke, P.C., and J.O. Muehrcke,(1992). Map Use: *Reading Analysis, Interpretation. Madison,* W.I: J.P. Publications.

Wiler, B.O. (1990) "The U.S. Fish and Wildlife Service's National Wetlands Inventory" US. Fish and Wildlife service, Biological Report, 90(18): 9-19.

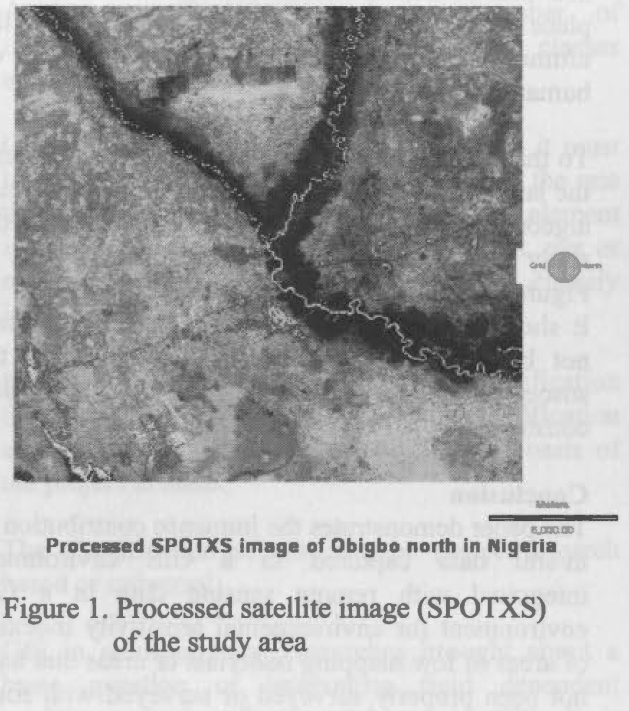

Meters
5,000.00
Processed SPOTXS image of Obigbo north in Nigeria

Figure 1. Processed satellite image (SPOTXS) of the study area

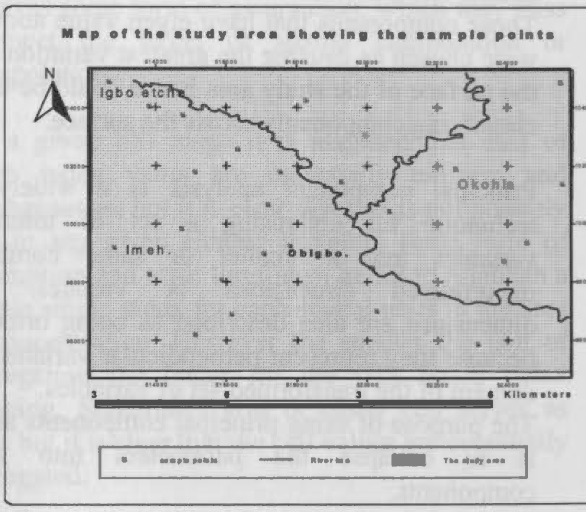
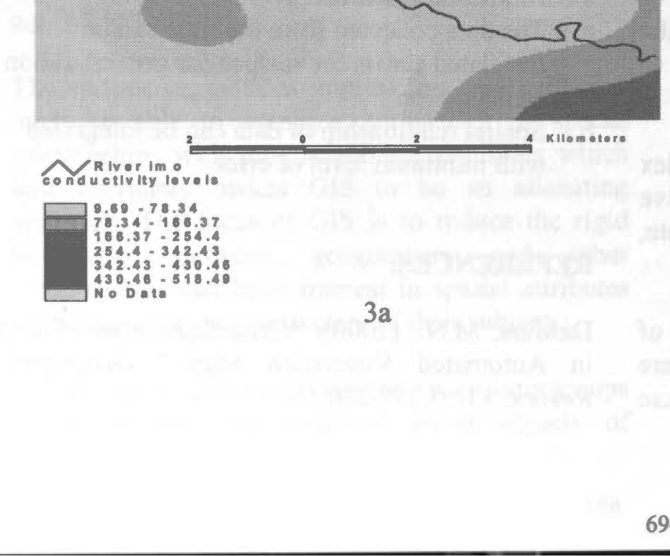
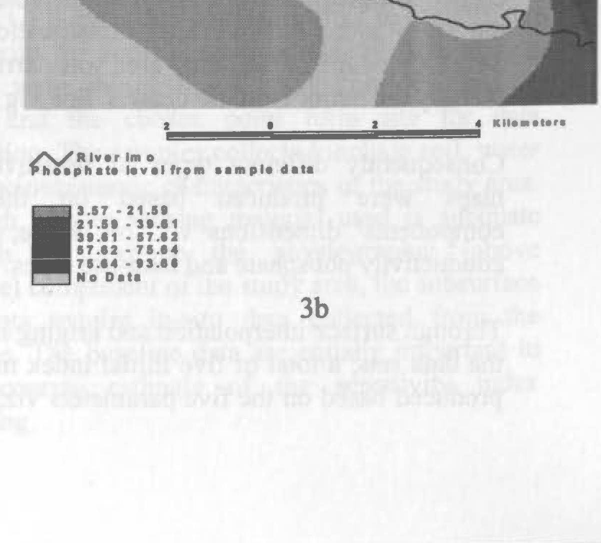

Map of the study area showing the sample points
N
Igbo etche
Okohia
Imeh.
Obigbo.
3
0
3
6 Kilometers
sample points
River line
The study area

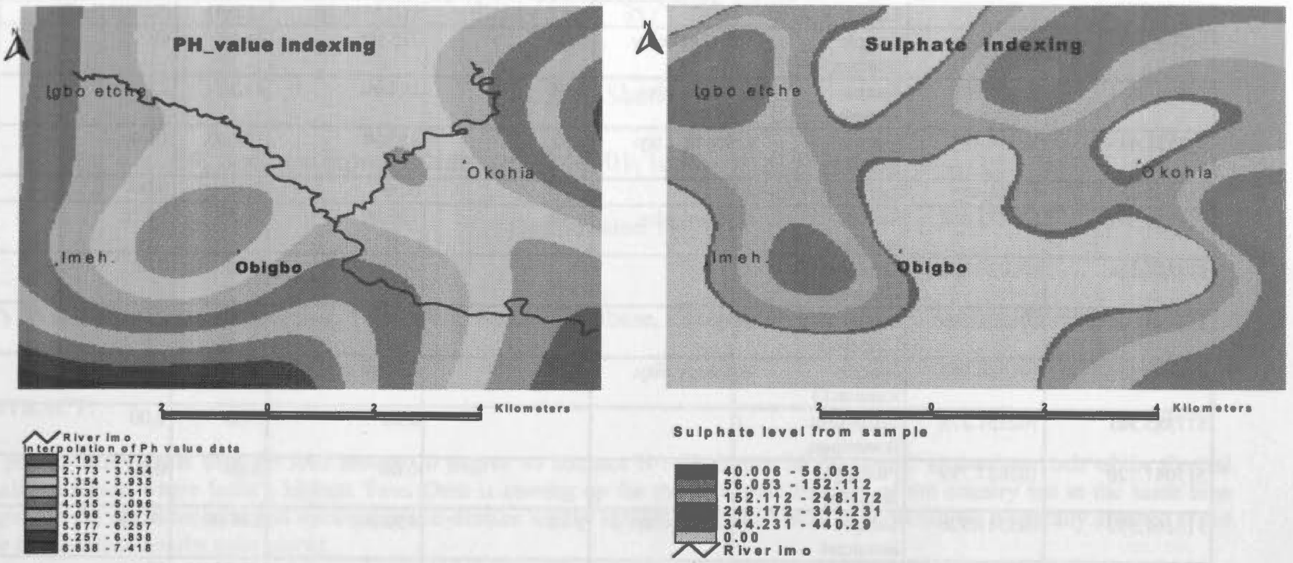
Figure 2. Location of the sample points


N
Sulphate indexing
Igbo etche
Okohia
Imeh.
Obigbo
2
0
2
4 Kilometers
River imo
conductivity levels
9.69 - 78.34
78.34 - 166.37
166.37 - 254.4
254.4 - 342.43
342.43 - 430.46
430.46 - 518.49
No Data

3a


N
Phosphate level
Igbo etche
Okohia
Imeh
Obigbo
2
0
2
4 Kilometers
River imo
Phosphate level from sample data
3.57 - 21.59
21.59 - 39.61
39.61 - 57.62
57.62 - 75.64
75.64 - 93.66
No Data

3b


PH_value indexing
Igbo etche
Okohia
Imeh
Obigbo
2 0 2 4 Kilometers
River Imo
Interpolation of Ph value data
2.193 - 2.773
2.773 - 3.354
3.354 - 3.935
3.935 - 4.515
4.515 - 5.096
5.096 - 5.677
5.677 - 6.257
6.257 - 6.838
6.838 - 7.418
Sulphate indexing
Igbo etche
Okohia
Imeh
Obigbo
2 0 2 4 Kilometers
Sulphate level from sample
40.006 - 56.053
56.053 - 152.112
152.112 - 248.172
248.172 - 344.231
344.231 - 440.29
0.00
River Imo

3c 3d

Figure 3(a-d). Index maps based on environmental components

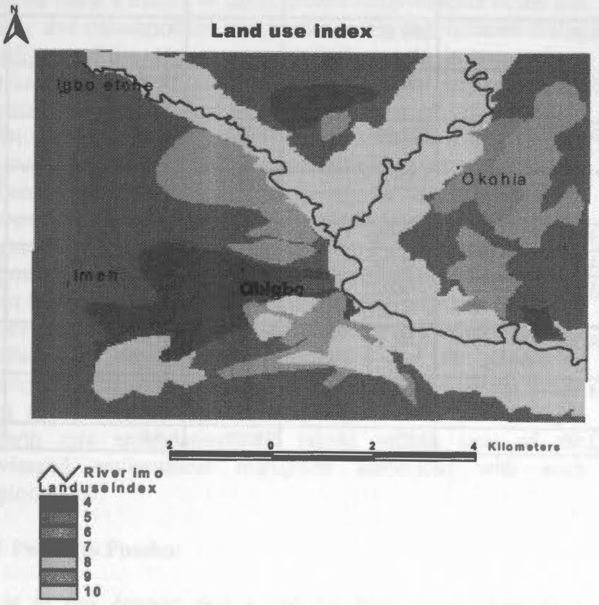

Land use index
Igbo etche
Okohia
Imeh
Obigbo
2 0 2 4 Kilometers
River Imo
Landuseindex
4
5
6
7
8
9
10

Figure 3e. Land-use index map of the study area from remote sensing data

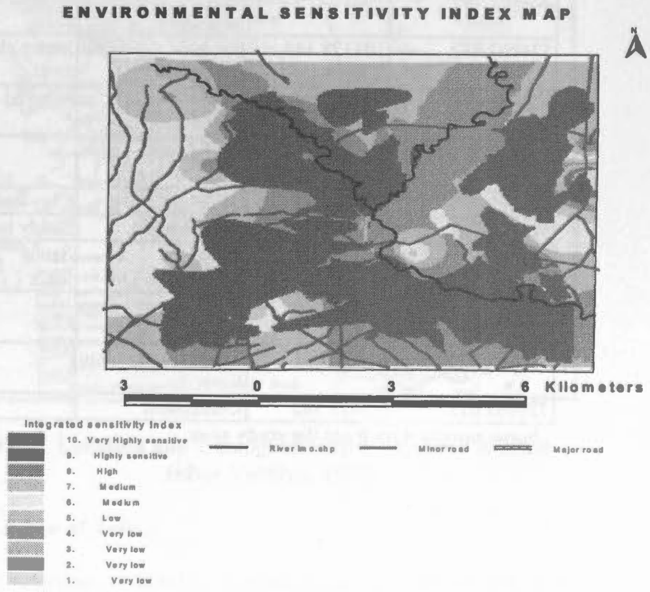

ENVIRONMENTAL SENSITIVITY INDEX MAP
3 0 3 6 Kilometers
Integrated sensitivity index
10. Very Highly sensitive
9. Highly sensitive
8. High
7. Medium
6. Medium
5. Low
4. Very low
3. Very low
2. Very low
1. Very low
River Imo.shp
Minor road
Major road

Figure 4. Final ESI map of the study area

Appendix 1

X	Y	SAMPLE_LOC	SOIL_TYPE	PH__VALUE	SULPHATE	COND	PHOSPHATE	NITRATE
517855.612	98807.829	Sample at location 8.	Loamy clay.	3.66	225.00	34.10	40.40	5.63
515401.831	98094.718	Sample at location3	loamy clay.	3.98	175.00	41.20	12.60	6.06
515521.689	98094.467	Sample at location 2	loamy clay.	3.22	325.00	138.00	17.40	6.37
516439.003	96034.403	Settlement / Iriebe.			0.00	0.00	0.00	0.00
516425.562	101573.018	Sample at location1	loamy clay.	3.99	150.00	76.30	12.10	5.63
518603.124	101221.380	Settlement / Umuebule.			0.00	0.00	0.00	0.00
518603.124	101221.380	Market at Umuebule			0.00	0.00	0.00	0.00
518363.521	101568.993	Sample at location13	loamy clay.	3.85	125.00	381.00	13.00	7.13
517385.381	102351.272	Settlement / Ikwerengo.			0.00	0.00	0.00	0.00
517041.720	103027.799	Pineapple farmland			0.00	0.00	0.00	0.00
516161.577	103511.926	Sample at location4	loamy clay.	3.98	150.00	508.00	19.40	6.07
515395.062	103817.657	Sample at location 5	loamy clay.	4.38	425.00	55.50	23.90	5.45
513468.607	99078.722	Settlement / Imeh.			0.00	0.00	0.00	0.00
517988.038	98777.452	Settlement / Oyigbo.			0.00	0.00	0.00	0.00
520544.525	96833.913	Oyigbo main market			0.00	0.00	0.00	0.00
523456.529	98539.085	Settlement/Ozuaku			0.00	0.00	0.00	0.00
520862.496	98664.078	Sample at location15	Sandy loam	3.47	125.00	133.00	19.30	7.30
524804.082	102231.773	Settlement /Okohia			0.00	0.00	0.00	0.00
524887.143	105275.743	Sample at location19	loamy clay.	4.85	250.00	85.70	70.30	7.67
524807.852	101178.145	Sample at location20	loamy clay.	5.95	150.00	217.00	30.50	6.81
523978.445	99152.409	Sample at location17	Alluvial clay .	3.56	375.00	80.20	80.20	7.07
512930.420	105315.858	Settlement/ Igbo etche			0.00	0.00	0.00	0.00
523002.390	94816.972	Sample at 14.	Clay loam.	3.54	225.00	150.00	13.80	7.71
517250.513	96736.185	Sample at 10.	Sandy loam	4.50	125.00	63.80	19.30	6.28
518363.521	101568.993	Sample 7(Quadrant).	Sandy loam loam	3.73	375.00	77.30	21.40	6.76
521963.035	97417.782	Sample at Imo River 1.		4.30	1.29	22.20	1.16	1.12
520455.038	98293.211	Sample at Imo River 2		4.50	2.97	22.30	1.18	1.34
521963.035	97417.782	Umuechem		3.50	2.27	33.20	0.86	1.11

Some sample data from the study area.

REMOTE SENSING AND GIS : AN AID TO SELECTING HABITABLE SITE IN CENTRAL HIMALAYA, INDIA

Dr. S.K. Sharma

24, National Road, Dehradun – 248001, India – sks105@rediffmail.com

Commission IV

KEY WORDS : Central Himalaya, Tehri Dam, Spatial Database, Geographic Information System, Thermatic map.

ABSTRACT:

The present study deals with the area around 30 degree 45 minutes N : 78 degree 50 minutes E in the inner belt of the Central Himalayan region where India's highest Tehri Dam is coming up for the economic benefits of the country but at the same time bringing in the environmental and socioeconomic stresses for the upstream people whose lives would be drastically affected if not made to migrate to nearby safer places.

It is in this context that the spatial database created through remote sensing information is converted into computer readable form using the Geographic Information System (GIS) for analyzing the geographically referenced desets of spatial and non-spatial origin. The recently acquired satellite images on 1:50,000 scale have been used to prepare thematic maps depicting the vegetation, watershed, soil cover, geomorphological features, drainage pattern and watershed areas which has helped when integrated with the Intergraph GIS system in distinguishing the localities, in an otherwise difficult terrain, suitable for variety of plantation/activities of economic value where upstream people can be made to migrate without impairing or unbalancing ecosystem.

1. INTRODUCTION

Despite India's extensive development achievements in the last nearly five decades since independence, the fact remains that is is still one of the poorest countries in the world, in terms of per capita income and consumption. The per capita energy consumption in India is about 0.25 tonnes of oil equivalent (TOE) as against 8.5 TOE of developed nation. If the standard of living has to be raised, an accelerated growth of economy becomes imperative which can greatly be facilitated through increased availability of energy, be a hydrocarbon or electricity. Fortunately, India has vast unexploited hydropower potential, estimated at about 84,000 MW which if properly harnessed can meet the domestic demand of ever-growing population and can relieve the country from the heavy burden of foreign exchange outflow and dependence on imported fuel for generating electricity. Therefore, the country has very little option but to step up its energy exploitation programme at a very rapid growth rate anticipating and taking utmost care of the envisaged environment disruption associated with such exploitation.

1.1 Points to Ponder

It is in this context that a 260.5m high Tehri Dam on a Himalayan tributary of the Ganga, the Bhagirathi river in the Central Himalayan Region is coming up (Fig.-1) for economic benefit of the country and of the local people in particular to delivery nearly 250 MW of electricity when completed.

But these economic benefits to the country are full of miseries to the local residents comparing nearly hundred villages around the Tehri Dam since there only source of livelihood this "land" is going to be submerged under its reservoir water. It is estimated that nearly 50,000 people have lost their homes and lands to reservoir as part of the worlds larges dam building programme. This has not only uprooted the local residents but has also adversely affected the agriculture produce, the only means of their livelihood.

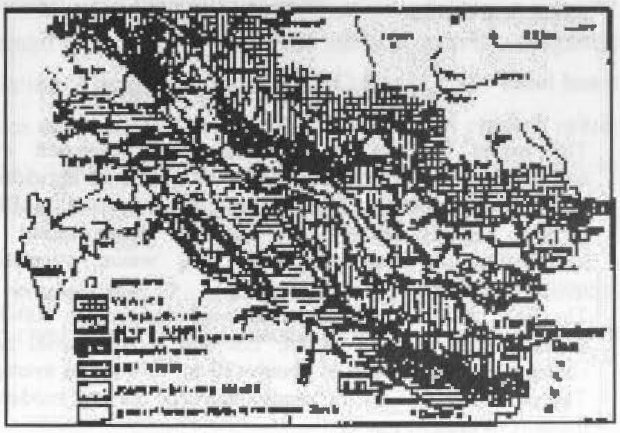


Fig. 1 Location and Geological map of the area studied (after Valdiya, 1997)

1.2 Force of Action

The selection of suitable alternatives site for the rehabilitation of uprooted native people forms the basis of present study.

2. ENVIRONMENTAL IMOPACTS ENVISAGED

The study area whose geographical extents are shown in figure 1 lies around $30^0 45'$N and $78^0 50'$E consisting of an area of about 2500 Km in the inner belt of Central Himalayan Region. Social and environmental impacts of this large dam have been a cause of concern over the last one decade as the dam is in the proximity of an active Srinagar Thrust (Prasad and Rawat, 1986). The evidence suggested that the elites living far from the dam site will enjoy the benefits of the irrigation and power, it is the indigenous people who with low social status and little potential power will suffer the negative effects of dam. In addition to the anticipated increase in the incidence of water-borne disease such as the debilitating schistosomiasis (nilharzias), malaria and onchocerciasis (river blindness). Other potential costs which can be triggered by the construction of this dam in the geologically sensitive area are the economic costs of submergence of valuable resources such as fertile land, lumber and medicinal plants and the loss of hardwoods and biological resources. Therefore, while selecting the alternative site nearby for the displaced people, the scientific knowledge of the soil, its stability and the ecological balance of the selected site will permit the appropriate eco-friendly use of the land.

3. INVESTIGATION PROCEDURE

The published topographic cultural/physiographic map by the Survey of India at 1 : 50000 scale is used to prepare the base map and the landset TM data obtained from IRS LISS II of November 11, 1993 is being used as remote sensing output. The geological map of the area is digitized at original mapping scale of 1 : 50,000 and the areas if misties have been solved using remote sensing data. The maps have been generated in Polyconic projection system having Everest Ellipsoid for $30^{0}45'$N Latitude and $78^{0}50'$E Longitude.

Such special data maps in the form of topography, hydrology, geology, soil types, forest/land cover stored as layers in digital form in the computer. And finally, the presentation of results based on the outcome of comparing/overlaying the spatial data with different sets of features, their distribution in space and actual field checks.

4. RESULTS AND DISCUSSIONS

The Indian Space Programme has been developed with multifacet vision for using space technology and the agriculture being the foremost. The Indian remote Sensing Satellite (IRS) picture LISS-II offers the valuable dissemination of information on improved land, soil, water, vegetation utilization capabilities.

The field checks show the variation of slopes from 100-300 m/km having southern aspect. The soil is of alluvial type containing soil moisture of about <20 to 40% on an average. The soil is of clayey to sandy loamtype having moderate quantity of organic matter.

On the basis of false colour composite (FCC), figure 2 has been generated from the digitized data of Derhadun and is surroundings covering an area of about 2500 Km between the two major river the Yamuna in the west and the Ganga in the east, in order to demarcate various themes distinguishable from different colours. The red colour covering and area of about 1500 Km^2 shows the forest envelope at higher altitudes (more than 2300m above sea level marked with A and B) which is vital for man, plant and animals, fulfilling the basic needs of fuel fodder, manure, medicines and raw materials for industries. The plant area (marked with X,Y and Z), the yellow colour depicts the agricultural area of about 700 Km^2 where the soil is very fertile having the moderate organic content. The green and orange colour together shows the plantation of various kind covering an area of about 350 Km^2. The Urban area is very less and the area covered by the water is about 20 Km^2 shown in blue colour.

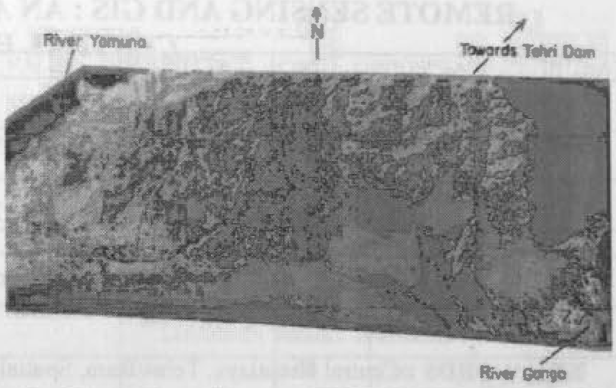


Fig. 2 Land Cover Themes – IRS LISS-II November 1993

5. CONCLUSION

Therefore, based on the latest available remote sensing information on water and land resources, integrated with the GIS provides an insight for planning the rehabilitation of about 50,000 uprooted people from Upstream which is about 60 Km North-East of Dehradun city, can be thought of, without endangering the existing environment.

6. ACKNOWLEDGEMENTS

The author is grateful to various local agencies who have helped him during field checks. Several useful discussions were held with the Staff of Garhwal vikas Mandal, Derhadun during the preparation of this paper. The help rendered by the staff is thankfully acknowledged.

7. REFERENCES

Pearce, F (1996) Superdams – have their altractions been oversheated? *Journal of Geosciences and development* No. 3, pp 7-10

Prasad, C. and Rawat, G.S. (1986) Studies in recent reactivation along the Srinagar Thrust. *Garhwal Himalaya Proc. Inter. Symp.* Neotectonics in South Asia, Survey of India, Dehradun pp 301-312.

Saxena, P.B., Bishnoi, Antium, Bhatia, D.K. and Saxena Bhuwan (1998) Investigations of eco-restorations of the Asarori-Mussoorie water divided to twin water sheds (Suswa-Asan) *in the Derhadun Jr. Himal. Inst. of Action Research & Development*, pp 19-29.

Tiwari, R.S. and Garg, P.K. (1995) Land form and Land uses of Doon Valley and their role in resource development. *Jr. Himalayan Geology,* Vil. 6(2), pp 9-15.

Valdiya, K.S. (1997) High Dams in Central Himalaya in context of active faults seismicity and societal problems. *Jr. Geol. Soc. of India*, vol. 49, pp 479-497.

PETROLEUM AND GAS RESEACH BY REMOTE SENSING

IN SOUTH CHINA SEA

Prof. Xu Ruisong

Guangzhou Institute of Geochemistry, Chinese Academy of Sciences
P.O.Box 1131, Guangzhou, China, 510640
E-mail: xurs@gig.ac.cn

KEYWORDS: South China Sea, Petroleum and Gas Anomaly, Remote Sensing Research, Correlation Analysis

ABSTRACT:

This study have been made SST processing by NOAA-11 CH4,5 band, false color image processing by NOAA-11 CH5,2,1 band, and more factor correlation analysis between gray value and thermal current value i.e. data in South China Sea, with computer image processing system, after weed cloud, atmospheric modified, and geometry rectification. The remote sensing data is from NOAA-11 in May 16-30, 1990, March 15-April 2, 1993 and April 12, 1994. The main study results as follow:

1. The temperature of sea surface is 1-3 ^{o}C higher than background value in petroleum and gas area of Yingo sea, Pearl river mouth, Taiwan strait, and Lilou beach in South China Sea.
2. The hydrocarbonphilous microorganism, algae, and plankton on Yingo sea, Pearl river mouth, Taiwan strait, Luzon sea, Lilou beach, and Honghe river mouth petroleum and gas area have been displayed special cherry red color on NOAA-11 CH5,2,1 band false color image different from background color.
3. The gray value with NOAA-11 CH1,4,5 band, temperature on South China Sea surface, gravity, magnetic anomaly and thermal current value have the best correlation. The gray value from NOAA-11 CH1 band has a positive correlation with temperature on South China Sea surface, gravity anomaly and thermal current value; has a negative correlation with magnetic anomaly. Gray value from NOAA-11 CH4,5 band has a positive correlation with magnetic anomaly; has a negative correlation with temperature on South China Sea surface, gravity anomaly and thermal current value.
4. According to research results, a few petroleum and gas prospect areas have been identified rapidly, economical and accurately in Luzon sea, Honghe river mouth, Pearl river mouth and Lilou beach i.e. areas on South China Sea.

1. INTRODUCTION

Since space age in 1960, remote sensing information have been serviced for ocean shipping, marine meteorology and marine engineering. World would going to marine remote sensing century after been launched the first ocean research satellite by U. S. A. in 1970. People study use marine remote sensing in areas of marine Physics, chemistry, meteorology, ecology, geology, environment, pollution and resources i.e.. In global change research, the remote sensing information that from different satellite station and more band have been wide for study in global marine environment, primary products, resources, water color, carbon amount, carbon circle, water and steam change over world, quantitative, automation, and all weather i.e.(Xu Ruisong et al. 1990, 1992,1993,1997).

Land provide limited space for humanity life. So, must study more nature rule for need more space and resources from marine. But marine too large, change fast at short time, just remote sensing is the best way for marine resource and environment research quickly, economy, quantitative, automation and all weather.

South China Sea located in S 3°-N 25°, E 98°-123°. It is the largest marginal sea in West Pacific and the biggest tropics sea basin in the world(Fig.1). The area is more than 360Km^2. Geomorphology consist with oceanic basin, trough, trench, continental slope, continental shelf, and Island. South China Sea basin is a rhombus one that long axle is direct to NE-SW. The water of basin is over 3600M deep. The marine trents are NE display, the deepest one is over 5559M. The continental shelf near by Chinese main land, India-China Peninsula, Indonesia archipelago is the smoothest continental in the world. The widest one is 285Km.The deepest one is 150m.The continental near by Philipines is narrow, high and precipitous. Water of continental slope is 150-3000m deep, the slope is 1-2 degree, the widest is 555Km. South China Sea is from tropics to subtropics, it has a striking tropical marine meteorology. There are rich regenerated and nonregenerated resources in South China Sea, include more than 500 Thousands Km^2 Petroleum and gas basin, the reserves is more than 150 billion barrels. Today, there are many petroleum companies to exploration and developed petroleum and gas in continental shelf around South China Sea. Because there are wide and far from main land in petroleum and gas basin of South China Sea, so take the greatest cost, get bear fruit from oceanic petroleum and gas exploration. This study stress the essentials problem, that are remote sensing theory, technology, and application model for oceanic petroleum and gas exploration. The aim is provide judging by oceanic petroleum and gas exploration through research and theory analysis.

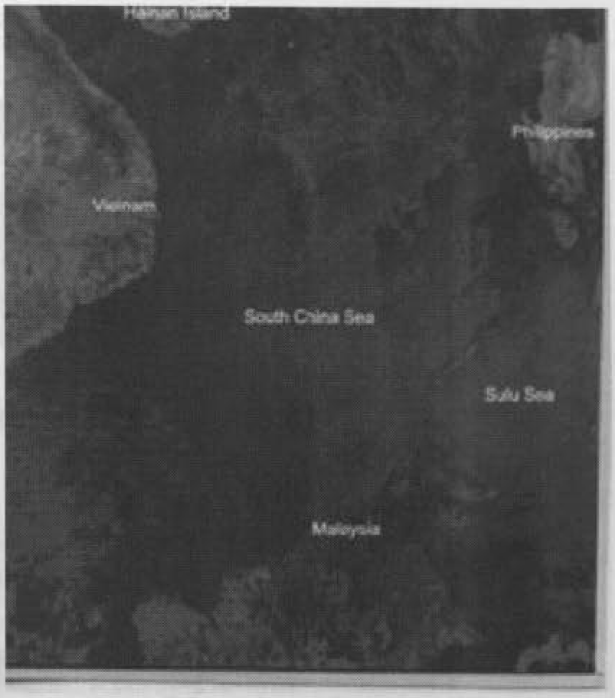


Fig. 1. NOAA-11 false color image and study area of South China Sea. Data is from NOAA-11 of 1993,3,15 to 1993,4,2. The false color image has been analysis with computer image processing system after weed cloud, atmospheric modified, and geometry rectification from NOAA-11 CH1, 2, 3, 4, 5. Red color is vegetation on land in image, yellow and green color is bear land, gray green color is cloud, light gray color is Pacific water in East South China Sea, blue gray color is India Oceanic water on West South China Sea, deep blue color is the stay long time water on middle South China Sea.

2. DATA RESEARCH AND ANALYSIS

2.1. Relationship Between Surfer Temperature and Marine Petroleum and Gas of South China Sea

This study had token NOAA-11 digit data from 1990,5,16 to 1993,3,15-4,2 on South China Sea for SST analysis with CH4, 5, after weed cloud, atmospheric modified, and geometry rectification by computer image processing system. Results display in Fig.2, 3, 4. SST model is:

$$T_s=C_1+C_2T_b+C_3T_b^2$$

T_s-research temperature on sea surfer, T_b-gray value of NOAA-11 thermal band, C-correlation coefficient. Temperature error is 1 °C in low latitude area by SST model.

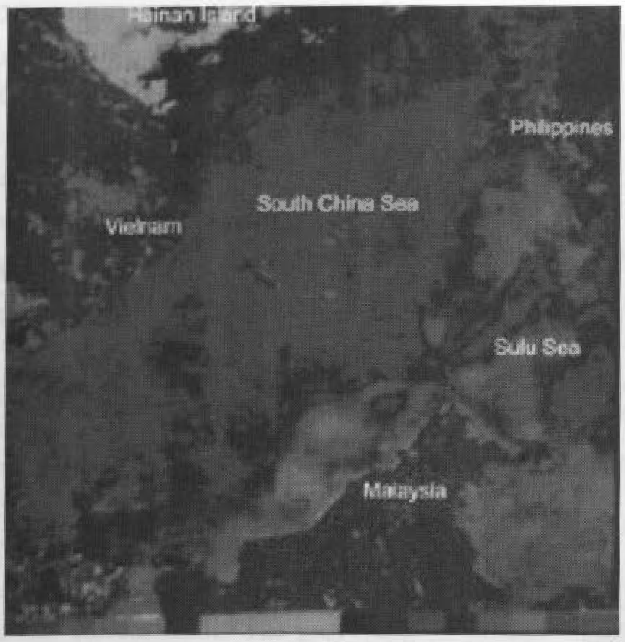

Hainan Island
Philippines
South China Sea
Vietnam
Sulu Sea
Malaysia

Fig.2. NOAA-11 false color image after SST analysis in South China Sea. Data is from NOAA-11 CH4,5 in 1993,3,15 to 4,2. Blue green color is 23°C. Red color is 33°C. It is 0.5^{O}C every color hue.

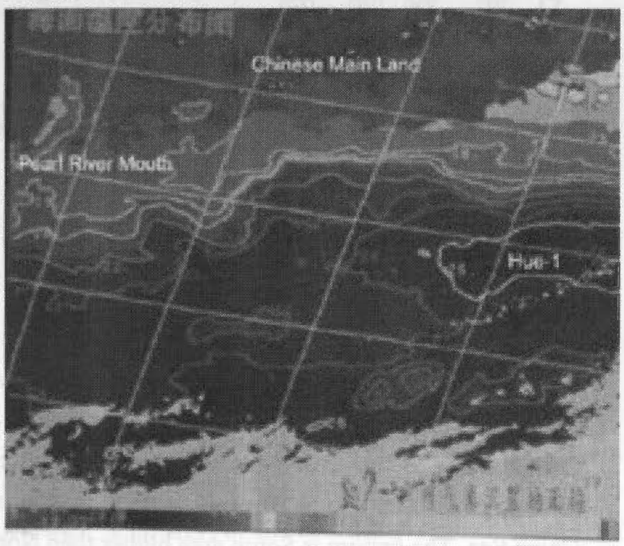

Chinese Main Land
Pearl River Mouth
Hui-1

Fig.4. NOAA-11 SST analysis image in Pearl river mouth to Taiwan strait of South China Sea. Data is from NOAA-11 CH4,5 in 1990,5,30. More than 25 is Huezhou-1 petroleum and gas deposit in image. More than 24 is petroleum and gas prospect area.

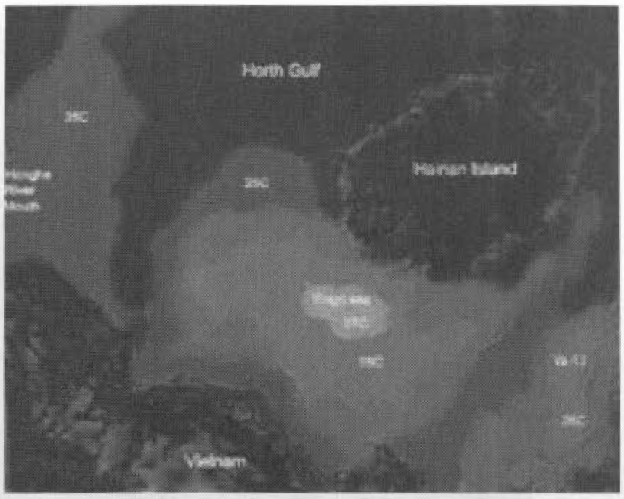

North Gulf
Hainan Island
Vietnam

Fig.3. NOAA-11 SST analysis image in North gulf of South China Sea. Data is from NOAA-11 CH4,5 in 1990,5,16 to 5,30. More than 27°C is Yingo sea petroleum and gas deposit in image. More than 26°C is Ya-13 petroleum and gas deposit prospect area.

2.2. Hydrocarbonphilous marine biology data research and analysis

According to trace percolation principle, there are rich hydrocarbonphilous microoganism, algae and plankton on sea water that is rich hydrocarbon. This study made out false color image with a line stretch from NOAA-11 CH5,2,1 in 1994,4,12. Hydrocarbonphilous marine biology display special cherry red color on NOAA-11 false image different from background sea water(Fig.5,6).

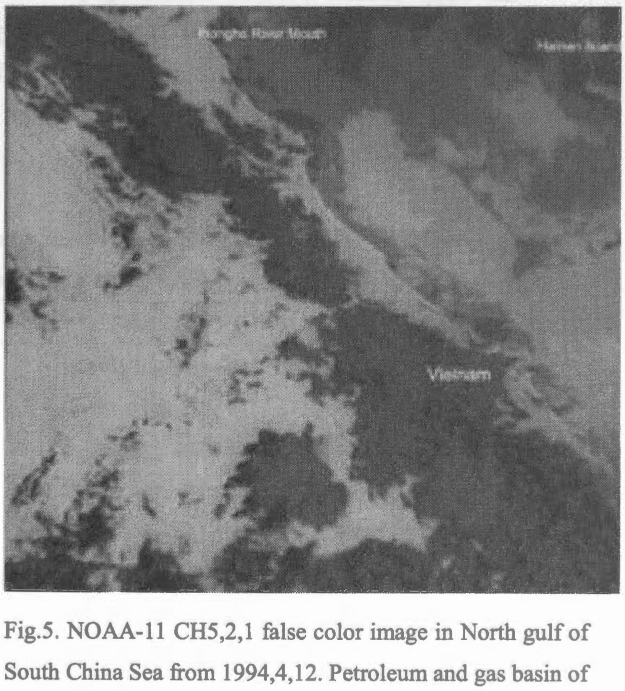

Honghe River Mouth
Hainan Island
Vietnam

Fig.5. NOAA-11 CH5,2,1 false color image in North gulf of South China Sea from 1994,4,12. Petroleum and gas basin of Yingo sea and Honghe river mouth is displayed deep yellow color in image. Hydrocarbonphilous marine biology is displayed special cherry red color line and block.

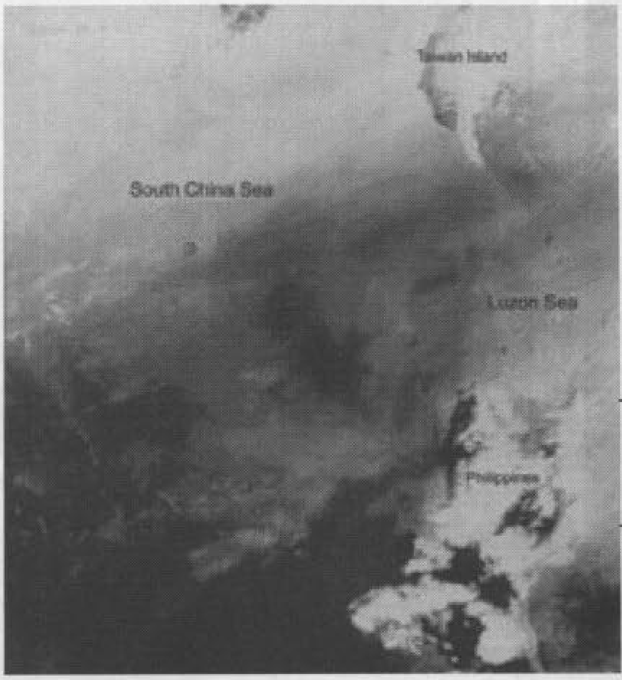


Fig.6. NOAA-11 CH5,2,1 false color image in Luzon sea from 1994,4,12. Hydrocarbonphilous marine biology is displayed special cherry red color in image.

2.3. More Factor Correlation Analysis Between Petroleum and Gas Research and NOAA-11 digit Data in Nansha sea area of South China Sea

This study select NOAA-11 CH1,4,5 gray value data, petroleum and gas research data at 17 same sample points from Nansha sea area of South China Sea in 1993,3,15-4,2 with more factor correlation analysis(Table 1). The study results at Table 2.

Table 1. Data of NOAA-11 and petroleum and gas exploration from Nansha sea area of South China Sea.

No	Longitude	Latitude	T	CH1	CH4	CH5	Gr	Ma	Th.
1.	112°20′	10°40′	26.30	6	23	21	-10	25	120
2.	113°30′	10°00′	26.50	5	23	22	30	-60	60
3.	115°30′	9°13′	26.70	4	24	24	10	-30	60
4.	112°50′	7°00′	26.90	6	22	20	30	20	60
5.	109°40′	7°30′	26.80	10	22	23	60	0	60
6.	109°40′	6°35′	26.90	9	19	18	0	-30	120
7.	110°30′	5°45′	26.20	8	20	18	30	20	90
8.	110°40′	5°35′	27.40	6	20	18	40	20	90
9.	111°25′	4°50′	27.30	7	21	19	38	-20	90
10.	109°45′	4°30′	27.20	8	23	21	30	10	120
11.	108°40′	5°40′	26.50	7	23	22	28	0	60
12.	116°30′	12°00′	26.70	6	23	21	90	-50	60
13.	116°30′	10°30′	26.60	4	24	23	-10	100	60
14.	111°10′	8°45′	27.00	6	21	19	0	-70	60
15.	111°30′	4°30′	27.30	6	20	18	30	-30	100
16.	108°35′	8°00′	27.00	7	20	17	30	-50	80
17.	112°35′	7°20′	26.00	6	22	20	20	-50	60

Note: T-temperature(°C). CH1,CH4,CH5-NOAA-11 CH1,4,5 gray value. Gr-gravity anomaly(mgal). Ma-magnetic anomaly(nT). Th-thermal current value(mw/m^2).

Table 2. The results of more factor correlation analysis

		Y	X1	X2	X3	X4	X5	X6
		T	CH1	CH4	CH5	Gr	Ma	Th
		°C	0-255	0-255	0-255	mgal	nT	mn/m^2
	MV	26.78	6.53	21.76	20.24	26.24	-17.35	81.76
	EM	1.61	0.634	6.25	8.43	9.963	1.895	9.485
	Y	1						
	X1	.133	1					
	X2	-.384	-502	1				
C	X3	.355	-.301	.910	1			
	X4	.202	.300	-.040	-.020	1		
	X5	.040	-.103	.131	.122	-.253	1	
	X6	.245	.422	-.359	-.390	-.243	.085	1

Note: MV-mean square value. EM-error of mean square. C-correlation coefficient. T,CH1,4,5,Gr,Ma,and Th-same as Table 1.

3.RESULTS AND DISCUSSION

3.1.Relation between sea surface Temperature and Petroleum and Gas Basin of South China Sea

On the basis of this study, sea surface temperature is different from South to North and East to West in South China Sea. The feature is South area higher than North, East area higher than West. The sea surface temperature different value is high to 10°C from South to North and East to West of South China Sea. Study results show, sea surface temperature can direct sea petroleum and gas exploration. For example, sea surface temperature of Yingo sea petroleum and gas basin is 1-2°C higher than background (Fig.3), Pearl river mouth and Taiwan strait is 2-3°C higher than background, (Fig.4), Lilou beach of Nansha sea, South China Sea is 2°C higher than background (Fig.2).

3.2.Remote sensing study of Hydrocarbonphilous marine biology

NOAA-11 CH 2.1 false color show, there are rich hydrocarbonphilous marine biology in North gulf and Luzon sea Petroleum and gas basins, south China Sea, and special cherry red color in petroleum and gas basins different from background (Fig.5.6).

3.3. More Factor Correlation Study

Modal of more factor correlation analysis is:

$$Y=47.65+0.23X_1-0.0918X_2-0.0181X_3-0.0435X_4-0.35X_5-0.464X_6$$

The mean of Y, X_{1-6} is same as Table 2. F is 0.5911.According to study results, sea surface temperature and NOAA-11 CH 1.4.5 gray value is the best correlation with gravity, magnetic anomaly and thermal current value. The gray value of NOAA-11 CH1 is positive correlation with temperature on South China Sea surface, graving anomaly and thermal current value; have a negative correlation with magnetic anomaly.

The gray value of NOAA-11 CH 4.5 is positive correlation with magnetic anomaly; have a negative correlation with temperature on South China Sea surface, gravity anomaly and thermal current value (Table 2).

3.4. On the basis of research results, a few petroleum and gas prospect areas have been identified rapidly, economical and accurately in Luzon sea, Honghe river mouth, Pearl river mouth, and Lilou beach i.e. areas on south China Sea.

3.5.This study results need more marine petroleum and gas exploration data to test.

REFERENCES

Xu Ruisong, 1990, Geobotanicall remote sensing in China, Proceedings of Remote sensing in the Minnig and petroleum Indstries, Landon.

Xu Ruisong, 1992, Cooperative developing and research in basin of South China Sea, Remote sensing information.

Xu Ruisong, 1992 Remote sensing study of Gold Biogechemical Effects is the Western. Guangdong-Hainan, Region, Acta Geological sinica, 5(4),411-425.

XU Ruisong, Remote sensing situation and trend in the world, Technology and application of remote sensing, 5, PP. 38-41.

Xu Ruisong, 1997, A Preliminarg study on chlorophyll remote sensing in Nansha Islands sea area, Science Press, PP. 133-142.

ACKNOWLEDGEMENTS:

I thank Chinese Academy of sciences, south China sea institute of Chinese Academy of and Chinese Meteorological Bureau who give me good service for my project.

PROJECTING GLOBAL RASTER DATABASES

E. Lynn Usery, Michael P. Finn, Douglas J. Scheidt

U.S. Geological Survey
Rolla, Missouri 65401

Phone: 573-308-3837; Fax: 573-308-3652
usery@usgs.gov; mfinn@usgs.gov; dscheidt@usgs.gov

KEY WORDS Global-Environmental-Databases, Decision Support, Transformation, Accuracy, Cartography, GIS

ABSTRACT

The need for regional and global models for spatially distributed phenomena and processes has driven the development of large raster databases of terrain elevation, population, land cover, vegetation, soils, temperature, and rainfall. The use of commercial geographic information system software to process these databases requires projection from spherical coordinates to plane coordinate systems and transformation from one plane system to another. Recent research has demonstrated that such projection and transformation of raster databases is error-prone and may generate data that corrupt the results from spatial models. This paper presents two approaches to handle the problems of raster database projection and transformation. The first is the development of a Web-based decision support system to aid users with the selection of appropriate projections based on data type, areal extent, location, and properties to be preserved. The second is a dynamic approach to projecting raster databases that adjusts the output pixel size according to the exact area of the corresponding pixel in the spherical coordinate system.

1. INTRODUCTION

Modeling regional and global activities of climatic and human-induced change requires accurate geographic data from which we can develop mathematical and statistical tabulations of attributes and properties of the environment. Many of these models depend on data formatted as raster cells or matrices of pixel values. Recently, it has been demonstrated that regional and global raster datasets are subject to significant error from mathematical projection and that these errors are of such magnitude that model results may be jeopardized (Steinwand *et al.*, 1995; Yang *et al.*, 1996; Usery and Seong, 2001; Seong and Usery, 2001). There is a need to develop methods of projection that maintain the accuracy of these datasets to support regional and global analyses and modeling.

Although recent research indicates that projection problems exist for raster databases at global and regional scales, there is little theoretical background for handling the relationships between the distortions that are due to projection methods. Also, there is little guidance for users of geographic information systems (GIS) and other software packages that offer the capability to project these regional and global raster datasets (Usery *et al*, 2001). In many cases, the software includes many projections, each projection with its own set of parameters and, according to projection theory, with its own particular strengths and appropriate areas of use. With this plethora of

Disclaimer: Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

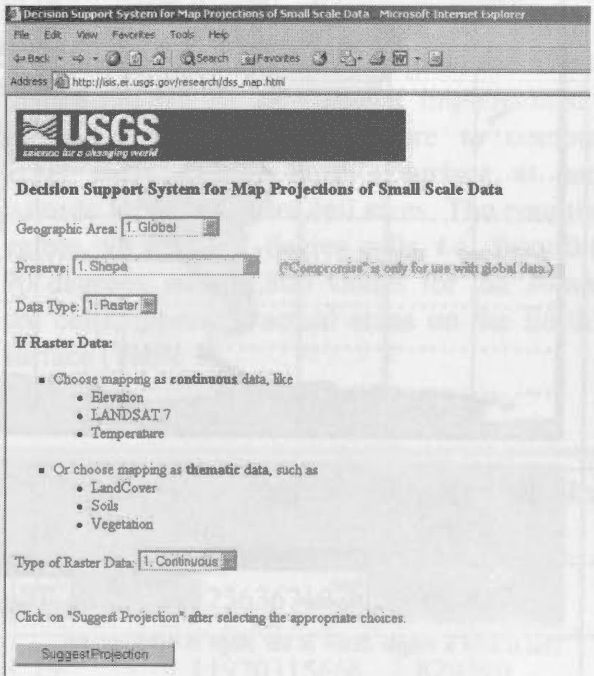
projection choices and the ability to customize each projection, most GIS users do not have the knowledge to select a projection appropriate to the area of the world, data type, or particular problem being examined. Couple these selection problems with the inherent errors found in commercial software packages for projection transformation, many of which were designed for small datasets that span only a few hundred kilometers of the Earth's surface, and the scenario is ripe for error to occur and erroneous model results to be generated. Most users will not know that an error has occurred or that their model, while good, is generating erroneous results because of the inaccurate transformation of data.

2. APPROACH

Our research has been accomplished in three stages: 1) empirical investigation of errors resulting from the projection of regional and global raster datasets, 2) development of a decision support system (DSS) to guide users in the selection of an appropriate projection, including a tutorial, and 3) new projection methods that result in less error when used with large regional and global raster datasets (Usery *et al.*, 2001). The empirical investigation has examined 10 different projections and 6 global raster datasets at various resolutions. The results indicate that variances in projection accuracy depend on the projection selected, the type of data (*i.e.*, categorical or continuous), the number and extent of the categories, the resolution of the output dataset, and latitude. Results of this empirical investigation are being used in the second and third stages of our research.

3. THE DSS

In the second stage of our research, we are developing a DSS for selecting an optimum projection considering various factors, such as pixel sizes, areal extent, number of categories, spatial pattern of categories, resampling methods, and error correction methods. The design and implementation is a Web-based DSS for map projection selection. We designed the DSS for map projections of small-scale datasets. The initial prototype is currently in a test phase. The design includes an interactive interface to solicit information from a user and guide the user to the selection of an appropriate projection (or alternate projections) based on the input (Figure 1).


Decision Support System for Map Projections of Small Scale Data - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address
USGS
science for a changing world
Suggest Projection

Address: http://isis.er.usgs.gov/research/dss_map.html

Decision Support System for Map Projections of Small Scale Data

Geographic Area: 1. Global

Preserve: 1. Shape ("Compromise" is only for use with global data.)

Data Type: 1. Raster

If Raster Data:

- Choose mapping as **continuous** data, like
  - Elevation
  - LANDSAT 7
  - Temperature
- Or choose mapping as **thematic** data, such as
  - LandCover
  - Soils
  - Vegetation

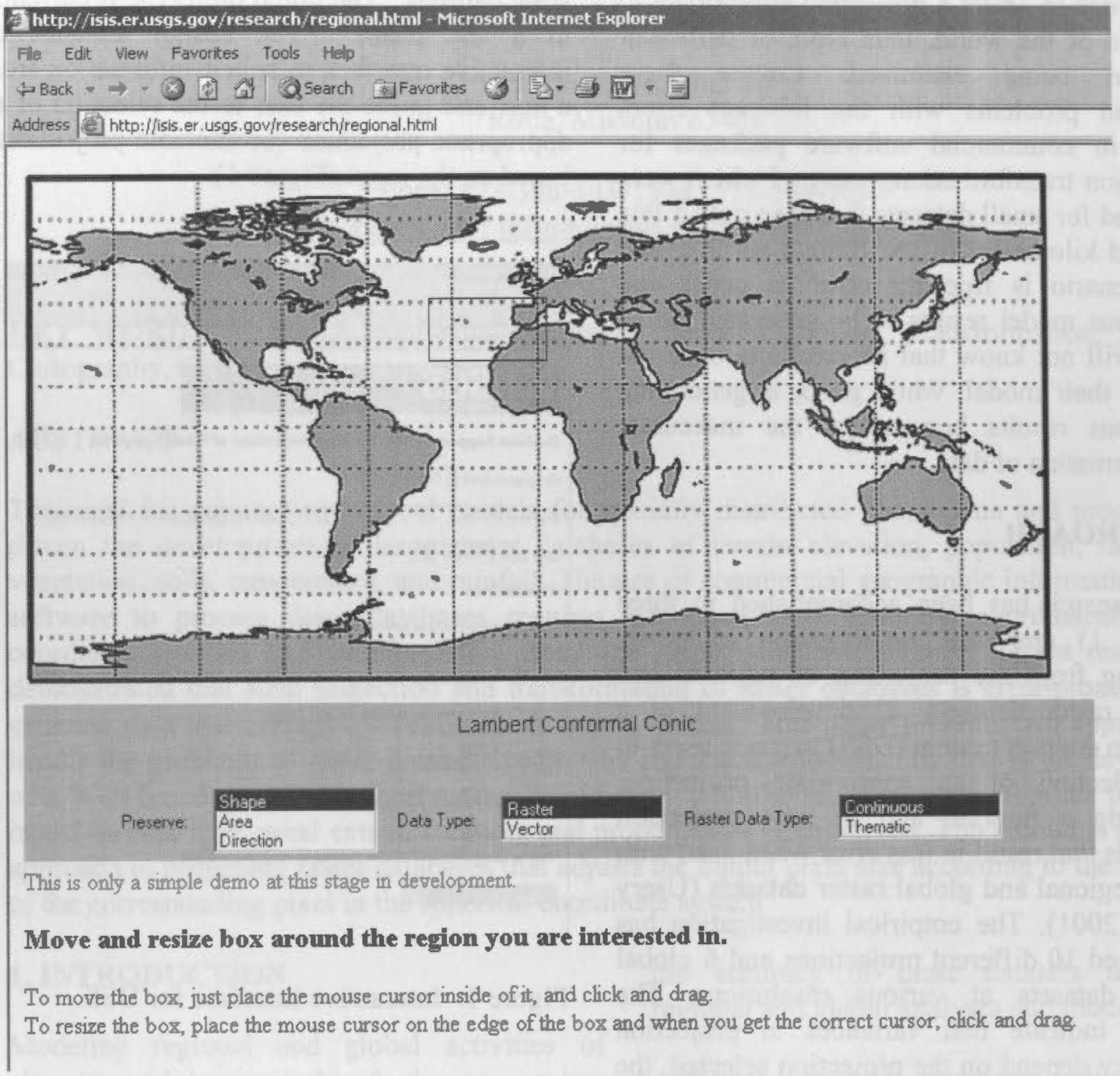
Type of Raster Data: 1. Continuous

Click on "Suggest Projection" after selecting the appropriate choices.

Figure 1. Interactive Interface for DSS.

The DSS poses questions interactively, in which the choice of the user determines the path in a decision tree to drive to a solution. The initial choice is among global, continental, or regional area of coverage. The second important choice is among the preservation of shape or area, or simply a compromise (such as the Robinson projection, which is neither conformal or equal-area, yet has nice visual appeal).

The system handles continents on the basis of their geographic location and directional extent. For example, the DSS handles continents with greater longitudinal range (Asia, Australia, and Europe) differently than those with greater


http://isis.er.usgs.gov/research/regional.html - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Address
http://isis.er.usgs.gov/research/regional.html

Lambert Conformal Conic

Preserve: Shape (selected) / Area / Direction

Data Type: Raster (selected) / Vector

Raster Data Type: Continuous (selected) / Thematic

This is only a simple demo at this stage in development.

Move and resize box around the region you are interested in.

To move the box, just place the mouse cursor inside of it, and click and drag.

To resize the box, place the mouse cursor on the edge of the box and when you get the correct cursor, click and drag.

Figure 2. Regional Area Selection Interface.

latitudinal range (Africa, North America, South America) or polar location (Antarctica). For regional cases, the DSS allows the user to interactively define an area of interest. It also recommends a projection in real time. Figure 2 shows the interface for regional area selection.In recommending a map projection for small-scale regional data, the system factors in geographic location and directional extent in addition to the user's choice of conformal or equal-area projection.

Somewhat similar to the design described by Jankowski and Nyerges (1989), further implementation of the DSS will expand the user input to include data type, volume or resolution, and other appropriate options. Refinements planned for the current prototype DSS that will constitute "Version 0" include other choices based on the raster versus vector data model and thematic versus continuous data type. While the DSS relies on conventional theory for the selection of projections (Voxland, 1978; Snyder, 1987; Snyder and Voxland, 1989), the design includes a specific decision tree, which handles regional and global raster projection. Decisions in this branch of the tree are based on results from our empirical research, which has examined many different projections with a variety of data types. The objective is to develop a "Version Beta 1" DSS that will include an integrated, Web-based map projections tutorial (Usery *et al.*, 2001).

The following describes the system flow of the DSS:

1. A user opens up the Web page in his or her browser.
2. The user selects among choices from the given fields on the Web page.
3. Upon submittal by the user, the DSS sends the responses in the fields to the server.
4. The Common Gateway Interface (CGI) program, written in Perl, processes the responses.
5. On the basis of the data, the CGI will then:
   a. Provide the appropriate projection for GLOBAL data.
   b. Create a new page displaying a clickable image of the world for CONTINENTAL data, then pass the user's choice of continent, along with the user's original responses, to another CGI (that acts much in the same way as the first CGI) and provide the appropriate projection.
   c. Send the user to a page that is running a Java applet that will interactively determine the appropriate projection using the geographic location and directional extent of the REGIONAL area that the user provides.[1]

4. PROJECTING EQUAL AREAS

The third stage of our work is to develop new projection methods for raster data that correct the transformation problems. Our approach includes dynamic projection, error theory models, and new resampling methods. This paper will discuss only the dynamic projection aspects, but an overview of our entire approach is available in Usery *et al.* (2001).

We are developing a dynamic projection method that adjusts projection methods for latitude on the basis of raster cell size to maintain cell sizes that match in area to the corresponding cells in geographic coordinates. We have computed the exact areas of raster pixels from geographic coordinates to establish a base from which to determine projection accuracies. Using raster cells in spherical coordinates of 1 degree by 1 degree, and also of 30 arc-sec by 30 arc-sec, we implemented a numerical integration procedure to compute exact areas on the Earth's surface at each latitude for these raster cell sizes. The resulting values, 90 for the 1-degree cells, *i.e.*, from 0 to 90 degrees, and 10,800 values for the 30-arc-sec cells, represent actual areas on the Earth's surface (Table 1).

Latitude (B)	1E Cell Size (m^2)	30O Cell Size (m^2)
0	12363671978	858631
15	11970315668	829390
30	10761202175	743627
45	8818730582	607188
60	6275277108	429370
75	3304173896	222291
90	107896706	63

Table 1. Areas of Cells from Sphere

These values have been applied to existing vegetation and land cover datasets to determine the exact areas of coverage of each of the vegetation types and land cover categories (Figure 3). Similar areas for datasets transformed in seven different projections are being computed and compared with the actual spherical areas. This process provides the extent of error for these seven projections with respect to these two datasets.

---

[1] Java is a trademark of Sun Microsystems, Inc.

The design of a projection procedure to maintain the areas determined in the numerical integration has taken the form of projecting each raster line independently. Using the sinusoidal projection equations as an example,

$$x = R(8- 8_0) \cos N \quad (1)$$
$$y = RN \quad (2)$$

where, N is the latitude,
8 is the longitude,
8_0 is the initial longitude,
R is the Earth's radius, and
x and y are the plane coordinates,

the procedure involves applying the equations to each of N and 8 for the four corners of the spherical cell to determine corresponding plane coordinates (x,y) (Snyder, 1987; Seong *et al*, 2002). The usual approach for raster data is then to resample all pixels to the same pixel size, thus sacrificing the areas. In our approach, we resample each raster line to the appropriate area determined by the numerical integration procedure and shown for specific latitudes in Table 1. The difficulty of this approach is that it results in a dataset that cannot be seamlessly mosaicked into a single raster image for display because each raster line contains pixels with sizes different from all other lines. This difference in sizes means that the approach works in analytical engines but cannot be used directly to produce a typical raster display (since pixels on different raster lines are of different sizes).

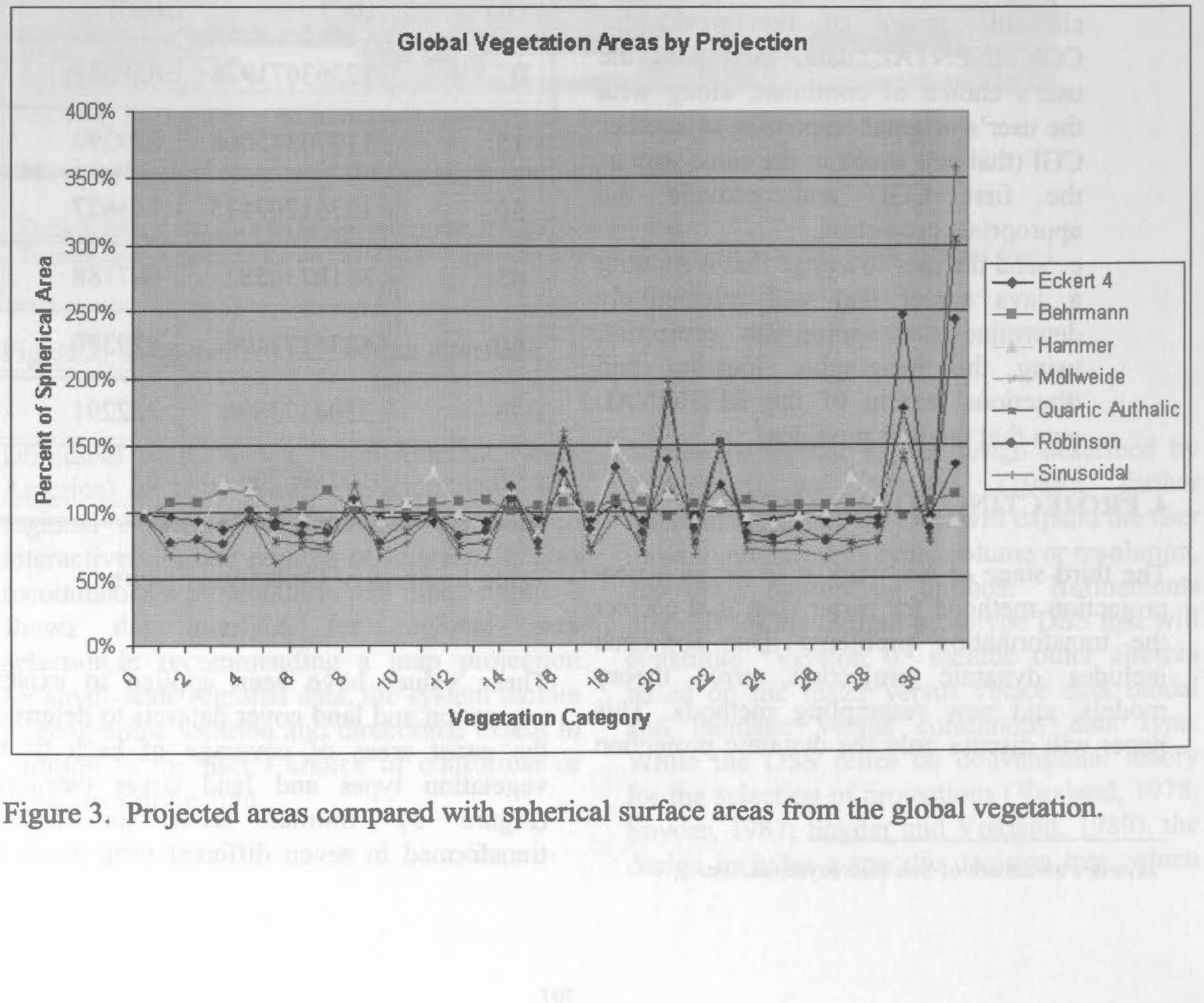


Figure 3. Projected areas compared with spherical surface areas from the global vegetation.

5. CONCLUSIONS

The projection of global raster datasets is a process that can result in considerable error for users unfamiliar with projection theory, implementation, and specific software problems. Our work has implemented a DSS to help users select an appropriate projection on the basis of factors commonly causing problems, including data type, location of regional area, extent direction, and characteristics to be preserved. We have also developed methods to improve accuracies for global raster projection for use with analytical engines. Future work will improve and expand the DSS to include a complete tutorial and examine the capabilities for displaying raster data of varying resolutions.

REFERENCES

Jankowski, P. and Nyerges, T., 1989. Design considerations for MaPKBS- Map Projection Knowledge-Based System. *The American Cartographer*, 16(2), pp. 85-95.

Seong, J.C. and E.L. Usery, 2001. Modeling raster representation accuracy using a scale factor model. *Photogrammetric Engineering and Remote Sensing*, 67(10), pp. 1185-1191.

Seong, J.C., K.A. Mulcahy, and E.L. Usery, 2002. The Sinusoidal Projection: A new meaning for global image data. *The Professional Geographer,* In press.

Snyder, John P., 1987. *Map Projections – A Working Manual.* U.S. Geological Survey Professional Paper 1395. Washington, DC.

Snyder, J.P. and P.M Voxland, 1989. *An Album of Map Projections.* U.S. Geological Survey Professional Paper 1453. Washington, DC.

Steinwand, D.R., J.A Hutchinson, and J.P. Snyder, 1995. Map projections for global and continental data sets and an analysis of pixel distortion caused by reprojection. *Photogrammetric Engineering and Remote Sensing*, 61(12), pp. 1487-1499.

Usery, E.L. and J.C. Seong, 2001. All Equal Area Map Projections are Created Equal, but Some are More Equal than Others. *Cartography and Geographic Information Science,* 28(3), pp. 183-193.

Usery, E.L., J.C. Seong, D. Steinwand, and M.P. Finn, 2001. Methods to achieve accurate projection of regional and global raster databases. *U.S. Geological Survey Open-File Report 01-383.*

Voxland, Philip M., 1978. *World Mapping and Projection Program: Reference Manual.* Regents of the University of Minnesota. Minneapolis, MN.

Yang, L., Z. Zhu, J. Izaurralde, and J. Merchant, 1996. Evaluation of North and South America AVHRR 1-Km data for global environmental modeling. *Third International Symposium on Integrating GIS and Environmental Modeling Proceedings*, Santa Fe, N.M.

WG IV / 9:

EXTRATERRESTRIAL MAPPING

USGS HIGH-RESOLUTION TOPOMAPPING OF MARS WITH MARS ORBITER CAMERA NARROW-ANGLE IMAGES

Randolph L. Kirk*, Laurence A. Soderblom, Elpitha Howington-Kraus, and Brent Archinal
Astrogeology Team, U.S. Geological Survey, Flagstaff, Arizona (rkirk@usgs.gov)

Commission IV, Working Group IV/9

KEY WORDS: Mars, topographic mapping, photogrammetry, photoclinometry, softcopy, extraterrestrial mapping

ABSTRACT

We describe our initial experiences producing controlled digital elevation models (DEMs) of Mars with horizontal resolutions of ≤10 m and vertical precisions of ≤2 m. Such models are of intense interest at all phases of Mars exploration and scientific investigation, from the selection of safe landing sites to the quantitative analysis of the morphologic record of surface processes. Topomapping with a resolution adequate to address many of these issues has only become possible with the success of the Mars Global Surveyor (MGS) mission. The Mars Orbiter Laser Altimeter (MOLA) on MGS mapped the planet globally with absolute accuracies <10 m vertically and ~100 m horizontally but relatively sparse sampling (300 m along track, with gaps of >1 km between tracks common at low latitudes). We rely on the MOLA data as the best available source of control and process images from the narrow-angle Mars Orbiter Camera (MOC-NA) with stereo and photoclinometric (shape-from-shading) techniques to produce DEMs with significantly better horizontal resolution. The techniques described here enable mapping not only with MOC but also with the high-resolution cameras (Mars Express HRSC, Mars Reconnaissance Orbiter HiRISE) that will orbit Mars in the next several years.

1. Introduction

Accurate topographic information, and, in particular, high-resolution digital elevation models (DEMs) are of intense interest at all phases of Mars exploration and scientific investigation, from landing site selection to the quantitative analysis of the morphologic record of surface processes. Unfortunately, the availability of extremely high resolution topographic data has hitherto been limited. The current "gold standard" for martian topographic data, the Mars Orbiter Laser Altimeter (MOLA; Zuber et al., 1992; Smith et al., 1999; 2001) has collected data globally with astonishingly high accuracy, but the resolution of this dataset is only about 300 m along track and, in many places near the equator, adjacent MOLA ground tracks are separated by gaps of one to several kilometers. Viking Orbiter images provide stereo coverage of the entire planet at low resolution and expected vertical precision (EP, a function of image resolution and stereo viewing geometry) but highest resolution stereo coverage only of limited areas (Kirk et al., 1999b). Given that the minimum separation of independent stereo measurements is about 3 pixels because of the necessity of matching finite-sized image patches, the highest resolution Viking images, at about 8 m/pixel, support stereomapping only at horizontal resolutions >24 m. Photoclinometry, or shape-from-shading (Kirk, 1987) can be used to produce DEMs at the pixel resolution from single images but the results depend on the accuracy of atmospheric and surface radiative transfer models (Kirk et al., 2000b, 2001). Calibration of photoclinometry against MOLA data (Herkenhoff, et al., 1999; 2002; Soderblom et al., 2002) or against stereo DEM data as reported here promises to reduce uncertainties about the inferred scale of relief, but albedo variations can still lead to artifacts in the resulting DEMs.

The MOC-NA camera, with a maximum resolution of 1.5 m/pixel (Malin et al., 1992; 1998; Malin and Edgett, 2001), offers the prospect of stereotopographic mapping at a horizontal resolution of ~5 m and EP ~ 1 m. MOC-NA stereo coverage is limited because most images are obtained with nadir pointing and are not targeted to overlap one another, but at least several tens of MOC-MOC stereopairs do exist. In addition, because of the urgent need to assess the safety and trafficability of candidate landing sites for the Mars Exploration Rover (MER) mission, further imaging, including stereo, of these sites is an important objective of the MGS extended mission. It is also likely that some MOC images will provide useful stereo coverage when paired with oblique Viking Orbiter images. A capability for stereomapping with the MOC-NA images is therefore highly desirable, but the pushbroom scanner geometry of the camera means that new software is required, as that developed for framing cameras, like those of the Viking Orbiter, will not suffice. The other main challenges in working with MOC-NA data are identifying suitable stereopairs and providing adequate geodetic control for such high-resolution images. In this paper we describe our methods and results for the Mars Pathfinder and candidate MER landing sites, expanding on a pair recent abstracts (Kirk et al., 2001b; 2002) that focused respectively on the methods and on the MER sites. We are not the only group working in this area; results of stereomatching uncontrolled MOC images are reported in a recent abstract by Ivanov and Lorre (2002)

2. Methodology

2.1 Software Implementation

The software packages, specialized hardware, and workflow for the MOC mapping reported here are the same as those we use for stereoanalysis of a wide range of planetary datasets (Kirk et al., 1999a; 2000a; Rosiek et al., 2001a; 2001b; Howington-Kraus et al., 2002a; 2002b). We use the USGS in-house digital cartographic software ISIS (Eliason, 1997; Gaddis et al., 1997; Torson and Becker, 1997) for mission-specific data ingestion and calibration steps, as well as "two-dimensional" processing such as map-projection and image mosaicking. Our commercial digital photogrammetric workstation, an LH Systems DPW-790 running SOCET SET software (© BAE

Systems; Miller and Walker, 1993; 1995) includes special hardware for stereo display of images and graphics, and is used mainly for such "three-dimensional" processing steps as automatic and manual measurement of image tiepoints; bundle-block adjustment of image orientations to conform to geodetic control; and automatic extraction and manual editing of DEMs. The ability to view DEM data as graphics overlaid on the images in stereo in order to detect and interactively edit errors is the single most important reason for using the commercial system.

In order to work with planetary data, we have written a set of translator programs drawing on both ISIS and the SOCET SET Developer's Toolkit (DEVKIT) to import images and geometric metadata from ISIS into SOCET SET and export DEMs and orthoimage maps back to ISIS. Images from planetary framing cameras (e.g., Viking, Clementine) can then be analyzed with the framing camera sensor model software supplied as a basic part of SOCET SET. (A sensor model consists of software that carries out the transformation between image and ground coordinates and vice versa, plus a variety of housekeeping routines.) The DEVKIT lets us implement and install sensor models for other instruments, such as the Magellan synthetic aperture radar (Howington-Kraus et al., 2002a). After beginning a similar in-house development of a sensor model for the MOC camera, we were able to take a substantial "shortcut" by making use of the generic pushbroom scanner model included in SOCET SET and writing only the software needed to import MOC images and set them up for use with this model.

The generic scanner model computes a physically realistic description of the process by which a scanner image is built up. It is "generic" in the sense that the following parameters must be specified and can be different for different cameras and/or images from the same camera:

- Image size; relation between line number and time
- Camera focal length and lens distortion polynomial
- Camera trajectory in the form of position and velocity at a series of equally spaced times spanning acquisition of the image, to be interpolated
- Camera orientation relative to the vertical and flight direction (nominally assumed constant)
- Corrections to the trajectory and orientation, normally initialized as zero and estimated as part of the bundle-adjustment process
  - Position and velocity offsets in the along-track, across-track, and vertical directions
  - Angular offsets around three orthogonal axes, angular velocities, and angular accelerations

These parameters suffice to describe not only the MOC-NA but also the wide-angle (WA) cameras, which have been used to obtain global stereo coverage with 240-m resolution (Caplinger and Malin, 2001); the infrared scanner of the THEMIS instrument on the Mars Odyssey Orbiter (Christensen et al., 1999) the Mars Express High Resolution Stereo Camera (HRSC) with multiple detector lines for single-pass stereo imaging at 10 m/pixel (Albertz et al., 1992); and the HiRISE scanner selected by NASA for the 2005 Mars Reconnaissance Orbiter mission (McEwen et al., 2002). Not only can the generic scanner model be used with images from any of these cameras, SOCET SET permits bundle-adjustment and stereo DEM collection with any combination of scanner and framing camera data in a single project. To date, we have written software to collect the necessary information from both MOC-NA and WA image labels, convert geometric quantities from the inertial coordinate system used by ISIS to the local Cartesian system used by SOCET SET, and write this supporting data in the needed format.

The main limitation of the software that affects its use with MOC-NA images is the nominally constant orientation of the camera. Images obtained (during the aerobraking phase of the MGS mission) by rotating the spacecraft do not fit this model, and our experiments with representing the spacecraft rotation by using the adjustable parameters have so far been unsuccessful. An enhancement of the sensor model promised for a future release will allow an arbitrary time history of camera orientation to be specified. The limited set of adjustable parameters in the model also has its drawbacks, and this is unlikely to be changed. The low-order (smooth) position and pointing corrections possible with these parameters cannot address the high-frequency undulations of the MGS spacecraft that plague some MOC stereopairs as discussed below. Nor does the SOCET SET bundle-adjustment software understand that images from a multiline scanner such as HRSC come from the same spacecraft and are subject to the same position and pointing corrections as a function of time. To address these shortcomings it is necessary to implement more capable bundle-adjustment software outside SOCET SET and then import the corrected geometric data derived with such software. We have proposed to NASA to add pushbroom scanner capability to the RAND bundle-adjustment software recently taken over by the USGS. We plan initially to model single-line scanners, including high-frequency pointing oscillations. Modeling of multiline scanners could be undertaken at a later date if adequate orientation data and/or adjustment software to produce such data are not produced within the Mars Express mission.

2.2. Identification of Stereoimagery

Identifying suitable pairs of MOC-NA images for stereoanalysis is a significant challenge, given that more than 30,000 images have been released to date but the typical image covers only about one one-millionth of Mars's surface area. We have pursued several approaches to identifying pairs for initial testing of our software. First, MOC press releases on the Malin Space Science Systems (MSSS) website http://www.msss.com/mars_images/moc/MENUS/moc_by_date.html include a number of anaglyphs made from NA stereopairs. Unfortunately, the most dramatic of these stereo views are also the most recent (press releases MOC2-256, 282, 283, and 287) and contain images not yet released. Earlier releases (Table 1) show the Viking 1 (MOC2-44), Mars Polar Lander (MOC2-255), and Mars Pathfinder (MOC2-46; Figure 1) landing sites. Of these, the last was selected for initial testing because we have previously mapped parts of the region with both Viking Orbiter (Howington-Kraus et al., 1995) and Mars Pathfinder IMP (Kirk et al., 1999a) images. The stereo pair unfortunately does not cover the actual landing point, but does include "Big Crater" (prominent in the lander images) and plains to the south and west. Non-stereo MOC coverage of the landing point shows that it and the plains south of Big Crater are extremely similar in texture and morphology.

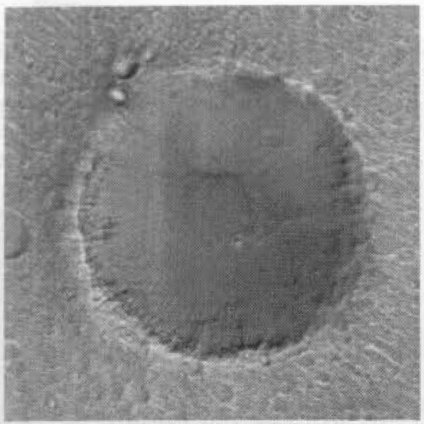


Figure 1. Anaglyph of "Big Crater" (1.5 km in diameter, located 2.2 km SSE of Mars Pathfinder landing point), taken from MSSS press release MOC2-46. Figure shows a small part of stereo overlap of images and SP1-23703SP1-25603, which extends to Sand W (Fig. 2) but unfortunately does not cover Pathfinder site. For correct stereo impression, view with red filter on right eye.

The optimal way to identify stereopairs is obviously through an exhaustive, automated search of catalog data. We previously conducted such a search for the ~45,000 Viking Orbiter images, which yielded ~360,000 potential pairs (i.e., each image overlaps about 8 others), only a fraction of which had useful EP (Kirk et al., 1999b). Here, we report on similar searches of the MOC-NA catalog (the cumindex.tab file on the official MOC releases, plus similar data for pre-released, extended-mission images of the MER landing sites). An initial search of catalog data, carried out in August 2001, generated 4,872 candidate pairs from the list of 31,901 images through mission phase E06. This list was narrowed by excluding those with stereo convergence angles too large or too small, and those with dissimilar illumination (criteria similar to those used by Kirk et al. (1999b) and adopted from the work of Cook et al. (1996)). This step yielded 158 candidate pairs, 18 of which were located in regions under consideration as MER landing sites; because of the urgency of assessing the safety of these sites, we limited further consideration to the images there. We interactively examined these images and eliminated the ones with little or no overlap, and those with low image contrast due to atmospheric opacity. The image pairs found for each of 4 sites are included in Table 2. Results of mapping and roughness analysis for these image sets (plus the Pathfinder site) were reported in a recent abstract (Kirk et al., 2002).

Unfortunately, a more recent search of the MOC catalog through mission phase E13 produced many new candidates but only a few additional usable pairs, all of which had already been found by interactive search as described below. The low success of this search may be due in part to a MOC targeting strategy that minimizes stereo overlap with past images in order to maximize new area covered, and in part to the substantial errors (comparable to or greater than the image width) in the image footprint locations catalogued on the basis of predicted spacecraft ephemerides. An additional source of error is the approximate nature of the criteria we used for automatically determining intersection of images: a test was made whether any of the corners of the minimum bounding rectangle (MBR) of one image in latitude and longitude fell inside the MBR of the other image. We plan in the near future to repeat the search for MOC image pairs, incorporating a rigorous test for the overlap of the actual image footprints (not their MBRs) and calculating footprints calculated from the mission ephemerides (rather than using those from the cumulative index) to reduce positional errors. Because coordinate errors cannot be eliminated entirely, this search will probably still require a final, manual step of checking the images for their actual overlap and contrast, and it may be useful to "pad" the image footprints by an amount corresponding to the coordinate errors in order to search for images that overlap in fact but not according to their SPICE data. We will also look for mixed pairs of MOC and Viking images that yield usable stereo.

Maps of MOC-NA image coverage of prospective MER landing sites (http://marsoweb.nas.nasa.gov/landingsites/mer2003/mer2003.html) provide another resource for locating stereopairs in these areas of highest current interest. Note that the interactive maps on this site do not always contain the most recent releases, but the site contains links to thumbnails of these images and context images for them that are also useful. Image footprints have been placed interactively in these maps and context images, so they are more accurate than the catalog information but not error-free. Pairs of images whose footprints appear to overlap must be screened to determine if they actually overlap, if they have useful stereo geometry (at a minimum, at least one off-nadir image), and if the illumination is consistent enough between images to permit automatic stereomatching. Additional image sets were subsequently rejected for a variety of reasons including inadequate overlap, obscuration of the surface by atmospheric dust, changes in surface appearance between the two images due to wind activity, missing spacecraft orientation data, and, in one case, an imaging strategy not supported by our software. Oblique image AB1-07704, which crosses nadir images M08-01457 and M09-01839 in the Hematite area, was obtained with the spacecraft rotating, and our attempts to model this image with the current software have been unsuccessful. In total, we have identified 16 usable stereosets within the MER landing sites and extracted DEMs from 10, including at least one at each of the six sites except Hematite (Table 2). It is noteworthy that all of these images were obtained by 2x2 or 4x4 summation of pixels, so that their resolutions are not as high as the MOC-NA camera is capable of. In all cases, however, the resolutions are better than the best Viking Orbiter images.

2.3 Geodetic Control

Our experience with map-projecting and comparing MOC-NA and WA images indicates that errors of position (combining both spacecraft position and pointing errors) are often <100 m but occasionally greater, especially for off-nadir images. This is adequate to produce uncontrolled mosaics of WA images (≥240 m/pixel resolution) but inadequate for the higher resolution NA data. In particular, 100-m relative horizontal errors between images of a stereopair will give rise to comparable vertical errors in the DEM. It is therefore highly desirable to use a bundle-adjustment process to bring the images into consistency with external control. This process is made challenging by the large gap in resolution between the NA images and the next-best datasets available for control. The MOLA dataset, with estimated absolute accuracies of <10 m vertically and ~100 m horizontally (Neumann et al., 2001), is the ultimate source of control, but the spacing of MOLA footprints is hundreds to thousands of MOC-NA pixels. Direct comparison of the MOC images with MOLA profiles or gridded MOLA data is therefore helpful in bringing the stereomodels into vertical correspondence with the altimetry but not very useful for improving their horizontal positioning. In our opinion, the best approach to refining the horizontal position of MOC-NA images and stereomodels would be to tie other images of intermediate resolution to the MOLA data and then tie the MOC images to these. At the moment, this necessarily means Viking Orbiter images—the best available resolution at each of the MER sites is indicated in Table 2—but over the next few years the visible light camera of THEMIS will provide 18 m/pixel CCD frame images of much of Mars. We routinely use intermediate resolution frames to transfer control from MOLA in our stereomapping with Viking images (Rosiek et al., 2001a) but we have yet to test the process with MOC data. The use of a large number of ties between intermediate-resolution images and MOC-NA will be essential to modeling and correcting the high-frequency oscillations of the MGS spacecraft with the advanced bundle-adjustment software we hope to develop. As discussed above, the bundle-adjustment capability currently available as part of SOCET SET does not include modeling of such high-frequency oscillations.

For the purposes of landing site selection (as opposed to precision landing), precise relative topographic data (from which slope estimates can be made) is more important than absolute accuracy. Our efforts to control the images listed in Tables 1-2 have therefore focused on bringing the stereomodels into vertical agreement with MOLA data and not on improving horizontal positioning. For each stereopair, we select a well-distributed set of points (typically 10–20) whose locations can be measured on both MOC images. We then assign each point the elevation interpolated from MOLA data at its *a priori* horizontal location. These heights are used as constraints in the bundle-adjustment process, but no constraints are placed on horizontal positions. The form of the MOLA data used in our control process is an in-house USGS product, gridded at 500-m ground sample distance and adjusted to conform to the most recent set of Mars cartographic constants recommended by the International Astronomical Union and International Association of Geodesy (Duxbury et al., 2002; Seidelmann et al., 2002). Similar products for the candidate MER landing sites are available online at http://webgis.wr.usgs.gov/mer/.

The control process for the stereopair near the Mars Pathfinder landing site was even simpler. Rather than determining an interpolated MOLA elevation for every tiepoint, we estimated the average MOLA elevation for the region of the stereomodel and loosely constrained a subset of tiepoints (away from Big Crater and other obvious relief) to have this elevation. The result is to underestimate any net regional slope of the stereomodel.

Our attempts to control the stereopairs listed in Table 2 have been complicated by transmission errors in one or both of the images. It is a characteristic of the Huffman coding used to compress these high-resolution images that signal degradation between the spacecraft and ground station can cause the loss of blocks of image lines; if the degradation is extensive, it may not be known how many lines were lost. When this happens, the correct acquisition times of the lines after the gap (needed in sensor model calculations) are lost. We are working on semiautomated approaches to reconstructing the image line count from information returned in the images, but for the time being our approach to working with MOC data after a data dropout is strictly empirical. We first compare the corrupted image with an uncorrupted image of the same region, estimate the number of missing lines, and insert this number of blank lines into the gap to approximately restore the image. We then control the image, being careful to place pass-points only in the section below the gap (if the section above the gap is needed for mapping, we treat it as an independent image) and allowing a larger than usual along-track adjustment of the spacecraft position to account for the error in reconstructing the gap size.

The final complication to the control process, encountered for most MER landing site stereopairs (but not for the Pathfinder site), arises from the high-frequency spacecraft pointing errors or "jitter" alluded to previously. Pointing oscillations in direction of the stereobase (i.e., across-track for most MOC pairs) mimic elevation-related parallax and result in artifacts in the DEMs in the form of "stripes" elongated across-track; these are also reported by other users of MOC stereoimagery (Ivanov and Lorre, 2002). From the dimensions of these DEM stripes we can infer the magnitude (highly variable but as much as 50 μRad or more than 10 MOC pixels) and frequency (0.1–1 Hz) of the oscillations that cause them. Oscillations of similar amplitude are seen in the spacecraft pointing data but the 2 second sampling interval of this dataset results in the jitter being aliased to much lower frequencies. Jitter in the direction orthogonal to the stereobase (normally, along-track) results in localized mismatches between the images that confound both automatic and interactive DEM extraction. It is important to note that the likely magnitude of high-frequency pointing errors was known to MGS engineers (if not to users of the data) even before the mission. The MGS spacecraft was not specifically designed for high resolution stereoimaging, and the MOC camera must share the platform with other instruments and articulated solar panels and high gain antenna, all of which are sources of high-frequency vibration at a magnitude that does not impact the primary MOC goal of imaging for geologic interpretation.

The rigorous and therefore desired approach to correcting for these spacecraft pointing oscillations is to collect a dense set of tiepoints between the MOC stereoimages, intermediate-resolution frame images, and the MOLA DEM, and to perform a bundle-adjustment with specialized software that includes an appropriate parameterization of the high-frequency motion. Until such software can be developed, we have been forced to develop alternative procedures to work around the problem. If the DEM can be collected, ad hoc processing to suppress the stripes in the direction of the parallax base is relatively straightforward. We first take the difference between the stereo-derived and MOLA DEMs at matching resolution, then apply a series of digital filters (lowpass across the image strip, highpass in the flight direction) to get an estimate of the pointing-related artifacts that can then be subtracted from the stereo DEM. Of course, this procedure also suppresses real topography if it consists of features (e.g., ridgelines) that happen to cross all or most of the image. The narrower the region of stereo overlap, the more difficult it becomes to make a useful distinction between artifacts and real topography.

For stereopairs with the most severe jitter, it may be impossible to collect a DEM based on a single control solution with the existing software: if corresponding image lines are registered in one part of the pair, they may be so misaligned elsewhere that automatic stereomatching fails and the brain is incapable of fusing them in stereo. In such cases we have found it necessary to break the images into smaller sections and adjust the across-base orientation angle separately for each section. Sectional DEMs can then be collected, edited as usual, and merged. The resulting product is necessarily imperfect, as a continuous oscillation of the across-base angle has been corrected in a piecewise fashion, but the discontinuities in the DEM that result can be minimized. Needless to say, this process is undesirable both because it is extremely time-consuming and because it is approximate and necessarily subjective.

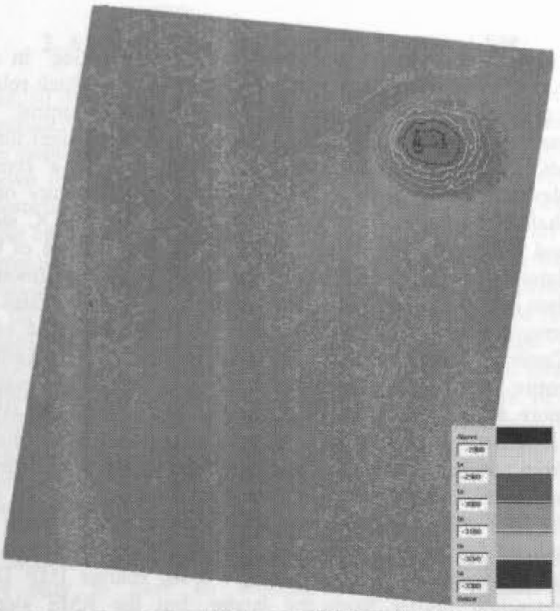


Figure 2. Orthorectified MOC image SP1-23703 with stereo derived contours (contour interval 50 m, color interval 100 m). Total relief from floor to rim of Big Crater is 300 m. Automatic stereomatching was successful except for a single artifact (red dot at center left).

2.4 Analysis of DEM Data

We describe the process of DEM analysis in some detail for the Pathfinder site, which was the first MOC stereopair mapped and for which there are a variety of datasets available for comparison. Figure 2 shows image SP1-23703, orthorectified, with superimposed contours derived from the stereo DEM. The low relief of this region is readily apparent: from the bottom to the top of Big Crater the total relief is 300 m. The most prominent feature in the unedited DEM, apart from the crater, is an apparent peak ~75 m high and 120 m across the base, located near the middle right of the image. This is not a real topographic feature, but an artifact caused by local failure of the automatic matching algorithm. It is also visible in a perspective view of the DEM (Figure 3). The high overall success of the SOCET SET matcher is gratifying, given the relatively low contrast of the images.

Figure 4 shows the image and DEM data for a portion of the stereomodel excluding both Big Crater and the matching artifact. This section was selected for statistical analysis to characterize the flat part of the landing site. The DEM in this area is consistent with our description based on IMP data of the landing site topography as dominated by ridges and troughs with a typical amplitude of a few meters and a wavelength of several tens of meters (Kirk et al., 1999a). The smallest ridges are not fully resolved in the DEM but a pattern of somewhat larger ridges can be seen. Comparison of the image and DEM tends to support our previous assertion that many of the bright albedo features in the images are associated with local topographic highs. These are probably strips of light-colored, rock-free sediment similar to those seen (also along ridges) near the lander and interpreted as dunes by Greeley et al. (2000).

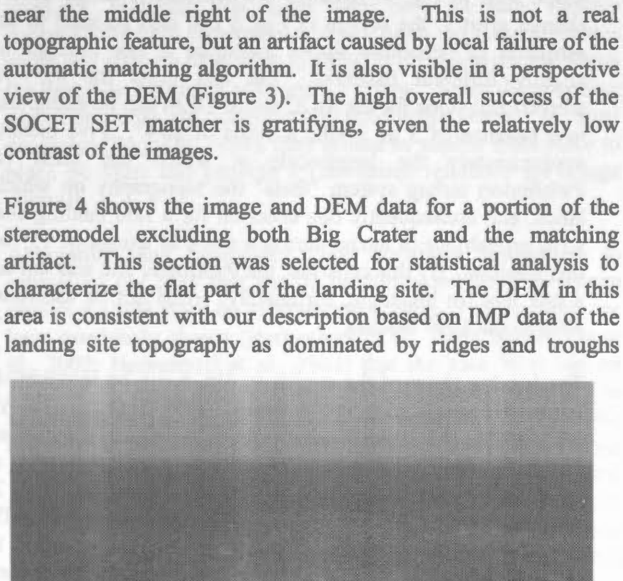


Figure 3. Perspective view of Big Crater from the northwest. 3 m/pixel MOC image SP1-23703 has been draped over 10 m/post stereo-derived DEM. Vertical exaggeration is 2. Matching artifact is visible as a small "hill" in the background. Colors are arbitrary, intended only to suggest the appearance of the martian surface.

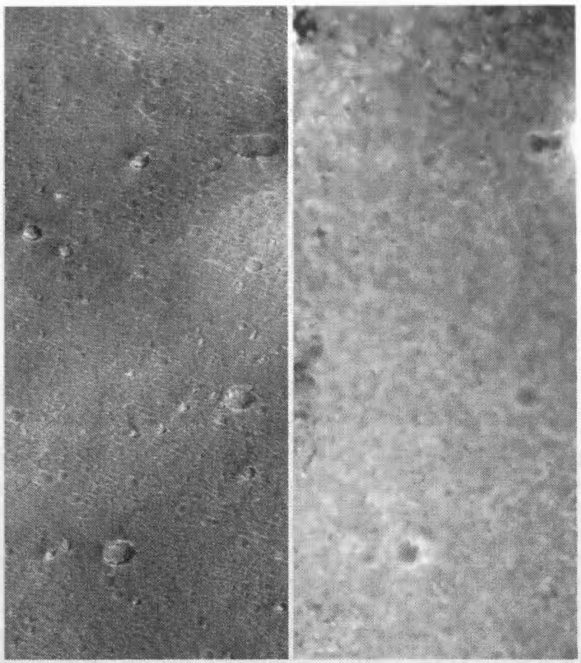


Figure 4. Orthoimage and DEM data (shown at right as grayscale) for ~2.6 x 7.5 km section of the MOC stereomodel SP1-23703/SP1-25603 to the W and S of Big Crater. This region was selected for slope analyses and comparison with results from the Mars Pathfinder lander, which is located on similar plains to the N. Total range of elevations in this area is ≤5 m.

Figure 5 shows the distribution of bidirectional slopes derived from the DEM area in Fig. 4, for a north-south baseline of one post (approximately 12 m). The root-mean-squared (RMS) slope is 4.2° but, as the histogram shows, the slope distribution has longer tails (i.e., extreme slopes are relatively more common) than for a Gaussian distribution. Slopes on this baseline are in the range ±14.1° for 99% of the test area. For the adirectional slope (maximum slope in any direction, or gradient) over the same baseline, the 99th %ile is 15.8°.

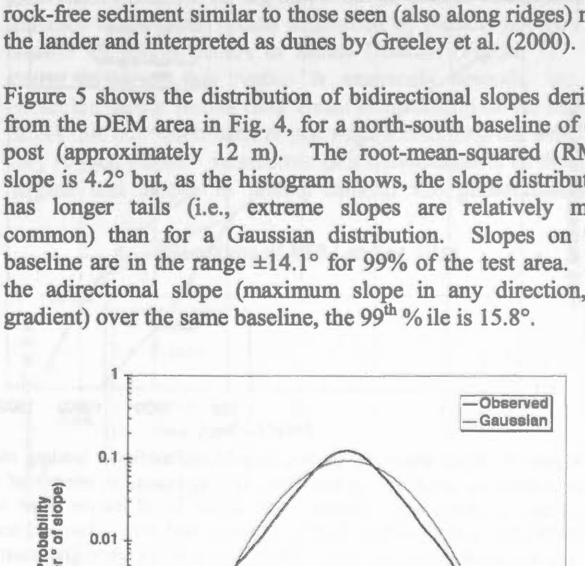


Figure 5. Histogram of bidirectional slopes over a 1-post (12 m) N-S baseline from the portion of the MOC DEM near Big Crater seen in Fig. 4. Gaussian distribution with the same RMS slope (4.2°) as observed is shown for comparison. Large slopes are significantly more common than the Gaussian model would suggest.

It is also of interest to look at the RMS slopes over a variety of distance scales. Not only do slopes over long baselines and local slopes have different implications for landing safety and rover trafficability, but this type of analysis lets us compare the MOC DEM with other topographic datasets for the region. If $z(x)$ is a profile of height as a function of horizontal coordinate, and Δ is a horizontal baseline ("lag"), then the one-dimensional autocovariance function $\rho(\Delta)$ is given by

$$\rho(\Delta) = \langle z(x)\, z(x+\Delta) \rangle$$

where the brackets $\langle\ \rangle$ indicate an ensemble average over both x and multiple parallel profiles. From the above definition, it is easy to show that the RMS (dimensionless) slope over the baseline Δ is given by

$$\Theta_{RMS}(\Delta) = \sqrt{\{\, 2\, (\rho(0) - \rho(\Delta)) / \Delta \}}.$$

Figure 6 is a plot of $\Theta_{RMS}(\Delta)$ calculated from autocovariance estimates obtained by fast Fourier transform techniques for four independent DEMs covering different parts of the Mars Pathfinder landing ellipse at a variety of scales. Dimensionless slopes have been multiplied by the conversion factor from radians to degrees, so the larger slope angles are slightly overstated here. Each of the curves shows a characteristic "dogleg" shape, with a steep section with $\Theta_{RMS}(\Delta) \propto \Delta^{-1}$ for large Δ and a shallower log-log slope at small Δ (the highest resolution dataset does not show an extended steep section but the beginning of a rolloff is nonetheless visible). The steep curves at baselines that are large relative to the respective DEM reflect the fact that each of these topographic models has been controlled essentially to be level overall, so slopes on the longest baselines are underestimated. The "dogleg" in each curve thus reflects the resolution at (and below) which relative topography and slopes are unaffected by errors in control. This scale is especially small (in relation to the DEM size) for our MOC stereomodel because we controlled it to a constant elevation rather than to MOLA data. The continuity of the shallow portions of the Viking photoclinometry, MOC stereo, and IMP stereo curves is striking, given that no two of these datasets cover precisely the same area. The photoclinometry data are taken from Viking image 004A72, which contains only smooth plains similar to those near the lander and south of Big Crater. In contrast, the Viking stereo data cover almost the entire 100x200-km Pathfinder landing ellipse, which contains more rugged features such as craters and streamlined islands (Howington-Kraus et al., 1995). Photoclinometry on images from rougher parts of the landing ellipse yields slope estimates as much as a factor of two greater than for 004A72. The curve for IMP data is derived from a DEM extending 10 m on each side of the lander (Kirk et al., 1999a). Slope estimates over larger baselines can be computed from coarser IMP DEMs extending farther from the lander, but the RMS slope is systematically underestimated in these datasets because much of the distant landing site was hidden behind ridges and the DEMs therefore contain (unrealistically) smooth patches interpolated from actual data for the visible areas. We have therefore not included these estimates in Fig. 6, but their lower RMS slopes are consistent with the value of 4.7° at 1-m baselines quoted by Kirk et al. (1999a).

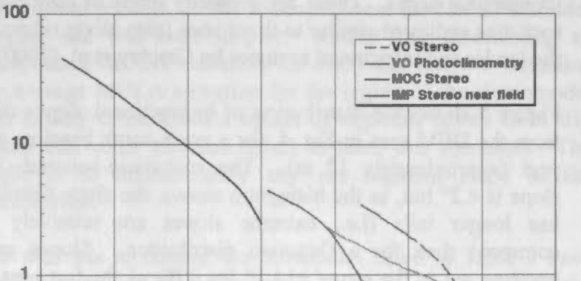


Figure 6. RMS slopes of regions near Mars Pathfinder landing site calculated as described in text from DEMs based on images of a variety of scales and sources. VO stereo DEM covers most of Pathfinder landing ellipse (including some high-relief features) and was interpolated from contours obtained by analytic photogrammetry of 38 m/pixel Viking images. VO photoclinometry DEM was obtained from 38 m/pixel image 004A72 of smoothest part of landing ellipse, highpass filtered to suppress artifacts of photoclinometry. MOC stereo DEM derivation is reported here. IMP stereo DEM covers the region from the lander to 10 m at GSD of 2 cm, and was interpolated from data collected on IMP stereopairs with highly variable ground spacing. Slopes over smaller baselines for each DEM are expected to be most accurate and are consistent between datasets. Slopes at longest baselines for each dataset are affected by the control process and systematically underestimate real slopes.

There is no necessary reason for the shape of the slope distribution to be independent of baseline, but it is likely that this assumption is at least approximately correct over modest baseline variations, and we have found in practice that slope distributions for many areas of Mars evaluated at various baselines almost always have the long-tailed, nearly exponential shape seen in Fig. 5. Because the distribution shape does not change greatly, the curves in Fig. 6 can also be used to scale estimates of percentile slopes measured at one baseline to a slightly different baseline. On this basis the 99th %ile adirectional slope for the MOC stereo DEM can be extrapolated to a baseline of 5 m, giving a value of 20.4°. Five meters is approximately the lengthscale at which the MER (and Pathfinder) airbag system "feels" the topography on which it lands, and consequently one criterion for a safe landing site is $\leq$1% probability of encountering a slope in excess of 15° over this baseline. By this criterion, the Pathfinder site was *not* safe, a fact that (in hindsight, fortunately) could not be ascertained prior to the 1997 landing.

The log-log slope of the curves in Fig. 6 can be interpreted in terms of fractal geometry (Turcotte, 1997): if $\Theta_{RMS}(\Delta) \propto \Delta^{H-1}$, where $0 \leq H \leq 1$ is the Hurst exponent, the fractal dimension of the surface is $D = 3 - H$. The Viking photoclinometry and MOC stereo datasets show a similar dimension $D \sim 2.3$, whereas for the IMP data $D \sim 2.4$. This difference may reflect the structural features (fluvial or eolian ridges vs. rocks) that dominate relief at different scales, but it should be borne in mind that the slope estimates are also affected by the noise properties of the data and method used to produce the DEM. In any case, we find it interesting that a straight-line extrapolation of the Viking photoclinometry curve from baselines $\geq$80 m yields slopes at centimeter scales that are within a factor of two of those measured *in situ*.

3. Application to Candidate MER Landing Sites

The 10 stereopairs from which we have collected DEMs of candidate MER landing sites are listed in Table 2. Analysis of the slope statistics for these sites was carried out as described above for the Pathfinder site. In addition, two-dimensional photoclinometry (Kirk, 1987) was used to generate DEMs of most of the sites at single-pixel (as opposed to three-pixel for stereo) post spacing. This technique, also known as shape-from-shading, amounts to the iterative, least squares adjustment of the DEM post elevations in order to make a shaded relief image calculated from the DEM fit the observed image. An important limitation of the method is that variations in surface reflectivity (albedo) are not generally accounted for in the shaded relief model. Areas of relatively constant albedo (at a minimum, more image contrast from topographic shading than from albedo variations) must therefore be selected for the results to be valid.

In order to obtain a quantitatively accurate DEM, it is necessary not only to have an accurate photometric model (i.e., bidirectional reflectance function) for the surface (Kirk et al., 2000b) but to account for light scattered in the atmosphere (Kirk et. al., 2001). Unless the degree to which image contrast is reduced by atmospheric scattering is correctly understood, the result will be a DEM with the correct qualitative shape but the wrong amplitude of relief, obviously useless for collecting slope statistics. A well-established approach to estimating the haze component of images for photoclinometry is to measure the brightness of a shadow and subtract this from the entire image. Unfortunately, the MOC orbit geometry precludes shadows in most images of the near-equatorial zone where the MER sites are located. We were able to use the shadow correction method for one nonstereo image (M07-01888, not in Table 2) of Athabasca Vallis. In the Hematite site, for which no stereo coverage and no shadows were available, we estimated the haze level by requiring that dunes there have the same relative contrast after haze correction as dunes (with the same incidence angle and azimuth relative to the Sun) in other images for which the haze was estimated as described below. Since the opposing slopes on dunes are controlled by relatively universal physical processes (Bagnold, 1954), they should be similar from place to place on Mars and provide a convenient reference for image contrast.

Alternative (and potentially more accurate) approaches to estimating the haze that should be subtracted from an image require a DEM from an independent source. For example, we have previously shown (Herkenhoff et al., 1999; Soderblom et al., 2002; Herkenhoff et al., 2002) that the haze level can be adjusted to bring the resulting photoclinometric DEM into agreement with a MOLA profile. Two important conditions for this method to work are, first, that the MOLA profile run close to the direction (i.e., up-sun) for which photoclinometric profiles are most accurate, and, second, that there be significant relief resolved by MOLA, which has a 300 m along-track sample spacing. The first condition is violated in the equatorial zone (illumination is roughly at right angles to the MOLA tracks) and the second is violated by many of the landing sites, which are deliberately chosen to be flat over large baselines. Use of a stereo-derived DEM circumvents both of these problems. We calculate a shaded-relief image from the stereo DEM by using a realistic surface photometric model but no atmospheric contribution. Provided the model resolves topographic features well enough that the shaded relief closely resembles the real image (this is readily evaluated visually so that a suitable subarea of the shaded relief model can be chosen for comparison), the intercept of a linear regression of image on shaded relief is the desired haze parameter.

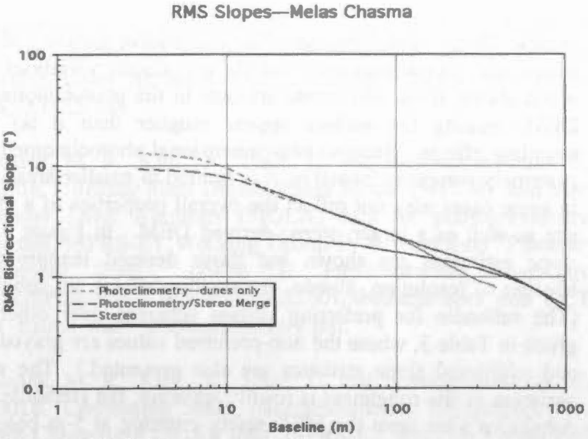


Figure 7. RMS slopes in a dune-covered region of Melas Chasma (stereomodel E02-00270/E05-01626) as a function of baseline; compare Fig. 6. Stereo DEM gives low slopes with little variation at baselines <100 m, suggesting that the site is extremely smooth, but photoclinometry reveals that dunes unresolved by the stereomatcher are much rougher. Figure also shows that stereo and photoclinometric DEMs can be merged to combine the absolute accuracy at long-baselines of the former with the high spatial resolution of the latter. (Curve for the merged DEM shows slightly lower slopes than for pure photoclinometry DEM at short baselines because it covers a larger area including some dune-free regions.)

With both stereo and photoclinometric DEMs available for most sites, it becomes important to determine which provides the most accurate slope estimates. The most important factor is the horizontal resolution of the two methods, which is readily addressed by Fourier analysis. Although the ~3-pixel post spacing of our stereo DEMs exceeds the 5-m baseline at which the slope must be estimated for the MER safety criterion, in most cases the slope-vs.-lag curve can be used to extrapolate the statistics to 5 m with results confirmed by photoclinometry. A notable exception occurs in Melas Chasma (Figure 7). The stereo DEM of this region is extremely smooth, but the photoclinometric results (and visual examination of the images) shows that the stereomatcher just misses resolving the small but very rugged dunes. Interactive spot measurements of the dune heights and slopes in stereo confirm the photoclinometric

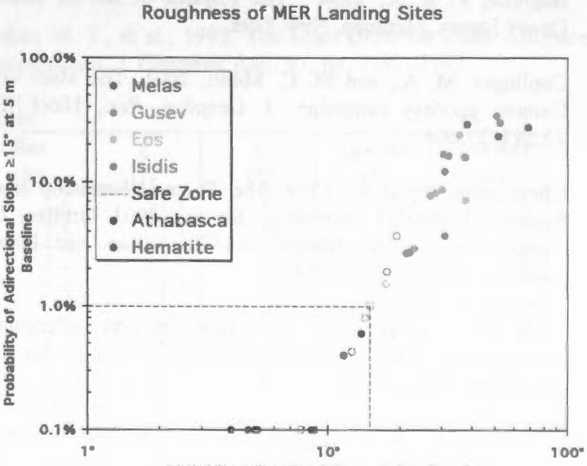


Figure 8. Summary of the roughness of regions of the 6 candidate MER landing sites, relative to the safety criterion of ≤1% probability of slopes in excess of 15° at the 5-m baseline sensed by the landing airbag system. Symbols for DEM areas that are unrepresentative or compromised by resolution effects or albedo variations are subdued.

results. Other conditions that affect the relative accuracy of the stereo and photoclinometric DEMs are albedo variations (as noted above, these will create artifacts in the photoclinometric DEM, making the surface appear rougher than it is) and sampling effects. Because two-dimensional photoclinometry is extremely computer intensive, it is limited to smaller areas that in some cases may not reflect the overall properties of a given site as well as a larger, stereo-derived DEM. In Figure 8, all slope estimates are shown but those deemed inappropriate because of resolution, albedo, or sampling effects are subdued. (The rationale for preferring certain subareas over others is given in Table 3, where the non-preferred values are grayed out and additional slope statistics are also presented.) The wide variation in site roughness is readily apparent; the Hematite and Athabasca sites meet the MER safety criterion at 5-m baseline but the others to a greater or lesser degree do not. The ranking of sites in terms of roughness shown here is in good qualitative agreement with results from the analysis of a variety of other disparate datasets (see http://marsoweb.nas.nasa.gov/landingsites/mer2003/mer2003.html for reports of the October 2001 and March 2002 MER Landing Site Workshops). As such results are published in greater detail, we plan to compare them with our analysis of the MOC images.

4. Conclusion

The newly developed software and techniques reported here are opening a door to a new realm of Mars topography. The ability to produce DEMs with horizontal resolutions of 10 m and better will be invaluable for selecting future landing sites and will contribute greatly to the study of surface processes. These capabilities will be almost immediately applicable to high-resolution stereoimagery from current and future missions such as 2001 Mars Odyssey, Mars Express and Mars Reconnaissance Orbiter, as well.

5. References

Albertz, J., et al., 1992. The camera experiments HRSC and WAOSS on the Mars 94 mission, *Int. Arch. Photogramm. Remote Sens.*, *29*(B1), 130–137.

Bagnold, F. R. S., 1954. The Physics of Blown Sand and Desert Dunes. Methuen, New York.

Caplinger, M. A., and M. C. Malin, 2001. The Mars Orbiter Camera geodesy campaign. *J. Geophys. Res.*, 106(E10) pp. 23,595–23,606.

Christensen, P., et al., 1999. The Thermal Emission Imaging System (THEMIS) Instrument for the 2001 Orbiter. *Lunar Planet. Sci.*, XXX, Abstract #1470, Lunar and Planetary Institute, Houston (CD-ROM).

Cook, A. C., et al., 1996. Clementine imagery: selenographic coverage for cartographic and scientific use. *Planet. Space Sci.*, 44(10), pp. 1135–1148.

Duxbury, T. C., Kirk, R. L., Archinal, B. A., and Neumann, G. A., 2002. Mars Geodesy/Cartography working Group recommendations on Mars cartographic constants and coordinate systems, *Int. Arch. Photogramm. Remote Sens.*, submitted (this conference).

Eliason, E., 1997. Production of digital image models using the ISIS system. *Lunar Planet. Sci.*, XXVIII, pp. 331–332, Lunar and Planetary Institute, Houston.

Gaddis, L. et al., 1997. An overview of the Integrated Software for Imaging Spectrometers (ISIS). *Lunar Planet. Sci.*, XXVIII, pp. 387–388, Lunar and Planetary Institute, Houston.

Greeley, R., Draft, J. D., Kuzmin, R. O., and Bridges, N. T., 2000. Mars Pathfinder landing site: Evidence for a change in wind regime from lander and orbiter data. *J. Geophys. Res.*, 105, pp. 1829–1840.

Herkenhoff, K. E., Bridges, N. T., and Kirk, R. L., 1999. Geologic studies of the Mars Surveyor 1998 landing area. *Lunar Planet. Sci.*, XXX, Abstract #1120, Lunar and Planetary Institute, Houston (CD-ROM).

Herkenhoff, K. E., Soderblom, L. A., and Kirk, R. L., 2002. MOC Photoclinometry of the North Polar Residual Cap. *Lunar Planet Sci.*, XXXIII, Abstract #1714, Lunar and Planetary Institute, Houston (CD-ROM).

Howington-Kraus, E., Kirk, R., Galuszka, D., Hare, T., and Redding, B., 2002a. Validation of the USGS sensor model for topographic mapping of Venus using Magellan radar stereoimagery. Lunar Planet. Sci., XXXIII, Abstract #1986, Lunar and Planetary Institute, Houston (CD-ROM).

Howington-Kraus, E., Soderblom, L, Kirk, R., Giese, B., and Oberst, J., 2002b. USGS and DLR topographic mapping of Comet Borrelly. *Int. Arch. Photogramm. Remote Sens.*, submitted (this conference).

Howington-Kraus, E., Kirk, R. L., Redding, B., and Soderblom, L. A., 1995. High-resolution topographic map of the Ares-Tiu landing site from Viking Orbiter data. In: Mars Pathfinder Landing Site Workshop II, LPI Tech. Rep. 95-01, pp. 38–39; map reproduced in Tanaka, K. L., 1997. Sedimentary Hhstory and mass flow structures of Chryse and Acidalia Planitia, Mars. *J. Geophys. Res.*, 102, pp. 4131–4149.

Ivanov, A. B., and Lorre, J. J., 2002. Analysis of Mars Orbiter Camera Stereo Pairs. *Lunar Planet. Sci.*, XXXIII, Abstract #1845, Lunar and Planetary Institute, Houston (CD-ROM).

Kirk, R. L., 1987. *III. A fast finite-element algorithm for two-dimensional photoclinometry*, Ph.D. Thesis (unpubl.), Caltech. pp. 165–258.

Kirk., R. L., et al.., 1999a. Digital photogrammetric analysis of the IMP camera images: Mapping the Mars Pathfinder landing site in three dimensions. *J. Geophys. Res.*, 104(E4), pp. 8869–8887.

Kirk, R. L., Barrett, J., and Howington-Kraus, E., 1999b. A database of Viking Orbiter image coverage for cartographic and scientific use. *Lunar Planet. Sci.*, XXX, Abstract #1857, Lunar and Planetary Institute, Houston (CD-ROM).

Kirk, R. L., Howington-Kraus, E., and Archinal, B., 2001a. High Resolution Digital Elevation Models fo Mars from MOC Narrow Angle Stereoimages. ISPRS-ET Working Group IV/9 Workshop "Planetary Mapping 2001", online at http://wwwflag.wr.usgs.gov/USGSFlag/Space/Isprs/Flagstaff2001/abstracts/Isprs_etm_OCT01_kirk_mars_moc_stereo. pdf.

Kirk, R. L., Howington-Kraus, E., and Archinal, B., 2002. Topographic analysis of candidate Mars Exploration Rover landing sites from MOC Narrow-Angle stereoimages. *Lunar Planet. Sci.*, XXXIII, Abstract #1988, Lunar and Planetary Institute, Houston (CD-ROM).

Kirk, R. L., Howington-Kraus, E., and Rosiek, M., 2000a. Recent planetary topographic mapping at the USGS, Flagstaff: Moon, Mars, Venus, and beyond, *Int. Arch. Photogramm. Remote Sens.*, XXXIII(B4), 476 (CD-ROM).

Kirk, R. L., Thompson, K. T., Becker, T. L., and Lee, E. M., 2000b. Photometric modelling for planetary cartography. *Lunar Planet. Sci.*, XXXI, Abstract #2025, Lunar and Planetary Institute, Houston (CD-ROM).

Kirk, R. L., Thompson, K. T., and Lee, E. M., 2001b. Photometry of the martian atmosphere: An improved practical model for cartography and photoclinometry. *Lunar Planet. Sci.*, XXXII, Abstract #1874, Lunar and Planetary Institute, Houston (CD-ROM).

Malin, M. C., et al., 1992. Mars Observer Camera. *J. Geophys. Res.*, 97, pp. 7699–7718.

Malin, M. C., et al., 1998. Early views of the martian surface from the Mars Orbiter Camera of Mars Global Surveyor, *Science*, 279, pp. 1681–1685.

Malin, M. C., and Edgett, K. S., 2001. Mars Global Surveyor Mars Orbiter Camera: Interplanetary cruise through primary mission. *J. Geophys. Res.*, 106(E10) pp. 23,429–23,570.

McEwen, A. S., et al., 2002. HiRISE: The High Resolution Imaging Science Experiment for Mars Reconnaissance Orbiter. LPS XXXIII, Abstract #1163, Lunar and Planetary Institute, Houston (CD-ROM).

Miller, S.B. and Walker, A. S., 1993. Further developments of Leica digital photogrammetric systems by Helava. *ACSM/ASPRS Annual Convention and Exposition Technical Papers*, 3, pp. 256–263.

Miller, S.B. and Walker, A.S., 1995. Die Entwicklung der digitalen photogrammetrischen Systeme von Leica und Helava. *Z. Photogramm. Fernerkundung*, 1/95, pp. 4–16.

Neumann, G. A., Rowlands, D. D., Lemoine, F. G., Smith, D. E., and Zuber, M. T., 2001. The crossover analysis of Mars Orbiter Laser Altimeter data. *J. Geophys. Res.*, 106(E10) pp. 23,753–23,768.

Rosiek, M. R., Kirk., R., Hare, T., and Howington-Kraus, E., 2001a. Utilizing Mars Digital Image Model (MDIM) and Mars Orbiter Laser Altimeter (MOLA) data for photogrammetric control. ISPRS-ET Working Group IV/9 Workshop "Planetary Mapping 2001", online at http://wwwflag.wr.usgs.gov/USGSFlag/Space/Isprs/Flagstaff2001/abstracts/isprs_etm_OCT01_rosiek_mars_photogrammetry.pdf.

Rosiek, M. R., Kirk., R., Hare, T., and Howington-Kraus, E., 2001b. Combining lunar photogrammetric topographic data with Clementine LIDAR data. ISPRS-ET Working Group IV/9 Workshop "Planetary Mapping 2001", online at http://wwwflag.wr.usgs.gov/USGSFlag/Space/Isprs/Flagstaff2001/abstracts/isprs_etm_OCT01_rosiek_moon_topography.pdf.

Seidelmann, P. K., et al., 2002. Report of the IAU/IAG working group on cartographic coordinates and rotational elements of the planets and satellites: 2000. *Celest. Mech. Dyn. Astron.*, 82, pp. 83–110.

Smith, D. E., et al., 1999. The global topography of Mars and implications for surface evolution, *Science*, 284, pp. 1495–1503.

Smith, D. E., et al., 2001. Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *J. Geophys. Res.*, 106(E10) pp. 23,689–23,722.

Soderblom, L. A., Kirk., R. L., and Herkenhoff, K. E., 2002. Accurate fine-scale topographic profiles in the martian south polar region from MOLA-Calibrated photometric modeling of MOC NA images. Lunar Planet. Sci. XXXIII, Abstract #1254, Lunar and Planetary Institute, Houston (CD-ROM).

Torson, J. and Becker, K., 1997. ISIS: A software architecture for processing planetary images. *Lunar Planet. Sci.*, XXVIII, pp. 1443-1444, Lunar and Planetary Institute, Houston.

Turcotte, D. L., 1997. *Fractals and Chaos in Geology and Geophysics*. Cambridge Univ. Press, New York.

Zuber, M. T., et al., 1992. The Mars Observer Laser Altimeter investigation, *J. Geophys. Res.*, 97, pp. 7781–7797.

Table 1. MOC Stereo Image Pairs of Past US Mars Landing Sites

Site	Image	Lon (°)	Lat (°)	Res (m)	ε (°)	ι (°)	Sun Az (°)	VO Res (m)
Viking 1	SP1-23503	48.30	22.59	2.6	31.4	50.5	53.4	16
	SP1-25403	48.30	22.49	2.5	22.3	54.7	43.9	
Mars Pathfinder	SP1-25603	33.50	19.30	3.2	30.7	56.1	38.0	38
	SP1-23703	33.60	19.20	2.6	21.4	50.9	46.6	
Mars Polar Lander	M11-01286	-76.96	200.15	1.4	0.0	55.1	228.3	130
	M11-01563	-76.65	195.55	1.4	1.3	54.9	228.2	
	M11-03519	-76.67	195.69	1.8	29.7	69.5	246.5	

In Tables 1 and 2 Lon, Lat are planetographic/west longitude and latitude, Res is resolution, ε and ι are emission and incidence angles, Sun Az is azimuth of Sun measured clockwise from east, and VO Res is approximate resoluton of best available Viking Orbiter images at given location.

Table 2. MOC Stereo Image Pairs of MER Landing Sites Mapped to Date

Site	Image	Lon (°)	Lat (°)	Res (m)	ε (°)	ι (°)	Sun Az (°)	VO Res (m)
Athabasca Vallis	M07-05928	205.00	7.64	5.85	0.09	44.01	149.7	130
	E10-02604	205.00	8.70	6.18	17.98	42.23	127.4	
	E05-00197	204.00	9.28	6.61	19.21	41.72	174.9	
	M07-00614	204.00	9.66	5.87	0.01	43.28	156.7	
Eos Chasma	E02-02855	41.47	-13.46	4.3	0.2	48.2	224.8	57
	E04-01275	41.50	-13.46	3.3	21.9	47.0	209.6	
	E02-02855	42.00	-13.47	2.87	0.16	44.28	214.0	
	E04-01275	41.90	-13.62	3.00	17.97	23.33	144.5	
Gusev Crater	E02-00665	184.06	-14.96	2.9	0.2	50.7	235.3	76
	E02-01453	184.05	-14.87	3.3	22.1	48.3	234.6	
	E01-00341	184.00	-15.11	2.85	0.07	52.32	238.7	
	E05-00471	184.00	-15.07	2.96	9.99	43.25	209.8	
Isidis Planitia	E02-01301	275.19	4.64	3.1	13.0	37.4	206.9	53
	E02-02016	275.21	4.66	2.9	0.2	38.6	214.1	
Melas Chasma	E02-00270	77.77	-8.87	2.9	0.2	46.8	219.9	68
	E05-01626	77.76	-8.82	3.0	12.8	40.2	196.0	
	M0-04367	77.40	-8.80	2.84	0.21	37.69	161.0	
	E09-02618	77.20	-8.74	3.01	18.09	32.72	137.2	
	M04-00361	77.80	-8.70	2.85	0.20	41.01	192.1	
	E12-00720	77.60	-8.86	3.01	17.94	22.73	137.6	

Table 3. Slope Statistics for candidate MER Landing Sites

Site	Set	Sub Area	DEM from	Baseline (m)	RMS Bidir Slope (°)	RMS Adir Slope (°)	99% Adir Slope (°)	Correction to 5 m Base	99% Adir Slope@ 5 m	P(Adir≥15°) @ 5 m (%)	Remarks
Athabasca	1	a	PC	5.87	1.26	1.72	5.02	1.020	5.12	0.001	NE of ellipse but similar
		b	PC	5.87	0.94	1.48	3.77	1.057	3.97	0.001	"
		c	PC	5.87	1.25	1.86	4.85	1.019	4.99	0.001	"
Athabasca	2	n	ST	10	3.39	4.72	15.67	1.125	17.64	0.019	S of ellipse, higher standing
Athabasca	3	a	ST	20	2.48	3.45	10.20	1.409	11.64	0.004	
		c	PC	5.87	3.99	5.35	13.79	1.007	13.88	0.006	
		d	PC	5.87	2.66	3.48	8.50	1.010	8.58	0.001	
Eos	1	nc	ST	10	6.27	9.22	34.39	1.092	37.56	0.072	
		nd	PC	3	5.82	7.07	23.50	0.927	22.95	0.029	PC area misses hills
Eos	2	a	ST	10	6.05	7.97	25.26	1.189	30.03	0.087	
		c	PC	2.87	8.10	9.61	28.20	1.005	28.33	0.082	
		d	PC	2.87	10.58	13.82	35.40	1.005	35.57	0.239	PC area dominated by hills
Gusev	1	a	ST	10	2.80	4.93	16.29	1.076	17.53	0.015	Smooth interior of small crater
		c	ST	10	5.63	8.20	24.95	1.066	26.61	0.078	Knobby S of small crater
		d	PC	3	4.20	5.23	15.31	0.982	15.03	0.010	Smooth interior of small crater
		e	PC	3	9.35	11.67	22.30	0.990	31.97	0.163	Knobby S of small crater
Gusev	2	a	ST	10	8.32	11.37	37.58	1.048	37.38	0.157	Gusev 2 areas similar to 1c/e
		b	ST	10	12.75	16.52	48.17	1.049	50.52	0.340	
		c	PC	3	9.00	11.65	30.80	0.989	30.45	0.166	
		d	PC	3	12.23	15.92	42.99	0.985	52.36	0.299	
Hematite	2	a	PC	2.9	4.89	9.45	24.38	0.791	19.29	0.037	Albedo variations, not slopes
		b	PC	2.9	1.25	1.82	4.94	0.946	4.68	0.001	Bland area, typical
		c	PC	2.9	2.21	3.38	9.46	0.933	8.83	0.001	Exposed rougher area
Isidis	1	nb	ST	10	4.66	6.39	25.60	1.202	30.78	0.037	
		nc	PC	3	5.70	7.45	22.32	0.983	21.93	0.027	
		sa	ST	10	4.12	5.80	20.08	1.058	21.24	0.027	
		sb	PC	3	8.49	10.78	31.18	0.987	30.78	0.121	
Melas	1	a	ST	10	2.72	4.86	14.34	1.000	14.34	0.008	Does not resolve dunes
		b	ST	10	1.56	2.66	7.74	1.000	7.74	0.001	"
		c	ST	10	2.43	4.11	12.61	1.000	12.61	0.004	"
		e	PC	3	13.19	15.85	41.37	0.923	38.17	0.289	Dunes resolved!
Melas	2	a	ST	10	9.96	12.89	43.42	1.187	51.52	0.233	Layers
Melas	3	a	ST	10	11.37	14.37	53.80	1.273	68.49	0.274	"

Set refers to MOC stereopair within a given landing site, in the order listed in Table 2, except that sets Athabasca 1 and Hematite 2 are single images used for photoclinometry. Subarea refers to different regions within each DEM selected for slope analysis. PC = photocliometry, ST = stereo. Bdir = bidirectional slope, Adir = adirectional slope (gradient). Remarks indicate reasons for assessing some results (subdued) as not representative of the landing site. Adirectional slopes scaled to 5-m baseline also appear in Fig. 8, where non-representative values are plotted with open symbols.

USGS AND DLR TOPOGRAPHIC MAPPING OF COMET BORRELLY

E. Howington-Kraus[1], R. Kirk[1], L. Soderblom[1], B. Giese[2], J. Oberst[2]

[1]United States Geological Survey, Astrogeology Team, Flagstaff AZ 86001, US (ahowington,rkirk)@usgs.gov
[2]DLR Institute of Space Sensor Technology and Planetary Exploration, D-12489, Berlin, Germany

Commission IV, WG IV/9

KEY WORDS: Borrelly, comet topographic mapping, photogrammetry, softcopy, extraterrestrial mapping

ABSTRACT:

In the fall of 2001, NASA's Deep Space 1 (DS1) probe imaged Comet Borrelly during a flyby encounter. Three of the Borrelly images have geometry suitable to photogrammetrically map the nucleus, which form two stereopairs with an expected precision (EP) of ~410 m and ~670 m each. DS1 team members at the USGS and DLR have independently produced digital elevation models (DEMs) of Borrelly. Automatic stereo-matching algorithms were used by both USGS and DLR, but the USGS DEM was additionally manually edited in stereo. We accomplished a quantitative statistical comparison of the DEMs and found they have a standard deviation of 120 m, which is small compared to the EP above. There are systematic differences in the DEMs attributable to manual versus automatic matching, but neglecting the systematic differences, we estimate the stereomatching error to only 0.20 pixel RMS, which is similar to the level of subpixel matching accuracy obtained in a wide variety of other mapping situations. The resulting DEMs enable a variety of applications such as perspective views, photometric modeling and studies of the energy balance of the nucleus. We hope to use the USGS DEM as a starting point to extrapolate the shape of the hidden side of the nucleus. This would not only let us determine the volume and moments of inertia of the nucleus, but would lead to a calculation of the insolation onto the nucleus averaged over an entire orbit, and thence to a model of the evolution of nuclear shape.

1. INTRODUCTION

On September 22, 2001, NASA's Deep Space 1 (DS1) probe successfully collected data and imaged Comet Borrelly during a flyby encounter (Figure 1). From a distance of ~90,000 to 3556 kilometers from Borrelly's nucleus, DS1's onboard MICAS (Miniature Integrated Camera and Spectrometer) CCD Sensor captured 20 black-and-white images of Borrelly in a span of 90 minutes (Soderblom et al., 2002). A variety of terrains and surface textures, mountains and fault structures, and darkened material are visible over the nucleus's surface (Britt et al., 2002).

Of the Borrelly images, "near_1" paired with "mid_5_4" and "mid_5_3" make up the optimal stereopairs, having the best compromise between increasing resolution and decreasing convergence angle to photogrammetrically map the nucleus at the highest resolution possible. These images were acquired at a range from the camera of 3556, 3962, and 4387 kilometers, respectively (Figure 1), yielding a ground sample distance of ~46.6 m/pixel for near_1, ~52.0 m/pixel for mid_5_4, and ~57.8 m/pixel for mid_5_3. The expected precision (EP) for stereomodel near_1/mid_5_3 (Figure 2) is ~410 m and that for near_1/mid_5_4 is ~670 m. These precisions are for a stereomatching error of 1 pixel, and smaller errors should be obtainable by automatic matching methods, except in very bland areas. This precision is a small fraction of the dimensions (~8x2 km) of the nucleus, indicating that it should be possible to gen-

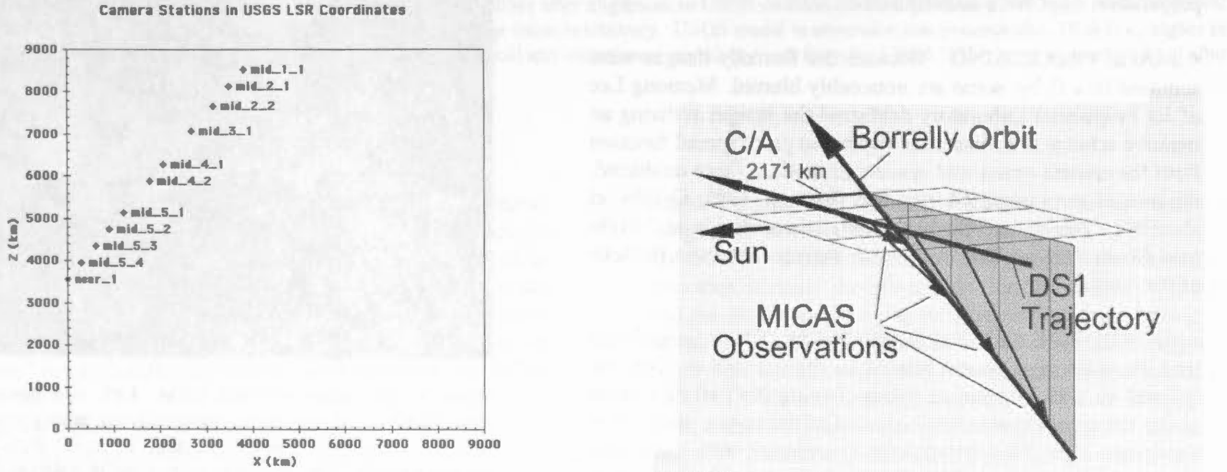


Figure 1. Geometry of DS1 encounter with comet Borrelly. Left, perspective view: spacecraft was looking down on comet nucleus in ecliptic coordinates. Images in this paper are oriented with direction to Sun (shown here by arrow) toward the left. Right, encounter geometry in X-Z plane of Borrelly-fixed LSR coordinates defined for USGS DEM (nucleus is at origin). The Sun is 51° left of the Z axis and ~6° out of the plane of the figure.

erate a relatively detailed stereo digital elevation model (DEM) of Borrelly. DS1 team members at the USGS and DLR have each produced such a model. It is important to note that, though we refer to "elevations" these are not measured in the direction of local gravitational acceleration as commonly understood, or even radially with respect to the center of the Borrelly nucleus. Instead, a Cartesian coordinate system is used, with "elevation" referring to displacement toward the spacecraft, relative to an arbitrary plane. We outline the methodologies used to produce each DEM and present a quantitative statistical comparison and examples of applications.

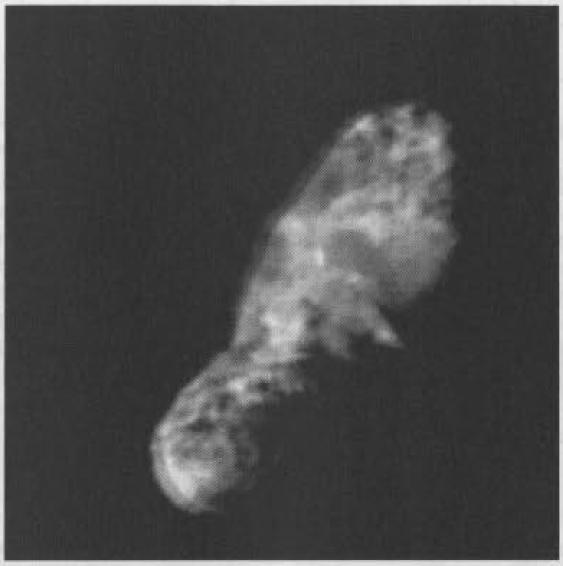


Figure 2. Stereoanaglyph of Borrelly nucleus based on deblurred versions of images near_1 and mid_5_3, which were used for DEM extraction by USGS and DLR. View with red filter on left eye.

2.1 USGS Stereo Analysis

The mapping phase of our work was accomplished on an LH Systems DPW-790 digital photogrammetric workstation running SOCET Set software (© BAE Systems; Miller and Walker, 1993, 1995). Prior to actual mapping, the following preparation steps were accomplished.

- IMAGE PROCESSING: Because the Borrelly images were acquired in a flyby, some are noticeably blurred. Memong Lee of Jet Propulsion Laboratory deblurred the images utilizing an iterative scheme to estimate the combined point spread function from the camera optics and spacecraft motion. Once deblurred, the images were imported into ISIS (Eliason, 1997; Gaddis et al., 1997; Torson and Becker, 1997) where divide and highpass filters were applied in order to sharpen and even the tone of the images.

- COORDINATE SYSTEM DEFINITION: To topographically map Borrelly in a system related to the comet's surface, we defined an LSR coordinate system having X=Y=0 at a point above the approximate center of mass of the comet, and Z=0 at the image limb. This system uses coordinates with the Z-axis toward the near_1 position, agreeing closely with the system used by DLR.

- IMAGE GEOMETRY: The known camera range and phase angle values were converted to the conventions expected by SOCET Set. Specifically, camera position in the LSR coordinate system and camera orientation angles omega, phi, kappa (rotations around X, Y, Z axes consecutively, going from world to camera) were computed.

With the images enhanced, the LSR coordinate system defined, and image geometry converted to photogrammetric terms, we imported the images into SOCET Set where bundle-adjustment and DEM extraction was performed.

The initial image geometry values calculated aimed the camera centers at the approximate LSR origin, therefore an adjustment of the orientation angles was required to aim the camera center at the defined LSR origin and to maximize superimposition of the images. Using SOCET Set's interactive point measurement tool, we measured 16 conjugate image points; of these, 14 were tie points. The remaining two points were a Z-only control point and horizontal-XY control point that established the origin of the LSR coordinates on Borrelly. Bundle-block adjustment was then performed by holding the camera positions and kappa for all images and only allowing angles omega and phi to adjust.

In order to optimize our chances of success in using SOCET Set's automatic matcher when extracting the DEM, it was necessary to adjust the matching strategy parameters to handle the small (155 x 115 pixel) imaged area of the Borrelly nucleus, and to supply an initial DEM "close" to the surface as a starting point for matching. We then collected a DEM over Borrelly at 150 meters/post (Figure 3) which is an elevation measurement at approximately every 3 image pixels of the highest resolution image, near_1. We manually edited the DEM for blunders and then generated an orthoimage of near_1. The image pair near_1/mid_5_4, which has about half the convergence angle of near_1/mid_5_3, was used to check details of the DEM, especially toward the "bottom" of the frame where the parallax is greatest and details are least clear in mid_5_3.

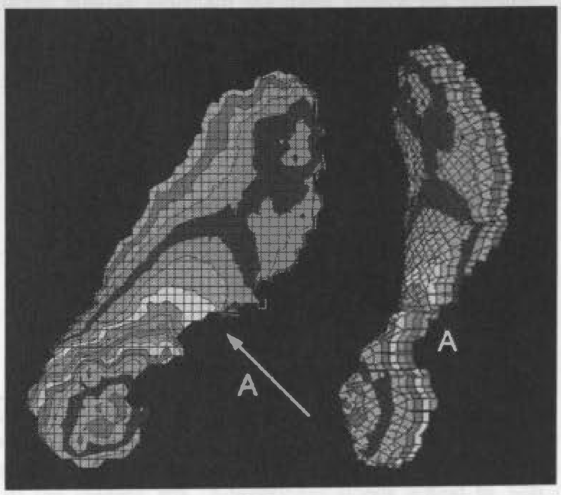


Figure 3. Two views of USGS DEM. Left, view down the Z axis of LSR coordinate system, essentially the same geometry as near_1. Right, view from the direction "A" at 45° from Z axis, no exaggeration. Grid crossings correspond to individual DEM posts with 150 m spacing. Contour interval is 200 m.

2.2 DLR Stereo Analysis

For the purpose of measuring stereo disparities, the deblurred stereo images were radiometrically calibrated and adjusted to have similar sizes, brightness and contrast. Stereo disparities were then measured for a set of 21 individual seed points, using a Zeiss/Phocus stereo system, equipped with a stereo mouse and featuring a 3-D cursor.

In the next step, a computer-based Digital Image Matcher was used to gather disparity points from stereo images near_1 and mid_5_3 (in their original unstretched and not-rotated format) where near_1 was the reference image. The manually collected seed points were used to define the gross perspective distortion of the images and help the matcher start the automatic sampling. Patch sizes of 10–12 pixels were used for the matching, and initial matching runs produced ~8500 disparity points, of which some were off target. After blunder removal, 6990 data points remained.

Owing to the small stereo angle, errors in the estimates for absolute elevations (which are typically made by computing intersection points of the viewing rays) were very large. Instead, the image disparities were converted to relative elevations using a constant conversion factor c, with c = spacecraft range * ifov / stereo angle. Thus, a typical measurement error of 1 pixel in disparity corresponded to an error of 410 m in height (ifov of the MICAS CCD sensor: 13 μrad). However, in some areas, the disparities can be determined to better than 1 pixel, and this error may be smaller. The elevation data were finally interpolated to fill small gaps yielding an estimated accuracy of 500 m horizontally and 150 m vertically.

3. DEM COMPARISON

The DLR and USGS DEMs were collected in different coordinate systems. DLR's DEM is relative to the near_1 image, with an elevation measurement at each near_1 pixel over Borrelly.

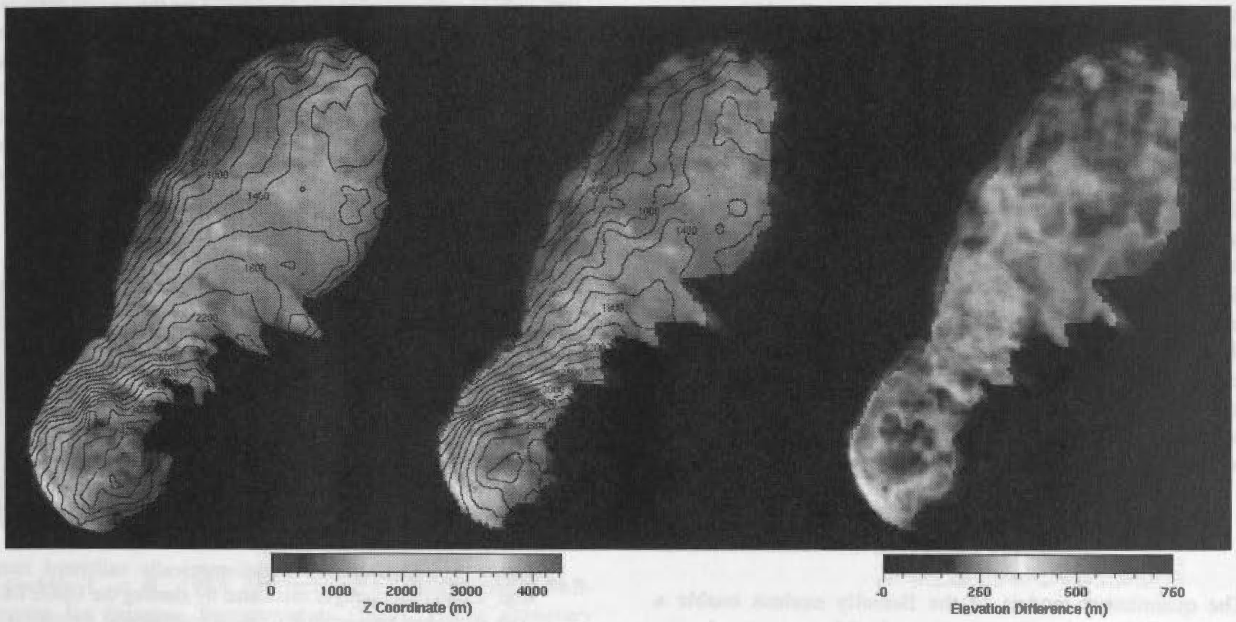


Figure 4. Comparison of USGS (left) and DLR (center) DEMs after alignment in USGS coordinates. Total relief is ~4500 m, contour interval 200 m. Difference (USGS - DLR) is shown at right; average value is arbitrary. USGS model is somewhat less concave than DLR (i.e., higher in the center of the nucleus). This and minor differences in detail and smoothness probably result from the USGS DEM being manually edited after automatic matching.

The USGS DEM is in the above defined LSR coordinates. Therefore, our first step toward comparison was to place the DEMs in the same coordinate system. SOCET Set provides a DEM Registration Tool that scales, translates and rotates the coordinates of one DEM to fit another via an affine transformation, but the DEM coordinates are required to be "close" for successful registration. Unfortunately, our DEMs were not close enough, so we "manually" scaled the DLR DEM to 48 m/post (the approximate resolution of near_1) and rotated it by 90 degrees around the Z axis. The resulting DEM was "close" to the USGS LSR coordinates so that the fine-tuned registration could be accomplished in SOCET Set.

Figure 4 shows both DEMs overlaid on a rectified version of image near_1 in the USGS coordinate system. The two models are very similar, but the USGS model appears slightly smoother (despite its being based on the deblurred images and the DLR model having been resampled from its original format). Elevations relative to an arbitrary zero (approximately the lowest point on the limb) range to approximately 4600 m, with only a small area above 4000 m. There is a trend from low areas at the "top" to high areas > 3 km at the "bottom". Hence, with its size of 8 km along the longest axis, the "bottom" of the nucleus appears to be tilted toward the camera.

The difference between the two DEMs is also shown in Figure 4. Difference values (USGS - DLR) range from -73 to 825 m with a mean of 300 m and standard deviation 120 m. Because the zero of elevation was set arbitrarily, the mean difference is not of great interest, but the standard deviation is surprisingly small compared to our estimates of precision above. There is a systematic pattern to the difference: the two models agree at the ends of the nucleus but the USGS DEM is higher in the center. This region of the surface is relatively smooth, so the difference may reflect the results of manual versus automatic matching on images that do not have adequate texture. Other differences between the DEMs can probably be related to manual vs. automated collection techniques also. The DLR DEM is smoother at the ends of the nucleus, where the USGS model has been edited to capture troughs in the surface (which are visible in the images as topographic shading); in the blander, center portion, the DLR model is rougher (i.e., noiser) whereas the USGS model has been edited to be relatively smooth.

Neglecting the above complexities, we can obtain an estimate of stereo matching error by attributing the 120 m standard deviation of the difference equally to each DEM. This yields the estimate of 85 m RMS range error. Given the image resolutions and convergence angle, this corresponds to only 0.20 pixel RMS stereomatching error, which is similar to the level of subpixel matching accuracy obtained in a wide variety of other situations. Unfortunately, this achieved precision is marginal to measure the relief of apparent topographic features such as mesas seen in the images (Britt et al., 2002) with any confidence, though these features are reflected in the DEMs to some extent as discussed below. The error in local surface orientation estimates, based on a pair of elevation points 150 m apart, will be ~30°. In order to obtain useful estimates of surface orientation for photometric normalization, thermal balance modeling, etc., it is therefore necessary to average over a relatively large number of DEM points (Oberst et al., 2002).

4. APPLICATIONS

The quantitative models of the Borrelly nucleus enable a variety of interesting investigations; in this paper we show a few examples of work in progress based on the USGS DEM. DLR is using their DEM to study the photometric properties of the nucleus (Oberst et al., 2002).

The DEM can be used to render perspective views of the nucleus as seen from any direction. Simulating images from early in the encounter by reprojecting near_1 data is a convenient way to check for errors in the model. Perspective visualizations are not limited to the geometries of the actual flight track, however. Figure 5 shows three views of the nucleus, two with the geometries of actual images and one looking from the "top" of these frames.

Figure 6 compares the observed image with a Lambert-shaded view of the DEM and a view shaded with a more realistic Minnaert model that was fitted (Kirk et al., 2000) to a Hapke model for C-type asteroid Mathilde (Clark et al., 1999), like Borrelly a very dark objec t. The Lambert model gives a better indication of the details of surface shape (as well as noise in the DEM, which was smoothed with a 5x5 lowpass filter before shading) but the Minnaert model, being slightly limb-brightened, is a better match for the image.

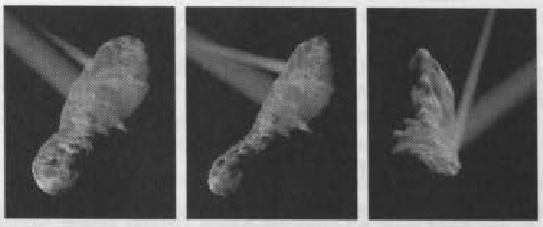


Figure 5. Three perspective views of the nucleus, based on the USGS DEM and near_1 image data. Left, from near_1 camera station; middle, from mid_1_2 station. Right, view from the "top" end of the nucleus (as oriented in the other images). Renderings include schematic representations of the alpha and beta jets (arbitrarily colored yellowish and bluish to distinghish them), based on their inferred 3D geometries (Soderblom et al., 2002). Simulated views like this provide a check on both the DEM and the reconstructed plume geometry.

The correspondence between the models and the image, though imperfect, suggests that some of the image texture is due to relief such as troughs (T) and scarps (S) but some reflects localized albedo variations (A). We are currently pursuing a more quantitative approach to this kind of photometric modeling, including the use of shape-from-shading (photoclinometry) to improve the DEM in bland areas and particularly near the terminator where image radiance is extremely sensitive to surface orientation.

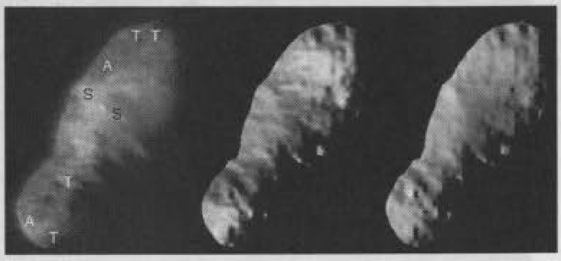


Figure 6. Comparison of radiometrically calibrated image near_1 (left) with images calculated by shading the USGS DEM with a Lambert photometric function (center) and a more realistic, limb-brightened Minnaert function (right). Correspondences (and lack thereof in some areas suggest image shows both topographic features (T=troughs, S=scarps) and local albedo variations (A).

The DEM also has implications for the energy balance of the nucleus. Figure 7 shows the instantaneous insolation and the diurnally averaged insolation, the latter based on the assumption that the alpha jet, whose direction does not change appreciably over time, is aligned with the spin axis of the nucleus (Soderblom et al., 2002). More speculatively, we hope to use the DEM as a starting point to extrapolate the shape of the hidden side of the nucleus. This would not only let us determine the volume and moments of inertia of the nucleus, but would lead to a calculation of the insolation onto the nucleus averaged over an entire orbit, and thence to a model of the evolution of nuclear shape. Obviously, such a program must include a careful assessment of how the results depend on uncertainties in the estimated hidden portion of the nucleus.

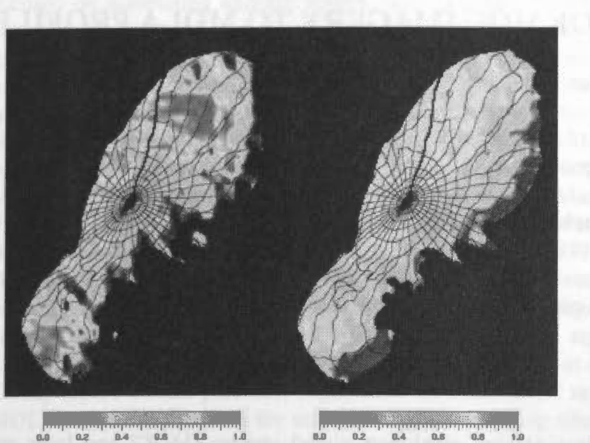


Figure 7. Comparison of instantaneous (left) and diurnally averaged (right) relative insolation onto the nucleus. Cosine of the incidence angle is shown, i.e., model of solar energy input ignores details of photometric function and nonuniform albedo. Overlaid on these models is a 10° (planetocentric) latitude-longitude grid. Coordinate system has polar axis aligned with alpha jet (the assumed axis of rotation (Soderblom et al., 2002)) and prime meridian at the "bottom" end. Location of origin is conjectural, at the approximate center of the nucleus as seen in stereo.

5. REFERENCES

Britt, D. T., et al., 2002. The geology of Comet 19/P Borrelly. *Lunar Planet. Sci.* XXXIII, Abstract #1686, Lunar and Planetary Institute, Houston (CD-ROM).

Clark, B. et al., 1999. NEAR photometry of Asteroid 253 Mathilde. *Icarus*, 140(1), pp. 53-65.

Eliason, E., 1997. Production of digital image models using the ISIS system. *Lunar Planetary Sci.*, XXVIII, pp. 331-332, Lunar and Planetary Institute, Houston.

Gaddis, L. et al., 1997. An overview of the Integrated Software for Imaging Spectrometers (ISIS). *Lunar Planetary Sci.*, XXVIII, pp. 387-388, Lunar and Planetary Institute, Houston.

Kirk, R. L., Thompson K. T., Becker T. L., and Lee E. M., 2000. Photometric modeling for planetary cartography. *Lunar Planet. Sci.* XXXI, Abstract #2025, Lunar and Planetary Institute, Houston (CD-ROM).

Miller, S. B. and Walker, A. S., 1993. Further developments of Leica digital photogrammetric systems by Helava. *ACSM/ASPRS Annual Convention and Exposition Technical Papers*, 3, pp. 256-263.

Miller, S. B. and Walker, A. S., 1995. Die Entwicklung der digitalen photogrammetrischen Systeme von Leica und Helava. *Z. Photogramm. Fernerkundung*, 1/95, pp. 4-16.

Oberst, J., Giese B., Soderblom L., and the DS1 Science Team, 2002. The nucleus of Comet Borrelly: A study of morphology and surface brightness. *Lunar Planet. Sci.* XXXIII, Abstract #1716, Lunar and Planetary Institute, Houston (CD-ROM).

Soderblom, L. A., et al., 2002. Encounter with Comet 19P/Borrelly: Results from the Deep Space 1 Miniature Integrated Camera and Spectrometer. *Lunar Planet. Sci.* XXXIII, Abstract #1256, Lunar and Planetary Institute, Houston (CD-ROM).

Torson, J. and Becker, K., 1997. ISIS: A software architecture for processing planetay images. *Lunar Planetary Sci.*, XXVIII, pp. 1443-1444, Lunar and Planetary Institute, Houston.

PHOTOGRAMMETRIC REGISTRATION OF MOC IMAGERY TO MOLA PROFILE

Jie Shan D. Scott Lee Jong-Suk Yoon

Geomatics Engineering, School of Civil Engineering, Purdue University, West Lafayette, IN 47907-1284, USA
jshan@ecn.purdue.edu

Commission IV, Working Group IV/9

KEY WORDS: Mars, Laser altimetry, Image registration, Ground control, Space mapping

ABSTRACT

The current Mars Global Survey (MGS) mission has been collecting high resolution digital images (MOC) and laser ranges (MOLA). The processing MOLA range data results in a global digital elevation model at ground spacings 1/64 degrees. However, many Mars exploration activities such as landing site selection, rover localization and site-specific studies require higher resolution topographic information. For this purpose, MOC images need to be processed based on photogrammetric principle along with MOLA data and other available ancillary data. As the first and also fundamental step, ground control needs to be selected for MOC images. This paper reports on our study on controlling MOC images by registering them to MOLA data. An approach is proposed which projects each MOLA footprint to MOC images using their exterior orientation parameters obtained from ancillary information about the time, location and orientation of the MOC images. Principles and technical details are presented. Using this approach, a stereopair of one of the selected candidate landing sites are associated to one MOLA profile. Consistent test results are obtained among the two images in the stereopair. Through this registration, further precision processing of MOC images is made possible.

1. INTRODUCTION

The surveying of Mars topography has been continuing for about three decades. The first Mars surveying missions could be dated back to Mariner 6 and 7 spacecrafts in early 1970's. This surveying was continuously updated as images from Mariner 9 and Viking Orbiters were obtained [Davies, et al, 1983; Zeitler et al, 1999]. As described by [Davies, 1999a], the first Mars control network is calculated by using nearly three thousand images along with trajectory and pointing data (the so called RAND network). Wu et al [1984] refine and augment the first results by using analytical photogrammetric triangulation. The refined control network (also called RAND-USGS network) is 3-D and has been used for the production of numerous Mars topographic products such as geometrically corrected Mars images, e.g., MDIM (Mosaicked Digital Image Model) 1.0 and 2.X [Kirk, et al, 2000], which are being used for many Mars related scientific investigations. To further refine the Mars control network, a joint effort from RAND and USGS recently started to recalculate the early control network with new sub-pixel image measurements from over four thousand MDIM images and precise reseau coordinates for correction of systematic image errors [Kirk et al, 1999; Davis, et al, 1999b]. Primary results of this RAND-USGS joint effort indicate that the estimated standard deviation of image measurements is 10.01 μm [Davies, et al, 1999a] or ~200m on the Martian surface [Davies, et al, 1999b]. Similar effort is made by [Zeitler et al, 1999] for local Mars topography. A combined 3-D bundle adjustment of Pathfinder images reportedly improves the precision to an estimated ~740m in each coordinate component for the Pathfinder landing site. [Zeitler et al, 1999].

The next generation of Mars topographic surveying is recently made possible by Mars Global Surveyor (MGS). Since 1999, the mission started returning high resolution digital images and range measurements (http://mars.jpl.nasa.gov/mgs/). Orbiting at a polar and circular orbit of altitude ~400km, MGS contains a mapping payload of two types of equipment: MOLA (Mars Orbital Laser Altimeter) and MOC (Mars Orbital Camera). MOLA calculates the range from MGS to Mars surface by measuring the round trip time a laser beam travels. Post processing of ranges based on orbital trajectory data and geodetic data can yield the three dimensional coordinates of the laser footprints [Neumann, 2001]. It is estimated that accuracy of elevations derived from MOLA data is ~30-40m [Smith et al, 1998]. Similar results are reported by [Zuber, et al, 1998] which indicate that MOLA elevation accuracy is ~5-30m, an increase of ~2 orders of magnitude over the previous network. MOLA topography data has been used to analyze the Mars topography in the northern hemisphere [Smith, et al, 1998, 2001] and northern polar region [Zuber, et al, 1998]. In these efforts, a 1km grid MOLA DEM (Digital Elevation Model) has been generated to analyze the regional topography. The global Mars topography was not known until May 1999 when a 1°×1° DEM with ~13m estimated elevation accuracy is released by [Smith et al, 1999]. Presently finer global DEMs up to 1/256°×1/256° interval are being produced by the MOLA scientific team.

MOC consists of three pushbroom digital cameras: one blue and one red wide-angle (WA) low resolution camera, and one narrow angle (NA) high resolution camera. The MOC WA cameras take images at 230-500m pixel resolution, whereas the MOC NA camera provides images at 1-5m pixel resolution. More results have been reported recently on using MOC images for Mars topographic mapping. A photoclinometric approach or shape from shading is used by [Beyer and McEwen, 2002; Kirk et al, 2002] to analyze the topographic characteristics for the selected candidate landing sites of Mars Exploration Rover in 2003. Similar analyses are carried out by using stereo pairs based on photogrammetric principles [Kirk, et al, 2001b, 2001c; Kirk, et al, 2002; Ivanov and Lorre, 2002]. The first global mosaic is produced using MOC WA images as reported by [Caplinger, 2002].

A primary difficulty in Mars topographic surveying is the acquisition of precision ground control. Current efforts are focused on associating existing results, such as DEM, MDIM, or the control network as well as newly collected MOC images to MOLA profile data, which provides accurate 3-D coordinates of discrete points on Mars surface. An early study shows local MOLA spots have a ~10-20km displacement in longitude from MDIM images [Smith et al, 1998]. Muller et al (1999) report a vertical offset up to ~1.8km between DEM generated from Viking Orbiter images and MOLA data. A link of Mars Pathfinder landing site images to the existing control network reveals a planimetric difference of ~4-5km [Zeitler, et al, 1999]. Despite all these discrepancies with previous studies, recent investigation suggests good consistency between MOC images and MOLA data. Anderson and Parker (2002), Ivanov and Lorre (2002), Kirk et al (2001a; 2001b, 2002), Soderblom et al (2002) report on the precision registration between MOC and MOLA data with focus on the selected candidate landing sites. The registration is achieved via a refined camera calibration model, instrument calibration parameters in SPICE kernel and trial and error studies. Archinal et al present (2002) their initial results on improving the RAND-USGS photogrammetric control network by using MOLA-derived radii and DIMs to improve control point absolute radii and horizontal positions. Using the extracted orientation information of MOC images from SPICE kernels, Kirk et al (2001b) generate high resolution DEMs over the selected candidate landing sites from MOC NA images. In addition, ortho images are produced for regional areas of interest using MOLA derived elevation data [Niedermaier, et al, 2002; Wählisch et al, 2002].

In this paper, we will present our current progress on precision registration of MOC images with MOLA ground profile. A new approach is proposed which projects each 3-D point in a MOLA profile onto the MOC images using pushbroom photogrammetric principals based on the extracted orientation information from SPICE kernels. The approach consists of several steps. First, information about MOC image acquisition is extracted based on their acquisition time. This information includes the location, orientation and scanning rate of MOC images. Second, corresponding MOLA range measurements are located from PEDR files and converted to 3-D coordinates referenced to the Mars surface (footprint). Third, these derived MOLA footprints are then projected onto MOC images using pushbroom photogrammetric principles so that the correspondence of each MOLA footprint to its MOC image point is determined. This process establishes a one-to-one correspondence between each MOLA footprint and its MOC image and provides the image coordinates for all MOLA footprints associated with the images. In this way, each MOLA footprint can be used as a ground control point and at least one MOLA profile with several dozens or about a hundred of footprints are associated with MOC images. This precision registration will allow us to carry out various geospatial processing of MOC images, including bundle adjustment, DEM and orthoimage generation, camera calibration, and the relative alignment of MOLA and MOC instruments.

The rest of the paper is organized as follows. Section 2 discusses the MOC imagery and the acquisition of its exterior orientation parameters from MOC image. Results are presented to show the information needed for subsequent photogrammetric processing. MOLA data is described in Section 3 where the approach to calculating ground elevation from MOLA range measurements is presented. MOLA footprints are overlaid with MOC image using existing tools developed by USGS. Section 4 proposes our approach that projects MOLA footprints on to MOC image using the derived MOC orientation parameters, followed the presentation and analysis of the test results. Section 5 concludes the paper by summarising the major findings and prospecting future studies.

2. MOC IMAGE AND ITS ORIENTATION

As described earlier, two types of cameras are aboard the MGS spacecraft. Because of its high resolution of up to ~1m, only MOC NA camera images are used in this study. Like most satellite mapping systems, the position and orientation of the MOC NA camera are tracked and archived for later use. This information is stored in corresponding SPICE kernels sorted by orbit number and can be retrieved via calculation using the SPICE library. SPICE stands for S – Spacecraft, P – Planet, I – Instrument, C - "C-matrix" and E - Events and contains comprehensive data for the MGS mission, including MGS ephemeris, time and time offsets, camera rotation angles, instrument calibration constants, Mars ellipsoid parameters and many other necessary data needed for processing MOLA and MOC data. It is prepared and provided by NASA Navigation and Ancillary Information Facility (NAIF).

In this study, we have successfully developed SPICE Library-based routines that access the corresponding SPICE kernels, and extract and calculate the necessary spacecraft locations and attitudes of MOC NA camera scan lines based on the acquisition time provided in the MOC image header. The calculation process involves the following steps. First, image acquisition time is converted to ephemeris time and spacecraft on-board time. A leapsecond kernel and a SPICE spacecraft clock kernel (SCLK) are used for this conversion. State vectors (position and velocity) of MGS relative to IAU-MARS frame are then calculated for each time epoch recorded in the MOC image header file by using ephemeris time and SPK kernels. The third step is to calculate the MOC orientations relative to Mars by using C-kernel and I-kernel. Once these quantities are obtained, the following collinearity equation can be established for each MOC scan line

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{Mars} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{MGS} + s\mathbf{R}_{MOC}^{Mars} \begin{bmatrix} x \\ y \\ -f \end{bmatrix}_{MOC} \tag{1}$$

where

$[X \ \ Y \ \ Z]_{Mars}^{T}$ are the Mars ground coordinates in Mars fixed coordinate system;

$[X \ \ Y \ \ Z]_{MGS}^{T}$ are the MGS coordinates in Mars fixed coordinate system;

$[x \ \ y \ \ -f]_{MOC}^{T}$ are the image coordinates in MOC image space;

f: focal length of MOC NA camera;

s: scale factor associated with a Mars ground point;

$\mathbf{R}_{MOC}^{Mars}$: the rotational matrix from MOC to Mars, which is represented as a sequential product of two rotational matrices

$$\mathbf{R}_{MOC}^{Mars} = \mathbf{R}_{MGS}^{Mars}\mathbf{R}_{MOC}^{MGS} \tag{2}$$

where

$\mathbf{R}_{MGS}^{Mars}$: rotational matrix from MGS to Mars;

$\mathbf{R}_{MOC}^{MGS}$: rotational matrix from MOC to MGS.

It should be noted that in Equation 1, the MGS coordinates, MOC image coordinates and the rotational matrix from MGS to Mars are time dependent, so each MOC scan line will have its own parameters. Such obtained parameters (location and orientation) are then used as exterior orientation elements in the subsequent photogrammetric processing of MOC images. Table 1 lists the calculated exterior orientation parameters for a MOC NA stereo pair.

Table 1. Exterior orientation parameters of a MOC stereopair

	Left image		Right image	
Ext. Orien.	Start line	End line	Start line	End line
X (km)	2638.162	2640.863	2743.679	2749.133
Y (km)	-2549.221	-2557.746	-2453.358	-2450.626
Z (km)	-891.875	-860.101	-895.114	-836.828
ω(deg)	-8.6296	-8.6392	11.3701	12.1810
φ(deg)	-75.9467	-75.9452	-75.9648	-76.8692
κ(deg)	-124.5666	-124.5345	-126.576	-125.6728

The exterior orientation for the central scan line of the image is calculated. In this way, the exterior orientation elements for any scan line x_i is written as

$$\begin{aligned} X_i &= X_0 + k_1 x_i & \omega_i &= \omega_0 + k_4 x_i \\ Y_i &= Y_0 + k_2 x_i & \varphi_i &= \varphi_0 + k_5 x_i \\ Z_i &= Z_0 + k_3 x_i & \kappa_i &= \kappa_0 + k_6 x_i \end{aligned} \tag{3}$$

where, Xo, Yo, Zo, ωo, φo, κo are the exterior orientation of the centerline of a MOC image, and $k_1 \ldots k_6$ are the changing rates of exterior orientation in the image. Equation 1 and 2 will be used to determine the image location of MOLA footprints. In this way, MOC images will be registered with and controlled by MOLA data and further precision processing can be pursued.

3. MOLA PROFILE AND ALIGNMENT WITH MOC

MOLA measures ranges from MGS to Mars ground surface at ground spacing ~340m along MGS orbit. Based on orbit trajectory data, initial post-processing of the original measurements can yield orthometric heights for Mars. The parameters used to compute Mars elevations is provided by the MOLA team in the form of PEDR (precision experiment data record) volumes. The PEDR includes the raw shot planetary radii (distance from Mars center of mass to MOLA range point on surface of Mars), the areoid radii (radius of reference areoid with 3396 km mean equatorial radius), the areocentric latitude and longitude of the frame mid-point on the Mars surface (along with deltas to calculate position of each MOLA shot), and associated correction factors (e.g., crossover residuals for planetary radii). The elevation of each MOLA shot, after all corrections have been applied, is

$$h = plrad - areoid \tag{4}$$

h: elevation of MOLA footprint or ground track
plrad: shot planetary radius (corrected)
areoid: areoid radius

Computer routines are developed based on existing tools by using the SPICE Library, PEDR accumulative index, PEDR reader. With these routines, we are able to find the corresponding MOLA PEDR file associated with a MOC image, calculate the areodetic coordinates (longitude, latitude, height) of MOLA ground track and convert them to Cartesian areocentric coordinates (X, Y, Z) for use in controlling MOC images. Table 2 shows a portion of extracted and calculated areocentric coordinates of MOLA footprints (ground track). The coordinates were computed by extracting the latitude, longitude, and elevation coordinates into an ASCII table using the PEDR2TAB utility, and then converting the coordinates to Cartesian format using the SPICE Library routine GEOREC.

Table 2. Cartesian areocentric coordinates of MOLA footprints

X (km)	Y (km)	Z (km)
2469.471	-2191.971	-792.691
2469.504	-2192.054	-792.363
2469.540	-2192.134	-792.035
2469.596	-2192.310	-791.380
2469.625	-2192.397	-791.052
2469.659	-2192.479	-790.724
2469.701	-2192.552	-790.395
2469.764	-2192.720	-789.740
2469.845	-2192.987	-788.758
2469.876	-2193.071	-788.430

In order to acquire the most MOLA data to obtain sufficient control for MOC images, an exhaustive search is carried out to find all MOLA profiles that are associated with a MOC image. This effort turns out to be a success. For the stereopair of Eos Chasma, one of the selected candidate landing sites, two MOLA tracks located were found. As a primary check on the consistency of MOLA and MOC data, the MOLA profile is overlaid with the MOC images. This is accomplished using the USGS developed ISIS (Integrated Software for Imagers and Spectrometers) package. ISIS recomputes the latitude and longitude of the MOLA ground coordinates using the same SPICE kernels used in the geometric processing of the MOC image to obtain consistency. The characteristics of this approach have been studied and its potential needs to be further explored. The initial overlay of MOLA elevation profile and MOC image is shown below in Figure 1, where the fine green line shows the elevation and the circular dots indicate the MOLA footprints.

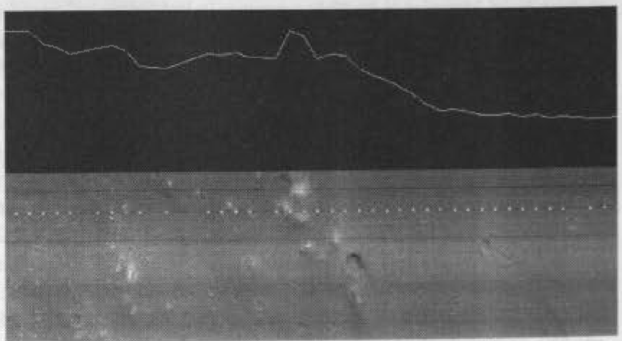


Figure 1. MOLA profile overlaid with a MOC image (Eos Chasma. The image is 90 degree rotated and only a portion is shown in a zoomed-in mode.)

4. PHOTOGRAMMETRIC REGISTRATION

The purpose of MOLA and MOC registration is to determine the location of a MOLA footprint on the MOC image. In this way, every MOLA point can then be used as ground control point for the subsequent MOC photogrammetric processing. To do so, we design and implement an iterative approach that projects a MOLA footprint to the MOC image. For each MOLA footprint, we first treat MOC as a frame camera with constant exterior orientation parameters. This set of parameters is taken from the ones for the central scan line of the MOC image. A MOLA footprint is then projected on to MOC image using this initial exterior orientation with Equation 5. The x coordinate along the flight direction is then used to update the exterior orientation parameters with Equation 3. In the next step, the updated exterior orientation is used to calculate the image coordinate correction dx with Equation 5. This process stops once the correction dx coordinate approaches zero. This calculation steps are shown as below.

Step1. Let $X_i = X_0, Y_i = Y_0, Z_i = Z_0$,

$\omega_i = \omega_0, \varphi_i = \varphi_0, \kappa_i = \kappa_0$

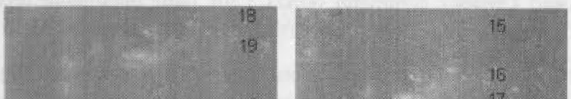
Step 2. Calculate (x, y) coordinates using the collinearity equation

$$x = -f\frac{a_1(X-Xi)+b_1(Y-Yi)+c_1(Z-Zi)}{a_3(X-Xi)+b_3(Y-Yi)+c_3(Z-Zi)}$$
$$y = -f\frac{a_2(X-Xi)+b_2(Y-Yi)+c_2(Z-Zi)}{a_3(X-Xi)+b_3(Y-Yi)+c_3(Z-Zi)} \qquad (5)$$

Step 3. Update the exterior orientation parameters using Equation 3.

Step 4. Use Equation 5 to calculate the x coordinate correction (dx) with the updated orientation parameters.

Step 5.Update the x approximate coordinate $x=x+dx$, if dx is beyond a given threshold, then return to step 3; otherwise the iteration terminates.



a. Image E02 b. Image E04

Figure 2. One MOLA profile (green dots) overlaid on a MOC stereopair (Eos Chasma area. Only a portion is shown in a zoomed-in mode. The labelled numbers are MOLA footprint number.)

Using the above approach, one MOLA profile is projected to two images of Eos Chasma area, one from E02 and one from E04 extended missions. The results are presented in Figure 2, with MOLA footprint numbers labelled beside its location. As can be seen from this figure, the projections of the MOLA profile are in general consistent. Initial visual check suggests the discrepancy of the two projections is at ~1-2 MOLA footprint spacing. Since the two projections are obtained independently using the exterior orientation parameters of the two images, this result also reveals the good consistence of orbital trajectory data (SPICE kernels) based on which the exterior orientation of the images is derived. It should be noted that this registration is conducted for the original MOLA footprint 3-D coordinates calculated using SPICE and PEDR. Every MOLA footprint is found its corresponding locations on the MOC images and therefore can be used as ground control for subsequent MOC image processing. Another distinction of this approach is that no knowledge on MOC and MOC instrument (internal) alignment is needed. This is made possible because of the use of precision photogrammetric geometry and the available exterior orientation. Taking advantage of this, an analytical approach may be developed to determine this alignment via in-flight self-calibration. Despite the aforementioned distinctions, this approach depends on the precision interior orientation of the MOC camera. In our initial tests, only normal focal length and aspect ratio are used. Further study will consider lens distortion and other calibrated values if available.

5. CONCLUSION

Accurate and sufficient ground control is of a necessity for using MOC images to generate DEM and control network. By using trajectory data and sensor geometry, MOC images can be registered to a MOLA profile at a precision of ~1-2 MOLA footprint spacing. This in turn suggests the good consistence of trajectory data. Through the use of the proposed approach, every MOLA footprint can be used as a ground control point in subsequent MOC image photogrammetric reduction, which is our current and future research focus.

REFERENCES

F.S. Anderson and T.J. Parker, 2002. CHARACTERIZATION OF MER LANDING SITES USING MOC AND MOLA. Lunar and Planetary Science XXXIII.

B. A. Archinal, T. R. Colvin , M. E. Davies, R. L. Kirk 1 , T. C. Duxbury, E. M. Lee , D. Cook, and A. R. Gitlin, 2002. MOLA-CONTROLLED RAND-USGS CONTROL NETWORK FOR MARS, Lunar and Planetary Science XXXIII.

Caplinger, M.A., 2002. Mars orbiter camera global mosaic, The 33rd Lunar and Planetary Science Conference, March 11–15, 2002, League City, TX.

Davies, M., Katayama, F., 1983, The 1982 control net of Mars, Journal of geophysical Research, Vol. 88, pp.7403-7404.

Davies M., Colvin T. et a l, 1999a. Status of the RAND-USGS control network of Mars, Second Workshop on Mapping of Mars, California Institute of Technology, U.S.A.

Davies, M. E., Colvin T. et al, 1999b. The RAND-USGS control network of Mars and the Martian prime meridian. Eos Trans. AGU (suppl.), 80, p. F615.

Ivanov, A.B., Lorre, J.J., 2002. ANALYSIS OF MARS ORBITER CAMERA STEREO PAIRS, Lunar and Planetary Science XXXIII.

Beyer, R. A. and McEwen, A. S., 2002. Photoclinometry measurements of Meter-Scale Slopes for the Potential Landing Sites of the 2003 Mars Exploration Rovers, http://www.lpi.usra.edu/meetings/lpsc2002/, The 33rd Lunar and Planetary Science Conference, March 11–15, 2002, League City, TX, USA.

Kirk R., Becker K. et al, 1999. Revision of the Mars control net and global digital image mosaic. Second Workshop on Mapping of Mars, California Institute of Technology, U.S.A.

Kirk, R. L., E. Howington-Kraus, and M. Rosiek, 2000, Recent planetary topographic mapping at the USGS, Flagstaff: Moon, Mars, Venus, and beyond, Int. Arch. Photogramm. Remote Sens., Vol. XXXIII, Part B4, pp.476-490.

Kirk, R. L., Becker, T. L., Eliason, E. M. , Anderson, J., Soderblom, L.A., 2001a. GEOMETRIC CALIBRATION OF THE MARS ORBITER CAMERAS AND COALIGNMENT WITH MARS ORBITER LASER ALTIMETER, Lunar and Planetary Science XXXII (2001), Huston, TX. USA.

Kirk,R.L., Howington-Kraus,E., Archinal,B.A., 2001b. HIGH RESOLUTION DIGITAL ELEVATION MODELS OF MARS FROM MOC NARROW ANGLE STEREOIMAGES, ISPRS WG Workshop, October, Flagstaff, AZ, USA.

Kirk R. L. Howington-Kraus E. Archinal B. A., 2002. Topographic Analysis of Candidate Mars Exploration Rover Landing Sites from MOC Narrow Angle Stereoimages, The 33rd Lunar and Planetary Science Conference, March 11–15, 2002, League City, TX.

Muller, J-P, Kim, J-R, 1999. Assessment of published Viking Orbiter DEMs derived from stereo photogrammetry using MOLA data, Second Workshop on Mapping of Mars, California Institute of Technology, U.S.A.

Neumann, G., Lemoine, F., Rowlands,D. et. al, 2001. Crossover analysis in MOLA data processing. Journal of Geophysical Research, Vol. 106, No. E10, pp.23753-23768.

G.Niedermaier, M.Wählisch, S.van Gasselt, F.Scholten, F.Wewel, T.Roatsch, K.-D.Matz and R.Jaumann, A Topographic Image Map of the MC-18 Quadrangle „Coprates " at 1::2,000,000 using Data obtained from the Mars Orbiter Camera and the Mars Orbiter Laser Altimeter of Mars Global Surveyor. http://www.lpi.usra.edu/meetings/lpsc2002/, The 33rd Lunar and Planetary Science Conference, March 11–15, 2002, League City, TX.

Smith, D., Zuber,M. et al, 1998. Topography of the northern heimsphere of Mars from Orbiter Laser Altimeter. Science, 279, pp.1686-1692.

Smith, D., Zuber,M. et al, 1999. The Global Topography of Mars and Implications for Surface Evolution. Science, May 28, pp. 1495-1503.

Smith, D., Zuber , M., et al, 2001, Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars, Journal of Geological Research, Vol. 106, No. E10, pp.23689-23722.

Soderblom, L. A., Kirk, R. L., and Herkenhoff, K.E., 2002. ACCURATE FINE-SCALE TOPOGRAPHY FOR THE MARTIAN SOUTH POLAR REGION FROM COMBINING MOLA PROFILES AND MOC NA IMAGES, Lunar and Planetary Science XXXIII.

M.Wählisch,G.Niedermaier,S.vanGasselt, F.Scholten, F.Wewel, T.Roatsch, K.-D.Matz, R.Jaumann, 2002. A NEW DIGITAL ORTHOIMAGE MAP OF THE MARTIAN WESTERN HEMISPHERE USING DATA OBTAINED FROM THE MARS ORBITER CAMERA AT A RESOLUTION OF 256 PIXEL/DEG, http://www.lpi.usra.edu/meetings/lpsc2002/, The 33rd Lunar and Planetary Science Conference, March 11–15, 2002, League City, TX.

Wu, S.C, Schafer, F.J., 1984. Mars control network. Technical papers of the 50th annual meeting of the American Society of photogrammetry, Vol.2, pp.456-463.

Zeitler, W., Oberst, J., 1999. The Mars Pathfinder landing site and the Viking control point network, Journal of Geophysical Research, Vol. 104, No. E4, pp.8935-8941.

Zuber,M., Smith, D. et al. 1998. Observations of the north polar region of Mars from the Mars orbiter laser altimeter. Science, Dec. 11, Vol.282, pp.2053-2060

ACKNOWLEDGEMENT

This study is supported by the National Aeronautical and Space Administration. The authors would like to thank their collaborators Drs. Charles Acton (JPL), Randolph Kirk (USGS) and Gregory Neumann (GSFC) for their invaluable technical support on accessing, understanding and using the MGS data, SPICE kernels and Library.

HIGH PRECISION LANDING SITE MAPPING AND ROVER LOCALIZATION BY INTEGRATED BUNDLE ADJUSTMENT OF MPF SURFACE IMAGES

Kaichang Di [a], Fengliang Xu [a], Ron Li [a], Larry H. Matthies [b], Clark F. Olson [c]

[a] Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University,
470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210-1275 - (li.282, di.2, xu.101)@osu.edu
[b] Jet Propulsion Laboratory, California Institute of Technology, Mail Stop 125-209, 4800 Oak Grove Drive,
Pasadena, CA 9110 - lhm@telerobotics.jpl.nasa.gov
[c] University of Washington, Bothell, Computing and Software Systems
18115 Campus Way NE, Box 358534, Bothell, WA 98011 - cfolson@u.washington.edu

Commission IV, WG IV/9

KEY WORDS: Mars Pathfinder Mission, Landing Site Mapping, Rover Localization, Bundle Adjustment

ABSTRACT:

High-precision topographic information from all available data is crucial to many landing site geological and engineering applications. At the same time, precise navigation and localization of the rover as it traverses the Martian surface is important both for its safety and for the achievement of its engineering and scientific objectives. In this paper, we investigate a new approach to high-precision Mars landing site mapping and rover localization based on integrated bundle adjustment of an image network built by linking ground-based images with automatically or manually selected tie points. The method and software are tested using lander and rover image data obtained from the Mars Pathfinder mission. An innovative method for automatic tie point selection is also presented.

1. INTRODUCTION

High-precision global and local topographic information is crucial for support of engineering operations and achievement of scientific goals in Mars exploration missions. In particular, landing site mapping will be extremely important for landing and rover navigation in future lander and rover missions such as the 2003 Mars Exploration Rovers (MER), the European Beagle 2 lander of the 2003 Mars Express, the 2007 Smart Lander and Long-range Rover, and sample return lander missions beyond 2010.

In Mars rover missions, accurate navigation and localization of the rover relative to the landing center are needed so that the rover can traverse the Martian surface safely, return repeatedly to the same location to perform operations, and support coordinated multidisciplinary high-precision scientific experiments. In the Mars Pathfinder (MPF) mission, the rover Sojourner principally used a heading sensor and wheel counters for localization. Sojourner accumulated 1m location errors in an area of about 10m x 10m during this mission. In the MER 2003 mission, rovers will traverse an extended distance of 600m up to 1,000m where the terrain may be more challenging. In the 2007 Smart Lander and Long-range Rover mission, the rover traverse will be even longer. More accurate rover localization will be highly desirable for achievement of the scientific objectives of future missions. In order to prevent the accumulation of localization errors, descent and orbital images should be used as well to provide global constraints.

Since 1998, the Mapping and GIS Laboratory at The Ohio State University and the NASA JPL Machine Vision Group have been jointly developing a bundle adjustment method with relevant techniques for the processing of Mars descent and rover images for rover localization and landing site mapping.

We missed an opportunity to test the concept using actual MARDI (Mars Descent Imager) data because of the failure of the Mars Polar Lander Mission. In order to verify our algorithm and software, field tests were conducted at Silver Lake, CA in April 1999 and May 2000. Based on the data obtained, various rover localization experiments were carried out. Using descent and rover images along with either an integrated or an incremental adjustment, rover localization accuracy was reached of approximately 1m for a traverse length of 1km from the landing center (Li et al., 2000, 2001; Ma et al., 2001). Experiment results also showed that it is feasible to localize the rover by using rover images only if no descent images are available (as with the MER mission), and yet still achieve a similar accuracy when issues such as optimal traverse design and image network generation are considered (Di et al., 2002).

In addition to using simulated descent and rover images, we tested our methods and software with actual Mars data. We downloaded rover and lander IMP (Imager for Mars Pathfinder) images from the PDS (Planetary Data System) web site. The German Aerospace Center (DLR) also provided us with a complete panorama chosen from a vast mount of IMP images. In the following section we briefly describe the bundle adjustment models used in this investigation and our other relevant studies. In subsequent sections, the processing of IMP panoramic images and rover images are presented.

2. BUNDLE ADJUSTMENT MODELS

The basic model for the bundle adjustment is based on the well known collinearity equations (Wolf and Dewitt, 2000). The linearized observation equation is expressed in matrix form as:

$$V = AX - L \tag{1}$$

In this bundle adjustment model, all of the unknowns (camera position, orientation of all the images, and 3D ground position of the tie points) are adjusted together after all the images are acquired. Therefore we call it the integrated bundle adjustment model.

Because there is no absolute ground control available on the Martian surface, the adjustment is a free net adjustment where the normal matrix is rank deficient. We used the Singular Value Decomposition technique to solve the normal equation in which the Minimum Norm is applied using the Least Squares principle. If certain distinctive features (such as mountain peaks or craters) can be observed from the orbital images, they are used as relative controls for the adjustment in the following form:

$$HX = W \tag{2}$$

In order to improve computational efficiency and make the bundle adjustment applicable to a real-time operation, we developed two incremental bundle adjustment models (Li et al., 2000, 2001; Ma et al., 2001). Model I is represented as:

$$\begin{aligned} V_{m-1} &= A_{m-1}X_{m-1} - L_{m-1}, \\ V_m &= A_m X_m + B_m Y_m - L_m \end{aligned} \tag{3}$$

Model II is represented as:

$$\begin{aligned} V_{m-1} &= A_{m-1}X_{m-1} + B_{m-1}Y_{m-1} - L_{m-1}, \\ V_m &= B_m Y_m + C_m Z_m - L_m \end{aligned} \tag{4}$$

In Model I, the unknowns are added gradually into the adjustment system when the new observations are available. After the step *m* adjustment, unknowns solved by step *m-1* are also updated so that the final results are ultimately the same as those from an integrated bundle adjustment model. In Model II, with the gradual increase of new unknowns some older unknowns are no longer taken into account, thus offering more flexibility and efficiency to the process.

There are several frames of reference that are used in the MPF image pointing data, including the camera head coordinate system, lander (L) frame, Martian Local Level (M) coordinate frame, Mars Surface Fixed (MFX) coordinate frame, and Landing Site Cartographic (LSC) coordinate system. Bundle adjustment and topographic products of our models are based on the LSC system as defined by the U.S. Geological Survey.

We developed a program to convert the pointing data of the PDS images to the exterior orientation data. This is accomplished by a chain of translations and rotations through the above frames of reference. Converted exterior orientation values were then used as initial values in the bundle adjustment.

3. PROCESSING OF IMP IMAGES

We used 60 pairs of IMP stereo images provided by DLR to explore automatic tie point selection, bundle adjustment, DEM (Digital Elevation Model) and orthoimage generation. These IMP stereo images form a complete panorama (as shown in Figure 1), in which the azimuth and the tilt angles of the images are represented in the x and y axes, respectively.

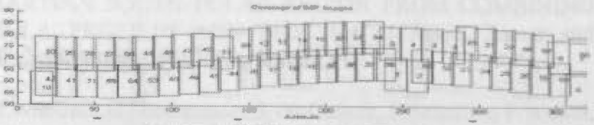


Figure 1. IMP image coverage

We developed a series of techniques to automatically select tie points within one stereo pair (intra-stereo) and between adjacent stereo pairs (inter-stereo).

For intra-stereo tie point selection, the Förstner operator is applied to extract interesting points. Then, an area-based matching is applied to match these interesting points using normalized correlation coefficients. Experiments with this data set show that in most cases over 90% of the matches can be found. Figure 2 shows the matched (white cross) and unmatched (black cross) interesting points in an intra-stereo pair. In this stereo pair, 319 and 312 interesting points were extracted from the left and right images, respectively, and 216 points were matched.

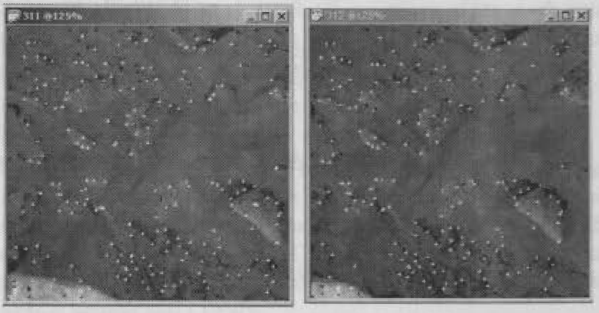


Figure 2. Interesting point matching

Within a set of matched points, some mismatched points may still exist that need further verification. Automatic verification of the matched tie points is based on the consistency of parallax. Assuming that the terrain does not vary significantly, parallaxes of the points should maintain an order throughout the image. In other works, if two points in the left image are located at the top left and bottom right corners, respectively, then their matches should preserve the same order (e.g. measured by a distance). Disordered points are considered as mismatches.

The order can be expressed simply as a function of parallaxes along the row direction, since in the column direction the parallaxes are almost a constant (close to zero). In Figure 3, we order the interesting points in the row direction and then draw their x (left) and y (right) parallax curves in the same order. Any abrupt changes of the parallaxes are assumed to represent mismatches. Mismatches are eliminated using a median filter to produce the smoother parallax curves found in Figure 4. These smooth parallax curves represent the correct order, and are used as reference parallaxes.

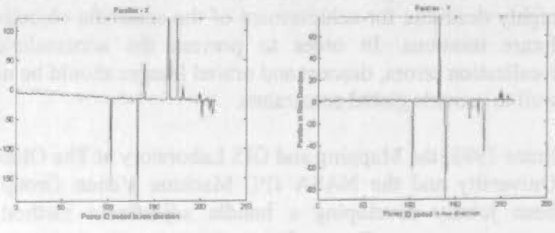


Figure 3. The x and y parallax curves

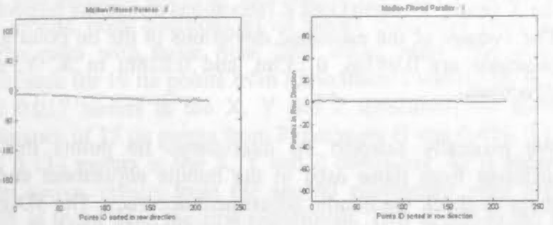


Figure 4. Smoothed x and y parallax curves

For each matched interesting point pair, the actual parallax is compared to the value of the reference parallax of the same row. If it is within a specified range (empirical values used are 10 pixels in x and 5 pixels in y), the pair is considered a correct match. 201 matches (93%) passed this parallax verification. A visual check was conducted to confirm the result.

To prepare an even distribution of the tie points for use in the later bundle adjustment and for improvement of computational efficiency, the final tie points are selected from the set of correct matches using a gridding scheme. Each image is evenly segmented into several patches, e.g., a 3 x 3 grid. Within each patch, the pair of correctly matched interesting points with the highest value of grayscale variance is chosen. Since this is the most distinguishable matched point within the patch, it can serve as a reliable tie point. In Figure 5, 9 pairs of final tie points have been selected for each stereo image pair.

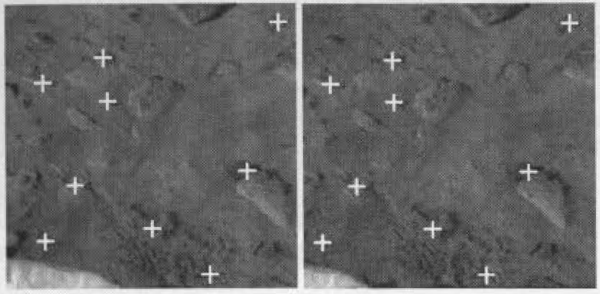


Figure 5. Nine evenly distributed final tie points are selected for each intra-stereo pair

Application of this strategy is much more difficult when selecting tie points between cross-stereo pairs or inter-stereo pairs, because disparities can often be very large. Or there may even be no overlapping across the stereo image pairs. In general, overlapping between the image margins should be very small (around 10%).

To get around this problem for inter-stereo pair images in the panorama, their relative orientations can be estimated to an acceptable level of precision by using information about the rotation of the cameras relative to the landing base. We can calculate 3D coordinates for matched tie points from the previous step and expect no large inconsistency. Figure 6 shows the distribution of 4,709 intra-stereo interesting points (matched and verified).

The 3D ground points are then used to interpolate a coarse DEM that, hopefully, will not differ very much from the final DEM. Figure 7 shows the generated coarse DEM. This is a global DEM for the entire panorama. More importantly, it can be used to bridge the stereo image pairs.

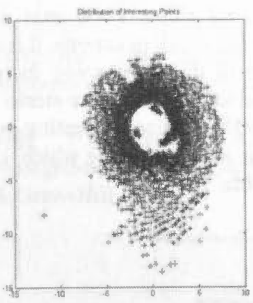


Figure 6. Distribution of all intra-stereo interesting points

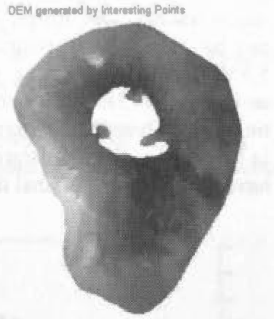


Figure 7. The Coarse DEM

With the coarse DEM, we can predict the overlapping of inter-stereo images. In addition, we can predict for an interesting point the approximate location of its conjugate point in its corresponding inter-stereo pair image. For each interesting point, we can find its 3D coordinates in the DEM and back project it onto its inter-stereo image. The back projected location should not be too far from the actual position. A small search range can thus be used around the back projected location. In this way, most interesting points in the inter-stereo pair margins can be matched. The success rate is 90% for most cases. Verification is accomplished by again using the row-based reference parallax method. Figure 8 illustrates 4 final tie points that were selected in an inter-stereo image pair after matching, verification and gridding selection.

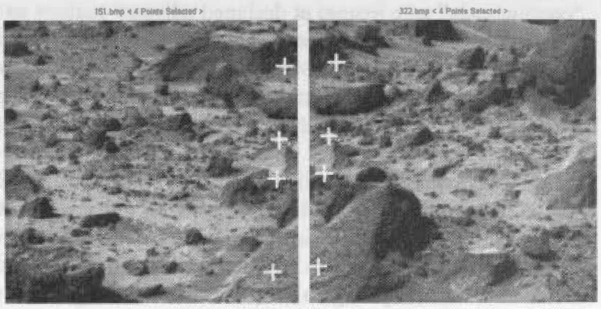


Figure 8. Selected final tie points in an inter-stereo pair

Difficulties can occur when the overlapping margins are too small and/or radiometric characteristics of the images differ greatly (e.g., when one image was taken under dim conditions while the other was taken in bright sunlight), or shadows have moved significantly from changing sun angles. However, most inter-stereo images are taken within a short time interval and the above challenge can be avoided.

Inter-stereo image pairs can be linked by three types of connections: in the horizontal, vertical, and/or diagonal direction. Because of the difficulties mentioned above, it cannot be assumed that every inter-stereo image pair will find all 3 links. If all the image pairs within a panorama can be linked by at least 1 of the 3 types, they can form a network and be bundle adjusted to estimate improved precision exterior orientation parameters. The refined parameters are then used to generate a high precision DEM.

The final links between the stereo pairs that were found through the automatically selected tie points are shown in Figure 9. Just as Figure 1, the x and y axes in Figure 9 stand for the azimuth and tilt angles of the IMP images. In total 84 pairs of inter-stereo images were linked by the tie points. Each inter-stereo pair has 6 selected tie points. For the entire panorama, there are 5,859 matched interesting points in the overlapping margins, among which 589 have been selected as final inter-stereo pair tie points. Overall, there are 24,903 matched interesting points in both intra- and inter-stereo pair images, among which 1,128 have been selected as final tie points.

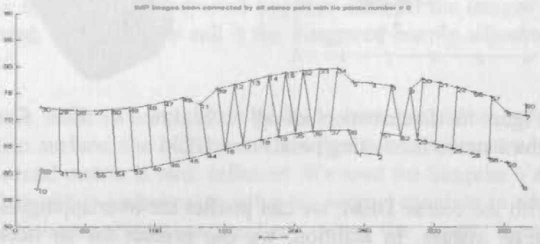


Figure 9. Connection map of the panoramic images

We are currently checking the quality of these tie points. We will then perform the bundle adjustment of the complete panorama with the automatically selected tie points and generate a precise DEM and orthoimage. Final mapping results will be presented at the symposium.

4. BUNDLE ADJUSTMENT OF IMP LANDER AND ROVER IMAGES

In order to form a rover network, the rover should "continuously" take images at designed adjacent stations so that the images can be linked together. Also the network configuration should be designed according to some network optimization principles to achieve a high accuracy in mapping and rover localization (Di et al., 2002). However, in the MPF mission, the rover did not take images continuously, so we could not test rover localization in a rover image traverse. Even for separate rover images, it is difficult to find sufficient tie points to link them to IMP images. We selected 10 stereo pairs of IMP images and 2 stereo pairs of rover images to test rover localization accuracy. Figure 10 shows the left images of 2 (out of the 10) IMP stereo images that have overlay with the rover images. Figure 11 shows the 2 rover stereo pairs used in the experiment. The rover was about 5 meters from the lander when these images were taken.

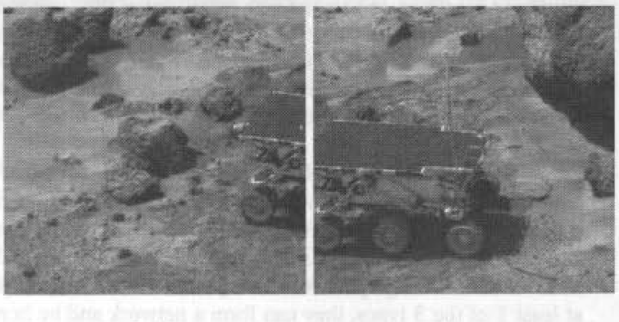


Figure 10. IMP images that have overlay with rover images

At first, we conducted a bundle adjustment by using the 10 IMP stereo pairs and no rover images. 155 tie points are selected to link the 20 images into an image network. The maximum distance from the tie point to the lander was about 25 meters, and the average distance about 6 meters. As a result of the bundle adjustment, the average of the estimated deviation of the lander camera positions were 0.013, 0.018 and 0.013 meters in the X (east), Y (north) and Z (elevation) directions respectively. The average of the estimated deviations of the tie point ground positions are 0.065m, 0.155m, and 0.028m in X, Y, and Z directions.

We manually selected 28 inter-stereo tie points that were different from those used in the bundle adjustment and used them to check the bundle adjustment accuracy. The 3D ground positions of these tie points were calculated by space intersection from one set of stereo pairs. Then we projected the computed ground points back onto the other stereo pairs. The differences between the back projected image points and those identified manually reflect the quality of the exterior orientation parameters. The space intersection and back projection were conducted using the exterior orientation parameters before and after the bundle adjustment. The average differences in x and y were 2.9 and 6.6 pixels before adjustment, and 0.4 and 0.3 pixel after adjustment. Maximum differences were 5.3 and 14.6 pixels in x and y before adjustment, and 1.0 and 0.8 pixel after adjustment. As the image coordinate differences were apparently reduced, this indicates that the accuracies of exterior orientation parameters were improved.

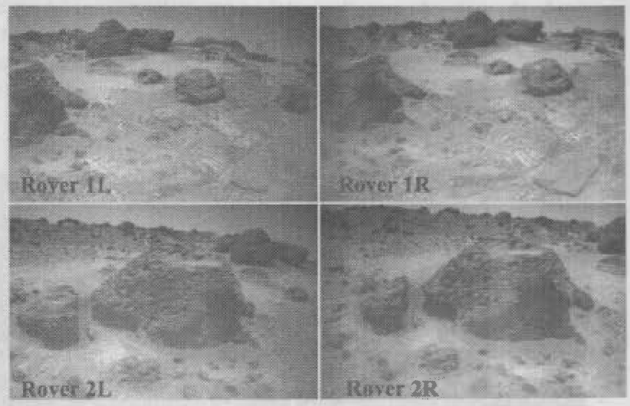


Figure 11. Rover stereo pairs that have overlay with IMP images

Next we conducted an integrated bundle adjustment by using the rover and IMP images together. An image network was built by linking the IMP and rover images: in addition to 155 tie points linked the IMP images, another 15 tie points were manually selected to link IMP and rover images and 20 tie points were selected to link the rover images.

In the image network, the distribution of the tie points between rover Stereo Pair 1 (top row in Figure 11) and IMP images is better than that of Stereo Pair 2 (lower row in Figure 11). For comparison purposes, we carried out 2 bundle adjustment experiments: one using just rover Stereo Pair 1 with all the IMP images and the other using both stereo pairs with all the IMP images.

From the results of the integrated bundle adjustment, we observed that the accuracy of IMP lander image orientation parameters is not significantly changed by adding the 4 rover images. From Experiment I, the average of the estimated deviations of the rover camera positions were 0.074, 0.085 and 0.031meters in the X, Y and Z (height) directions respectively. In Experiment II, the estimated deviations of the rover camera positions were 0.083, 0.081 and 0.035 meters in the X, Y and Z directions. We can see that the rover localization accuracy estimated from the two experiments is very similar (about 2% of the distance from the lander).

We also assessed the rover localization accuracy by comparing the ground point positions of tie points calculated from rover stereo pairs with those from IMP stereo pairs. The average difference for 10 tie points from Experiment I was 0.118, 0.051, and 0.017 meters in the X, Y and Z directions; the average difference of 15 tie points from Experiment II was 0.426, 0.178, and 0.134 meters in the X, Y and Z directions. As is apparent from above, results from the second experiment were not as good as those from the first experiment. This is due to the fact that the geometry between the second rover stereo pair and the lander images is relatively weak.

5. SUMMARY

We propose a landing site mapping and rover localization method based on bundle adjustment of an image network built by linking ground-based images. Integrated and incremental bundle adjustment models and automatic tie point selection methods have been developed. Sixty MPF IMP images of a complete panorama were used to test the effectiveness of automatic intra- and inter-stereo tie point selection, which turned out to be very promising. Ten IMP stereo pairs and two rover pairs were used to test the capabilities of the bundle adjustment. The results demonstrated that the accuracy of exterior orientation parameters of the IMP images was improved, consequently the differences between measured image points and those back-projected from adjacent stereo images were reduced from several to more than ten pixels to a sub pixel level. With this method, a rover can be localized to an accuracy of about 2% of the distance from the lander (depending on the strength of the link between rover and lander images). Mapping products from the complete panorama of IMP images, in particular precision DEM and orthoimage, are currently being processed and will be presented at the symposium.

REFERENCES

Di, K., R. Li, L.H. Matthies, C.F. Olson, 2002. A Study on Optimal Design of Image Traverse Networks for Mars Rover Localization. *2002 ASPRS-ACSM Annual Conference and FIG XXII Congress*, April 22-26, Washington DC. (CD-ROM).

Li, R., F. Ma, F. Xu, L. Matthies, C. Olson and Y. Xiong, 2000. Large Scale Mars Mapping and Rover Localization using Descent and Rover Imagery. *Proceedings of ISPRS 19th Congress*, Amsterdam, July 16-23, 2000 (CD-ROM).

Li, R., F. Ma, F. Xu, L.H. Matthies, C.F. Olson, and R.E. Arvidson, 2001. Localization of Mars Rovers using Descent and Surface-based Image Data. *Journal of Geophysical Research-Planet*, Accepted.

Ma, F., K. Di, R. Li, L. Matthies, and C. Olson, 2001. Incremental Mars Rover Localization using Decent and Rover Imagery. *ASPRS Annual Conference 2001*, April 25-27, St. Louis, MO. (CD-ROM).

Olson, C.F., L.H. Matthies, Y. Yiong, R. Li, F. Ma and F. Xu, 2001. Multi-resolution Mapping Using Surface, Descent and Orbital Imagery. *The 2001 International Symposium on Artificial Intelligence, Robotics and Automation in Space*, Montreal, June 2001.

Xu, F., F. Ma, R. Li, L. Matthies and C. Olson, 2001. Automatic Generation of a Hierarchical DEM for Mars Rover Navigation. *Proceedings of the 3rd Mobile Mapping Conference*, Cairo, Egypt, January 3-5, 2001.

Wolf, P.R., and B.A. Dewitt, 2000. *Elements of Photogrammetry with Applications in GIS*, Third Edition, McGraw-Hill, 608p.

AN ORTHOIMAGE MAP USING DATA OBTAINED FROM THE MARS ORBITER CAMERA OF MARS GLOBAL SURVEYOR

M. Wählisch, G. Niedermaier, S. van Gasselt, F. Scholten, F. Wewel, T. Roatsch, K.- D. Matz, M. Hoyer, R. Jaumann
DLR Institute of Space Sensor Technology and Planetary Exploration, Berlin, Germany
E-mail: Marita.Waehlisch@dlr.de

KEY WORDS: Extraterrestrial, Digital, Mapping, Orthoimage, High resolution, Mosaic

ABSTRACT:

A basic requirement for the planning of future Mars missions are precise and high resolution maps, especially, of the landing site area. We present a new digital orthoimage map of Mars using data obtained from the Mars Orbiter Camera (MOC) of the Mars Global Surveyor (MGS). The new map covers the Mars surface from 180° E (180° W) to 360° E (0° W) and from 60° South to 60° North with a resolution of 231.529 m/pixel (256 pixel/degree). The mosaic was divided into 8 parts, according to the digital size of the Mars Digital Image Mosaics (MDIM2). They are available digitally at http://solarsystem.dlr.de/PG/MOC/. In addition, we announce the release of a printed map of Coprates (MC 18) based on MGS data. For map creation, digital image processing methods have been applied. Furthermore, we developed a general processing method for creating image mosaics based on MOC data. This method can be used for creating image mosaics using CCD (Charged Coupled Device) line camera data and it is applicable also for other Mars missions, whenever a CCD line camera is employed.

1. INTRODUCTION

In this paper, we report our efforts processing MOC wide-angle red images, derived from the orbits M00-M18 in 1999/2000 at a resolution of about 250 m/pixel or less. The data set has two big advantages in comparison to the 30 years old Viking data set. First, the better knowledge of the navigation data: As it was reported in a paper of Smith (2001), the spacecraft pointing has an accuracy of 1 to 3 mrad (400 m to 2000 m, depending on the spacecraft altitude) and absolute spacecraft position uncertainties in order of 100 m. Second, the Mars Orbiter Camera has an 8 bit dynamic range instead of the 7 bit dynamic range of the Viking vidicon. Due to this technical progress, the processing of the images offers the opportunity to get a new map of Mars with better radiometric and geometric quality.

2. INPUT DATA

The MOC wide-angle (WA) camera is a line scanner camera operating in the push-broom mode. In order to map the whole planet, stripes of images (4° longitude range by 17° latitude range) were obtained at the beginning of the MGS mission during the Geodesy Campaign (Caplinger, 2001). When the mapping with the narrow-angle camera started, WA- context images (2° by 2°) were obtained. Inspecting the available images of the red WA camera, it was obvious, that the dynamic range of the recently obtained MOC context images is better than the dynamic range of images of the Geodesy Campaign. To take advantage of the 8 bit dynamic range of the MOC camera, we decided to use not only the long strips of the Geodesy Campaign, but all available context images. We found 4,339 context images and 183 Geodesy images of good quality in the investigated area and with a resolution better than 250 m/pixel. Additionally, we had to use 313 images of the Geodesy Campaign with a resolution > 250 m/pixel and < 435 m/pixel. Approximately 10 % of the visually inspected images were sorted out for lack of quality.

3. METHODS

Image data processing has been performed using multiple VICAR (Video Image Communication and Retrieval) and IDL (Interactive Data Language) programs, developed by the JPL (Jet Propulsion Laboratory), DLR (German Aerospace Center) and the TUB (Technical University of Berlin) (Scholten, 1996). Furthermore, ISIS (Integrated Software for Imagers and Spectrometers software), developed by the USGS (U.S. Geological Survey), was applied (see figure 1).
First, each MOC image was corrected for radiometric camera errors. After visual inspection, some images were edited manually to remove image artifacts (stripes of pixel errors, etc.). Images containing too many artifacts, were not included. The correction of images with major differences in brightness was performed using IDL routines developed at the DLR. After all radiometric and brightness corrections, the images were Mars referenced, geometrically corrected (Kirk, 2001) and orthoprojected using a global Martian Digital Terrain Model (DTM), developed by the DLR and based on MGS Mars Orbiter Laser Altimeter (MOLA) data (Smith, 2001): We used all released MOLA binary data. Reading out these files, we got about 588,000,000 measurements of planetographic latitude, east longitude (referenced to the radii: A=B=3,396.19 km, C=3,376.2 km) and Mars geoid heights. All longitudes were shifted by a value of +0.238 to convert the longitudes from International Astronomical Union (IAU) 1991 (the MOLA reference system) to the IAU 2000 reference system (Seidelmann, 2001). Finally, a gridded DTM from all these object points was computed using DLR/TUB software (Scholten, 1996). The DTM has a resolution of 64 pixel/degree and is Mercator projected. The images were sinusoidal map projected onto this DTM to get orthoimages. As a reference system the IAU 2000 reference system was adopted. For the 0° to 90° W region, the longitude 45° W represents the reference meridian, for the 90° W to 180° W region the longitude 135° W. To eliminate major differences in brightness between the individual images of the mosaics, high- and low-pass filter processing techniques were applied for each image after the map projection.

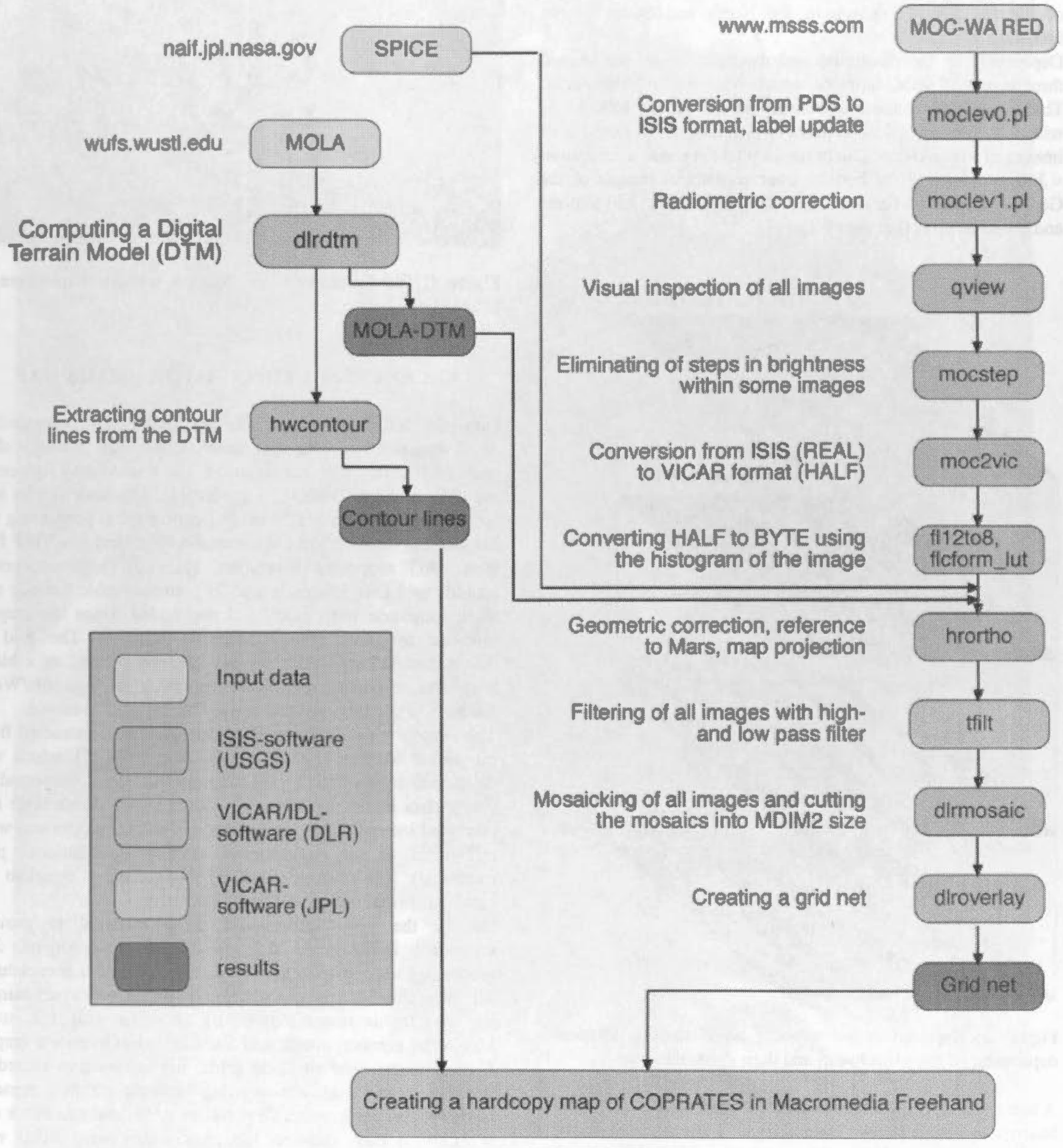

naif.jpl.nasa.gov
SPICE
wufs.wustl.edu
MOLA
Computing a Digital Terrain Model (DTM)
dlrdtm
MOLA-DTM
Extracting contour lines from the DTM
hwcontour
Contour lines
Input data
ISIS-software (USGS)
VICAR/IDL-software (DLR)
VICAR-software (JPL)
results
www.msss.com
MOC-WA RED
Conversion from PDS to ISIS format, label update
moclev0.pl
Radiometric correction
moclev1.pl
Visual inspection of all images
qview
Eliminating of steps in brightness within some images
mocstep
Conversion from ISIS (REAL) to VICAR format (HALF)
moc2vic
Converting HALF to BYTE using the histogram of the image
fl12to8, flcform_lut
Geometric correction, reference to Mars, map projection
hrortho
Filtering of all images with high- and low pass filter
tfilt
Mosaicking of all images and cutting the mosaics into MDIM2 size
dlrmosaic
Creating a grid net
dlroverlay
Grid net
Creating a hardcopy map of COPRATES in Macromedia Freehand

Figure 1: Image data processing scheme

After filtering the images, we mosaicked the images together. No registering or block adjustment techniques were used in order to improve the geometric quality. We recognized that the accuracy of the navigation data has such a good quality, that the orthoimages fit very well to each other except for some images of the Geodesy Campaign in the North and South of the investigated region.

Depending on the resolution and dynamic range, we created three layers of MOC mosaics, which were stacked afterwards: The upper layer consists of context images (orbits M00-M18) with a resolution < 250 m/pixel, the middle layer consists of images of the Geodesy Campaign (orbit M01) with a resolution < 250 m/pixel and the bottom layer consists of images of the Geodesy Campaign (orbit M01) with a resolution > 250 m/pixel and < 435 m/pixel (see figure 2).

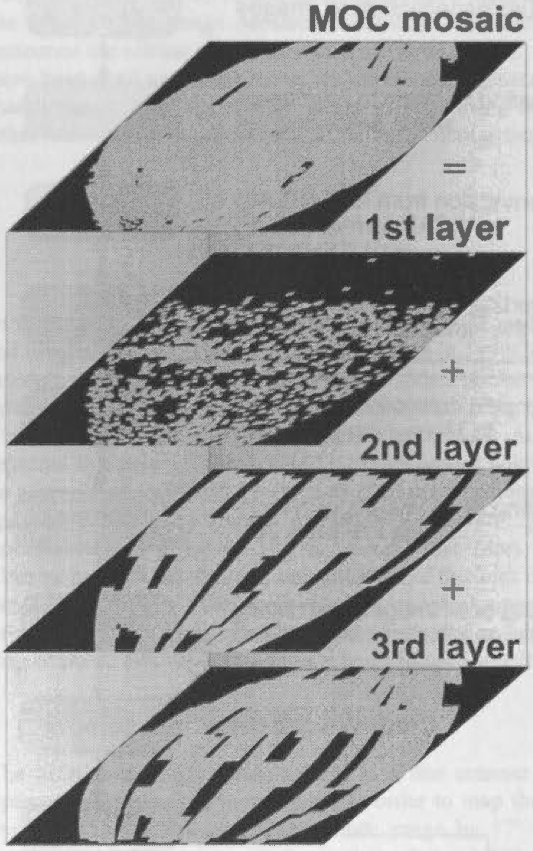


Figure 2: Tree layers of mosaics were stacked afterwards depending on their resolution and their dynamic range.

A few remaining gaps in the coverage were filled with MDIM2, based on 7 bit VIKING-Data. Figure 3 (next page) shows a small part of the Valles Marineris in full resolution. For comparisons with existing maps of Mars, the mosaic was divided into 8 parts, according to the digital size of the MDIM2 (see figure 4). The 8 map parts with a resolution of 231.529 m/pixel (256 pixel/degree) are available digitally at http://solarsystem.dlr.de/PG/MOC/.

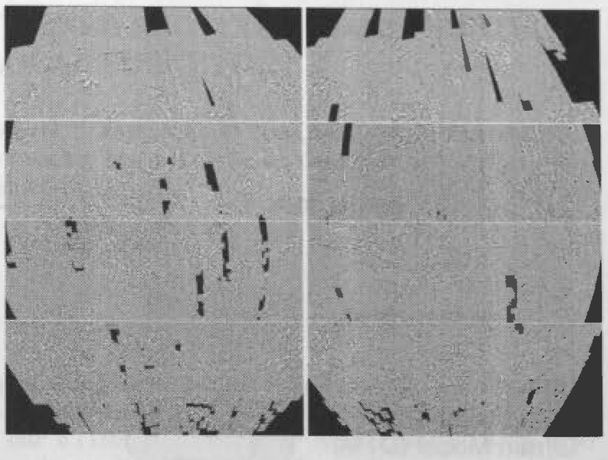


Figure 4: The 8 mosaics of the Martian western hemisphere in MDIM2 resolution.

4. CREATING A TOPOGRAPHIC IMAGE MAP

One part, MC-18 Coprates, was cartographically processed in detail and printed using a commercial oversize plotter with a scale of 1 : 2,000,000 (see figure 5). The printed map represents the left part of the MDIM2 j quadrangle. The scale results in a map field of 0,89 m x 1,33 m. For cartographic processing the mosaic was imported into Macromedia FreeHand as a TIFF file. Both IAU supported coordinate systems: 1) planetocentric latitude and East longitude and 2) planetographic latitude and West longitude were calculated and added, since the map is intended to serve several scientific interests. The grid in planetocentric/East is the primary grid-net printed as a black line. The secondary coordinate system (planetographic/West) has been printed in cyan and is used for historical reasons.

The contour lines calculated for this map were extracted from the global Martian Digital Terrain Model (DTM), which was developed by the DLR. The contour data were imported as vector data into Macromedia FreeHand as a separate layer and corrected interactively. The heights are areoid heights and were referenced to an equipotential surface (gravitational plus rotational). The average radius of this surface is equal to the mean equatorial radius of 3396.2 km.

Besides the basic information, it is essential to provide additional information such as camera data, digital data processing steps, map projection parameters, and nomenclature. The map sheet consists of 6 fields: i) the map field containing the topographic image mosaic, ii) the cover with title, scale, map serial number, author and a shaded relief overview map of Mars with latitude/longitude grids, iii) information regarding the camera and mosaic processing methods within a separate text field, iv) the quadrangle sheet, an additional text block and scale, v) a map showing the gaps which were filled with MDIM2 in the "main map", vi) an outline of the system of coordinates and the imprint.

For layout reasons, the legend was placed in the lower map part. Due to the map size, the best folding system resulted in a final folded map size of 20 cm x 27.8 cm.

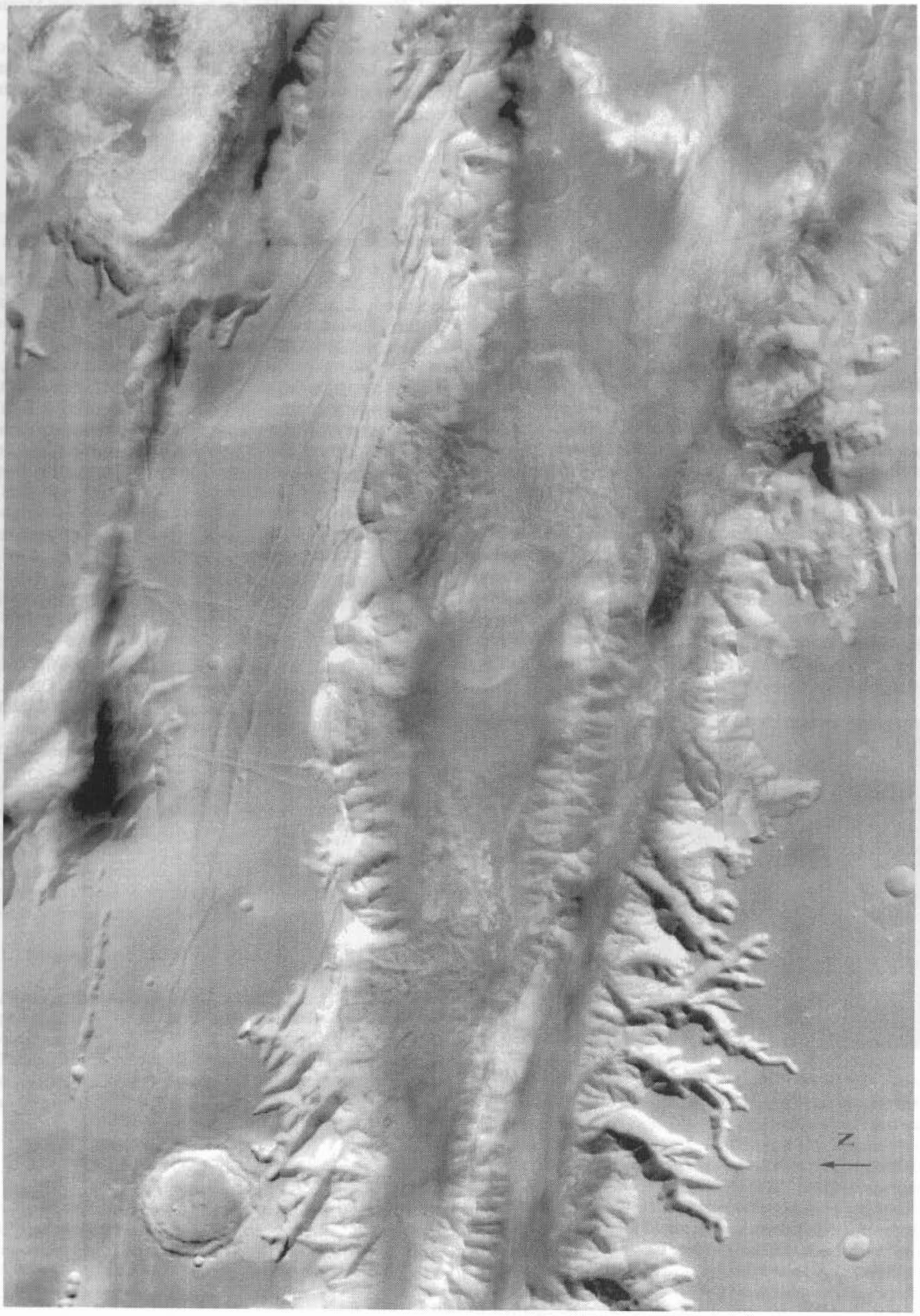


Figure 3: Detail of the orthomap of Melas Chasma in full resolution

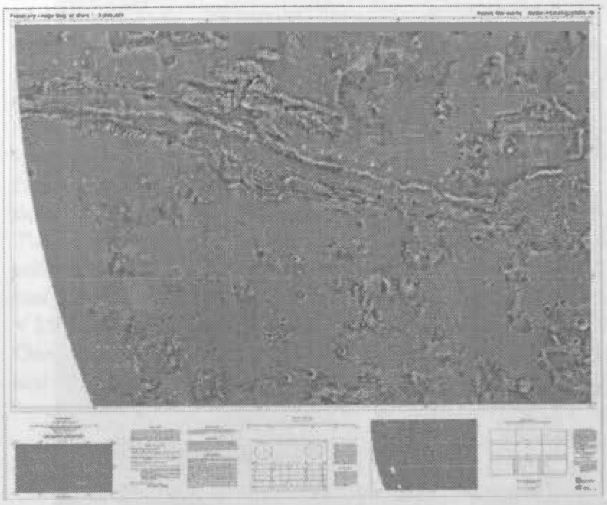


Figure 5: Topographic image map of the MC-18 quadrangle "Coprates" 1:2,000,000.

5. RESULTS AND OUTLOOK

We present a new digital orthoimage map of the Martian western hemisphere with a resolution and map projection parameters, except the radii, according to the MDIM2. The reference system is the new defined IAU 2000. We see good correspondence between MOLA and MOC datasets by merging the MOC mosaic with the MOLA data using IHS-transformation (see figure 6).

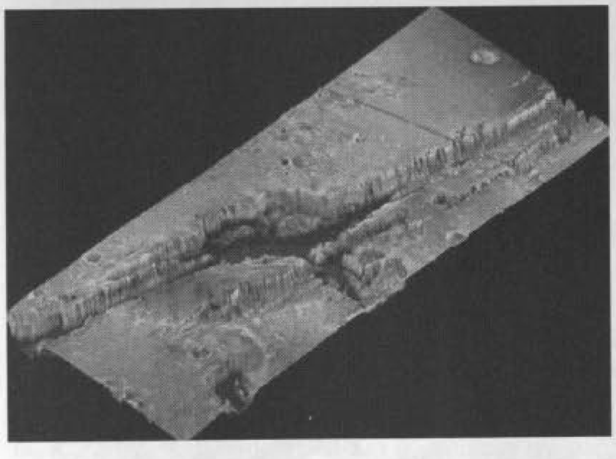


Figure 6: Detail of MOC mosaic merged with MOLA (IHS transformation), Valles Marineris region, North is on the bottom right.

The 8 derived mosaics are new geometrically precise orthoimage maps in MDIM2 resolution. They will be used for the targeting of future lander missions to Mars and in the planning of imaging sequences from orbit, e.g. within the Mars Express mission in 2003. This satellite will carry the HRSC (High-Resolution Stereo Camera), a multiple-line multispectral stereo scanner instrument (Neukum, 1995). The developed method for creating orthoimage maps from line scanner data is also applicable for the HRSC data. Further activity is planned to fill the remaining gaps with MOC Geodesy images of lower resolution (< 435m) or MDIM2 data. It is necessary to process also the Martian Eastern hemisphere and the pole regions depending on the available data in order to get a global Martian mosaic of the same quality.

It is still important to use both datasets for photogeological interpretations due to the difference of the photometric conditions of the MOC and MDIM2 images. The photometric correction of the MOC images still needs to be done.

6. REFERENCES

Caplinger, M.A. and Malin, M.C., 2001, Mars Orbiter Camera geodesy campaign, JGR, Volume 106, pp. 23,595-23,606.

Kirk, R.L., Becker, T.L., Eliason, E.M., Anderson, J., and Soderblom, L.A., 2001, Geometric Calibration of the Mars Orbiter Cameras and Coalignment with Mars Orbiter Laser Altimeter, LPSC XXXII, 1863.

Neukum, G., J. Oberst, G. Schwarz, J. Flohrer, I. Sebastian, R. Jaumann, H. Hoffmann, U. Carsenty, K. Eichentopf, and R. Pischel, 1995, The Multiple Line Scanner Camera Experiment for the Russian Mars 96 Mission: Status Report and Prospects for the Future, Photogrammteric Week 95, Wichmann Press, pp. 45-61.

Scholten, F., 1996, Automated Generation of Coloured Orthoimages and Image Mosaics Using HRSC and WAOSS Image Data of the Mars96 Mission, International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B2, S.351-356, Wien.

Seidelmann, P.K., Abalakin V.K., Bursa, M., Davies, M.E. (died on April 17, 2001), de Bergh, C., Lieske, J.H., Oberst, J., Simon, J.L., Standish, E.M., Stooke, P., Thomas, P.C., 2001, Report of the IAU/IAG/COSPAR Working Group on Cartographic Coordinates and Rotational Elements of the Planets and Satellites: 2001.

Smith, D.E. et al., 2001, Mars Orbiter Laser Altimeter: Experiment summery after the first year of global mapping of Mars, JGR, Volume 106, pp. 23,689-23,722.

wufs.wustl.edu

naif.jpl.nasa.gov

www.msss.com

MARS GEODESY/CARTOGRAPHY WORKING GROUP RECOMMENDATIONS ON MARS CARTOGRAPHIC CONSTANTS AND COORDINATE SYSTEMS

T. C. Duxbury[1], R. L. Kirk[2,*], B. A. Archinal[2], and G. A. Neumann[3]

[1]Jet Propulsion Laboratory, Pasadena, CA (Thomas.C.Duxbury@jpl.nasa.gov), [2]U.S. Geological Survey, Flagstaff, AZ (rkirk@usgs.gov), [3]NASA Goddard Space Flight Center, Green Belt, MD (neumann@tharsis.gsfc.nasa.gov)

Commission IV, WG IV/9

KEY WORDS: Mars, Extraterrestrial Mapping, Coordinate Systems

ABSTRACT:

NASA's Mars Geodesy/Cartography Working Group (MGCWG), established in 1998 and chaired since 2000 by one of us (TCD), consists of leading researchers in planetary geodesy and cartography at such diverse institutions as JPL, NASA Ames and Goddard Centers, Purdue and Ohio State Universities, Malin Space Science Ssystems, the German Center for Aerospace Research DLR, and the US Geological Survey, as well as representatives of the current and future Mars mission teams that are the customers for Mars maps. The purpose of the group is to coordinate the activities of the many agencies active in Mars geodesy and cartography in order to minimize redundant effort and ensure that the products needed by mission customers are generated. A specific objective has been to avoid repeating the experience of the 1970s–80s, when competing researchers produced geodetic control solutions and maps of Mars that were mutually inconsistent. To this end, the MGCWG has recently assembled a set of preferred values for Mars cartographic constants, based on the best available data. These values have been transmitted to the International Astronomical Union and appear in the report of the IAU/IAG Working Group on Cartographic Coordinates and Rotational Elements of the Planets and Satellites as the officially recommended constants for Mars(Seidelmann et al., 2002). The MGCWG has also recommended to NASA that the USGS adopt the IAU-approved coordinate system of planetocentric latitude and east longitude for future maps of Mars, in place of the (also IAU-approved) planetographic system with west longitude positive. This recommendation has recently been approved by NASA. In this paper we present the preferred values for Mars cartographic constants with discussion of the process by which they were derived, then discuss the rationale and implications of the use of east/planetocentric coordinates in future Mars maps.

1. Cartographic Constants

The parameters that must be defined in order to carry out mapping of a planet such as Mars with spacecraft data are the orientation of the spin axis (including precession), rotation rate, rotational orientation at a specified time, and size and shape of an ellipsoidal reference surface. In addition to these parameters and their uncertainties, the IAU/IAG Working Group reports, where available, statistics on the deviation of the planetary surface from an ellipsoidal model.

1.1 Spin Axis and Rotation Period.

Parameters describing Mars' rotation rate and direction of its polar axis in space have been adopted from those values determined by Folkner et al. (1997), based on tracking of Mars Pathfinder and the Viking 1 and 2 landers. The right ascension α and declination δ in degrees at a given time *t* are given by the expressions:

$$\alpha = 317.68143^\circ - 0.1061^\circ/\text{century} * T$$
$$\delta = 52.88650^\circ - 0.0609^\circ/\text{century} * T$$

where T is the number of Julian centuries of *t* from the standard epoch of J2000.0 TDB. The uncertainties in these values estimated by Folkner et al. are 0.00001° in offset and 0.0007°/century in rate for α, and 0.00003° in offset and 0.0004°/century in rate for δ.

The rotation rate of Mars is assumed to be W_dot = 350.89198226°/day, for which Folkner et al. cite an uncertainty of 0.00000008°/day.

1.2 Orientation at Epoch.

The orientation of the prime meridian of Mars in space at a given epoch *t* is specified by the angle W, which is measured along the equator to the east between the 0° (or prime) meridian and the equator's intersection with the celestial equator. It is given (in degrees) by the expression:

$$W = W_0 + W\_dot * d$$

where W_dot is defined as above and *d* is the number of days from *t* to the standard epoch J2000.0. W_0 is the value of W at the standard epoch J2000.0.

Because the location of the prime meridian is defined by the center of the crater Airy-0 (Seidelmann et al., 2002; de Vaucouleurs et al., 1973) it is necessary to estimate W_0 based on the expected position of Airy-0 at J2000.0. The determination of this angle is critical, as any change in this value will result in a change in longitude of any point on the planet whose position has been determined in inertial space. It is therefore desirable to determine a value for this that is as accurate as possible, and yet also a value that is unlikely to be improved on (thus changing longitudes again) in the near future. In the past this has been done as part of photogrammetric bundle block adjustment solutions of Viking

---

[*] Corresponding author.

Orbiter and Mariner 9 images (Davies and Arthur, 1973; Davies and Katayama, 1983; Wu and Schafer, 1984; Davies et al., 1992; Davies et al., 1996; Zeitler and Oberst, 1999); however, the advent of Mars Global Surveyor (MGS) Mars Orbiter Laser Altimeter (MOLA) and Mars Orbiter Camera (MOC) data has provided an alternative and more accurate method of determining the value of W_0. The MOLA observations allow for surface coordinates to be estimated with fairly high accuracy (order 100 m in horizontal position) in the same inertial coordinate system in which the spacecraft position is determined. Matching the MOLA ground track observations to images of the surface will therefore allow the position of surface features shown on the images to be determined at a similar level of accuracy in inertial space. Using this method, we have taken the center position of Airy-0 (measured by M. Caplinger of Malin Space Science Systems to be sample 281, line 5498 in MOC narrow angle image M23-00923; see http://www.msss.com/mars_images/moc/01_31_01_releases/airy0 /index.htm) and then determined the position of this pixel relative to MOLA data obtained at the same time in the area of this image. This comparison made use of the USGS calibration of the relative alignment of the MOC and MOLA instruments to an estimated precision of ~100 m on the ground (Kirk et al., 2001b). From this analysis a value of W_0 = 176.634° was derived. Similarly, we have compared mosaics of high-resolution Viking Orbiter images in the vicinity of Airy-0 with a MOLA derived digital image model (DIM) and determined a value W_0 = 176.627°. Additional work was done to verify that these determinations of W_0 were consistent with the positions of the Mars Pathfinder and Viking 1 lander, as determined in inertial space and on global image mosaics. A final value of W_0 = 176.630° was then chosen as the simple average of these two determinations, and in general agreement with photogrammetric solutions and other checks by others and other working group members (e.g. by M. Davies, T. Colvin, Kirk, W. Zeitler, J. Oberst, and M. Wählisch). Given these comparisons we expect the uncertainty of this value of W_0 (which derives mostly from the difficulty of relating the images to the MOLA ground track or DIM) to be ±0.003° to 0.004°, or about 250 m on the Martian surface. We believe it unlikely that this value can be improved upon until a radio transmitter (lander) can be placed in the vicinity of Airy-0, presumably at some point in the distant future.

1.3 Reference Surface:

In order to derive best-fitting reference surfaces for Mars, a number of least squares fits of MOLA data (as of 2001 Spring) were performed. These solutions were done for spherical, rotational ellipsoid, and triaxial ellipsoid reference surfaces. Possible assumptions for these solutions included: a) using either a sinusoidal projection with 0.5 or 1.0° data spacing, or a cylindrical projection with 0.4° or 0.8° data spacing; b) using data up to ±90°, ±85°, and ±80° latitude; and c) solving for Cartesian offsets of the chosen spheroid from the center of mass or not solving for them. The cylindrical projection solutions were done in order to see the effect of over sampling data near the poles, in comparison to the sinusoidal solutions. The solutions with different latitude limits were done to look at the effect of data (e.g. elevation changes reflecting the permanent polar caps) near the poles. The solutions with and without offsets were done mostly to determine what the difference is in the north and south polar radii, assuming an ellipsoid whose center is offset from the center of mass of Mars. The MOLA data used were points from a grid including from 40726 to 230400 points, depending on the solution assumptions (spacing, projection, and latitude limits). These solutions were done mainly to check what type of variation occurred in the solutions given the different assumptions used and to make sure the basic rotational ellipsoid solution was reasonable in comparison to the others. The solutions of primary interest were ones using all the data (i.e. to ±90°), using the sinusoidal projection (so data near the poles would not be over sampled), and with the densest grid spacing (of 0.5°). Accuracy estimates were made based entirely on the variation between the different solutions. These estimates come to order 100-200 meters, although the formal uncertainties of the individual solutions were at the few meter level. At this estimated accuracy level, fits to a triaxial ellipsoid showed no improvement over those to a rotational ellipsoid.

The results are, for a best fitting sphere with no offsets from the center of mass, a radius of R = 3389.50 ±0.2 km. The best fitting rotational ellipsoid, again with no offsets included, results in a semi-major (equatorial) axis A = 3396.19 ±0.1 km and a semi-major (polar) radius of B = 3376.20 ±0.1 km. If an identical fit is done, but with the polar axis allowed to have an offset from the center of mass, this offset amounts to 3.01 ±0.1 km and the fitted minor axis is unchanged, resulting in a north polar radius of such an offset ellipsoid of 3376.20 - 3.01 = 3373.19 ±0.1 km and a south polar radius of such an offset ellipsoid of 3376.20 + 3.01 = 3379.21 ±0.1 km. For the best fitting rotational ellipsoid, with no offset, the RMS deviation from the spheroid is 3.0 km, the maximum elevation is 22.64 ±0.1 km (Olympus Mons), and the maximum depression is 7.55 ±0.1 km (Hellas). It is unlikely that these values can *ever* be improved upon, given that the uncertainties are now almost entirely due to the fact that Mars' topographic surface simply does not match these geometric reference surfaces at accuracy levels of 0.1 km and better.

The rotational ellipsoid with best-fit dimensions A = 3396.19 km, B = 3376.20 km is being adopted by the USGS as the reference surface for map projections for a variety of Mars image mosaics and maps currently in production and will be used in products generated in the foreseeable future. These mosaics include a final version of the global 1/256° monochrome digital mosaic of Viking Orbiter images (MDIM 2.1) and global 1/256° color and stereo mosaic of MOC wide angle images (MDIM 3.0) (Kirk et al., 1999; 2000; 2001a). A revised global 1/64° color mosaic based on Viking Orbiter data is also a possible future product. It is important to note that the ellipsoid will be used only as the reference surface for defining map projections. Other surfaces can be adopted for other purposes, with no inconsistency. In particular, the detailed MOLA topographic model will be used for accurate projection of images and other remote-sensing data onto the planet, and a spherical-harmonic representation of an equipotential surface will be used as the reference for elevations.

The MOLA team also plans to use the reference surface and other parameters given here in the final archived version of the MOLA dataset. Adoption of a common set of parameters for the altimetry and image-mosaic archives is necessary but not sufficient to ensure direct compatibility of the dataproducts. A common coordinate system must also be adopted.

2. Coordinate System

The International Astronomical Union, meeting in 1970 (IAU, 1971), formalized the use of two types of coordinate systems for planets. The first, sometimes called "planetocentric coordinates," consists of longitude measured positive eastward and planetocentric latitude, defined as the angle between the equatorial plane and a line from the center of the body to a given point. This system is right-handed and identical to spherical polar coordinates as commonly defined; it is used almost universally in cartographic calculations and is generally preferred by geophysicists for display of their data. The other system measures longitude positive in the direction opposite the planetary rotation, so the sub-Earth longitude increases with time, and planetographic latitude. Planetographic latitude is defined as the angle between the equatorial plane and the normal to a spheroidal reference surface at the given point. (If the point of interest is above or below the reference surface, the planetographic latitude is determined for a point on the ellipsoid whose normal passes through the given point.) This coordinate system conforms to earlier astronomical usage for Mars and was extended by the IAU to the other planets and satellites. It has been used for all USGS (and many other) planetary maps to date, including digital and paper maps of Mars (Inge and Batson, 1992). The second IAU-approved system is sometimes called "planetographic coordinates" but we prefer to explicitly include the direction of positive longitude (e.g., "west/planetographic" or just "west/'ographic" for Mars and other bodies rotating prograde) because some software packages use a non-IAU-approved combination of planetographic latitude and east longitude. Similarly, we refer to the other IAU coordinate system as "east/'ocentric".

The east/'ocentric coordinate system has a number of technical advantages such as being right-handed, independent of the definition of reference ellipsoid (or of the height of a point above/below such an ellipsoid), and identical to the spherical coordinate system used in gravitational potential calculations, spacecraft navigation, etc. The primary justification of the west/ 'ographic system has always been the weight of historical usage. A second potential advantage of planetographic latitude is that it is referenced to the local ellipsoid normal, which is close to the local vertical. Thus, on Earth, a surveyor with a leveled theodolite can measure planetographic (geographic) latitude directly to a good approximation. This supposed advantage is largely negated by the absence of surveyors on Mars and by the considerable deviations of the Martian solid surface and equipotential surface from an ellipsoid (so that the local vertical determined by a hypothetical surveyor would not correspond accurately to the ellipsoid normal that defines planetographic latitude). Nevertheless, the importance of consistency over mere technical advantages led to the use of west/'ographic coordinates in all Mars maps produced from 1970 through the late 1990s.

From a practical standpoint, the situation was radically changed by the decision of the MOLA team to use east/'ocentric coordinates for its products. As a result, the best global map of Mars—the MOLA gridded elevation data (Smith et al., 2001), with a final grid spacing of ~1 km and ~100 m horizontal accuracy (Neumann et al., 2001)—is in east/'ocentric coordinates and there are no image maps that register to it. This is the "gold standard" of Mars geodetic control for the foreseeable future, hence the basis for all other maps. Another important result is that the existence of different maps in both IAU coordinate systems is forcing users to be careful about specifying the direction of positive longitude and to convert coordinates between systems as needed. Unfortunately, whereas converting individual latitude values such as target coordinates between systems is straightforward and can be done with a pocket calculator, converting maps between systems requires specialized software. Until quite recently not even the leading planetary cartography software packages (ISIS and VICAR) were able to resample Mars image mosaics to the MOLA grid or vice versa.

The USGS has recently enhanced its ISIS software to work with datasets gridded uniformly in either planetographic or planetocentric latitude, as well as with data labeled in terms of either east or west longitude. In addition, future Survey data products will clearly distinguish the latitude-longitude system used and will provide information to assist use with data in the other system. For example, USGS I-maps will be printed with dual east/'ocentric and west/'ographic coordinate grids, distinguished by different colors. Digital products will include tables with boundary coordinates in both systems.

Furthermore, in 2001 the USGS, with the support and partly at the instigation of the MGCWG, formally proposed to NASA to make future digital and paper Mars maps in the east/planetocentric coordinate system. This proposal was unanimously supported by the NASA Planetary Cartography and Geologic Mapping Working Group (PCGMWG) and in early 2002 was approved at NASA headquarters. Digital products currently in production that will be affected are the final Viking Orbiter global image mosaic (MDIM 2.1), global color and stereo mosaics of MOC wide-angle data (MDIM 3.0) (Kirk et al., 1999; 2000; 2001a) and future products in the series of large-scale topographic maps and geologic map bases.

The USGS proposal was motivated by two practical considerations, rather than by the technical merits of the coordinate systems. First, many potential users lack the time, interest, and knowledge (even if the necessary software is available) to resample data from one coordinate system to another, so it is desirable for the USGS to do this resampling work. Second, MOLA has been joined by the MGS TES team, the entire 2001 Mars Odyssey mission, the Mars Exploration Rovers (MER) mission and MER site selection process, as well as a majority of instrument teams on the Mars Reconnaissance Orbiter in adopting east/'ocentric coordinates. In addition, the HRSC team of the ESA Mars Express mission indicated in 2001 that it would follow the USGS lead in adopting east/'ocentric coordinates. Producing USGS image mosaics in east/'ocentric coordinates will thus reinforce the trend of widespread use of the system and will require much less work than resampling and rearchiving the MOLA and numerous other datasets in the west/'ographic system.

It is important to note that although the projections for digital products will be different in the future (equally sampled in planetocentric rather than planetographic latitude), the projections of printed maps will not be affected by the adoption of east/ 'ocentic coordinates. Whereas digital products are made in "database" projections (sinusoidal and simple cylindrical) defined by simple formulae in which either kind of latitude can be substituted, printed maps use conformal projections (Mercator, polar stereographic, etc.). Definition in terms of conformality (i.e., no distortion of angles) ensures that these projections are the same regardless of whether planetographic or planetocentric latitudes are adopted as the primary system. The only difference in future printed maps will be the location of quadrangle boundaries at "round" numbers of degrees planetocentric rather than planetographic.

3. Conclusion

The adoption of cartographic constants based on high-quality MGS data and therefore unlikely to be superseded for many years, as well as the adoption of east/planetocentric coordinates by the USGS and the majority of mission teams, will ensure that the many Mars datasets produced over the coming decade will share a common cartographic reference frame. As a result of these advances (and of the availability of the MOLA data as a source of highly accurate control), it will be possible to compare disparate datasets with confidence that they are coregistered to order 100-m accuracy.

4. Acknowledgement

We acknowledge the help and assistance of all of the members of the Mars Cartography and Geodesy Working Group, without whom the results presented here could not have been compiled.

5. References

Davies, M., and Arthur, D. W. G., 1973. Martian surface coordinates. *J. Geophys. Res.*, 78, pp. 4355–4394.

Davies, M. and F. Katayama, 1983. The 1982 control network of Mars. *J. Geophys. Res.*, 88, pp. 7503–7504.

Davies, M., et al., 1992. Report of the IAU/IAG/COSPAR Working group on Cartographic Coordinates and Rotational Elements of the Planets and Satellites: 1991. *Celest. Mech. Dyn. Astron*, 53, pp. 377–397.

Davies, et al., 1996. Report of the IAU/IAG/COSPAR Working group on Cartographic Coordinates and Rotational Elements of the Planets and Satellites: 1994. *Celest. Mech. Dyn. Astron*, 63, pp. 127–148.

de Vaucouleurs, G., Davies, M. E., and Sturms, F. M., Jr., 1973. Mariner 9 areographic coordinate system. *J. Geophys. Res*, 78, pp. 4395–4404.

Folkner, W. M., Yoder, C. F., Yuan, D. N., Standish, E. M., and Preston, R. A., 1997. Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars Pathfinder. *Science*, 278, pp. 1749–1752.

IAU, 1971. Proceedings of the Fourteenth General Assembly (Ed. C. de Jagger, A. Jappel), D. Reidel, Dordrecht.

Inge, J. and R. Batson, 1992. "Indexes of Maps of the Planets and Satellites" http://wwwflag.wr.usgs.gov/USGSFlag/Space/mapbook/TOC_nf.html.

Kirk, R. L., Archinal, B. A., Lee, E. M., Davies, M. E., Colvin, T. R., and Duxbury, T. C., 2001a. Global image mosaics of Mars: Assessment of geodetic accuracy. *Lunar Planet. Sci*, XXXII, Abstract #1856, Lunar and Planetary Institute, Houston (CD-ROM).

Kirk, R. L., Becker, K., Cook, D., Hare, T., Howington-Kraus, E., Isbell, C., Lee, E., Rosanova, T., Soderblom, L., Sucharski, T., Thompson, K., Davies, M., Colvin, T., and Parker, T., 1999. Mars DIM: The next generation. *Lunar Planet. Sci.*, XXX, Abstract #1849, Lunar and Planetary Institute, Houston (CD-ROM).

Kirk, R. L., Becker, T. L., Eliason, E. M., Anderson, J., and Soderblom, L. A., 2001b. Geometric calibration of the Mars Orbiter Cameras and coalignment with Mars Orbiter Laser Altimeter. *Lunar Planet. Sci.*, XXXII, Abstract #1863, Lunar and Planetary Institute, Houston (CD-ROM).

Kirk, R. L., Lee, E. M., Sucharski, R. M. , Richie, J. , Grecu, A., and Castro, S. K., 2000. MDIM 2.0: A revised global digital image mosaic of Mars. *Lunar Planet. Sci*, XXXI, Abstract #2011, Lunar and Planetary Institute, Houston (CD-ROM).

Neumann, G. A., Rowlands, D. D., Lemoine, F. G., Smith, D. E., and Zuber, M. T., 2001. The crossover analysis of Mars Orbiter Laser Altimeter data. *J. Geophys. Res.*, 106(E10) pp. 23,753–23,768.

Seidelmann P. K., et al., 2002. Report of the IAU/IAG working group on cartographic coordinates and rotational elements of the planets and satellites: 2000. *Celest. Mech. Dyn. Astron.*, 82, pp. 83–110.

Smith, D. E., et al., 2001. Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *J. Geophys. Res.*, 106(E10) pp. 23,689–23,722.

Wu, S., and F. Schafer, 1984. Mars control network. *Tech. Papers 50th Annual Mtg. ASP*, pp. 456–464.

Zeitler, W. and J. Oberst, 1999. The Mars Pathfinder Landing Site and the Viking Control Point Network. *J. Geophys. Res* 104, pp. 14051–8942.

IC WG II / IV:

SYSTEMS FOR AUTOMATED GEO-SPATIAL DATA PRODUCTION AND UPDATING FROM IMAGERY

AN OBJECT-WISE CLASSIFICATION APPROACH

Volker Walter

Institute for Photogrammetry (ifp), University of Stuttgart, Geschwister-Scholl-Straße 24 D, D-70174 Stuttgart
Germany
Volker.Walter@ifp.uni-stuttgart.de

Commission IV, IC WG II/IV

KEY WORDS: update, classification, change detection, remote sensing

ABSTRACT:

Keeping spatial data up-to-date is a very time and cost intensive task. Every object of the database has to be checked by a human operator (for example by comparing it with an up-to-date orthophoto) to see if there has been a change in the landscape. Therefore the amount of work of updating a spatial database is nearly as high as the primary acquisition. But a large number of GIS applications rely on up-to-date data in order to solve their tasks. The higher the number of objects in the database the more difficult the problem. New satellite systems offer high resolution multispectral data in high quality with high repetition rates. This data can be used as an input for automatic change detection procedures. One approach is for example to classify each pixel of an image to one of several predefined landuse classes. Afterwards, the classification result can be compared automatically with the GIS database in order to detect updates. Whereas the classification is a very well understood and manageable problem, the matching is still a difficult task. Problems arise for example by objects which are not captured according their real shape but according ownership structures or by objects which have a very inhomogeneous appearance. Nevertheless a human operator can deal with that problems and distinguish between correct and incorrect acquisition with high certainty. The reason for this is that human image interpretation is not based on the interpretation on single pixels but on whole object structures and their relations between them. In this paper an approach is introduced that classifies not only single pixels but groups of pixels which represent already existing object geometry's in a GIS database. This object structured classification result is than compared with the existing GIS objects and all objects are marked where the result of the classification is not the same as the object class of the object which is stored in the GIS database. The result is not only a change detection but also a classification into the most likely class.

1. INTRODUCTION

In (Walter, 1999; Walter, 2000) a concept for the automatic update of GIS databases using multispectral remote sensing data is introduced. This approach can be subdivided into two steps. In a first step, remote sensing data are classified with a supervised maximum likelihood classification into different landuse classes. The training areas are derived from the already existing GIS database in order to avoid the time consuming task of manual acquisition. This can be done, if it is assumed that the number of changes in the real world is very small compared with the number of all GIS objects in the database. Because we want to realise update cycles in the range of several months, that assumption can be seen as true.

In a second step the classified remote sensing data have to be compared with the existing GIS objects in order to find those objects where an update occurred or which were collected wrongly. We solved this task by measuring object-wise the percentage, the homogeneity and the form of the pixels which are classified to the same object class as the object is stored in the database (Walter, 2001). All objects are classified into the classes *full verified, partly verified* and *not found.* This classification is done by using thresholds which can be defined interactively by the user.

The problem by using thresholds is, that they are data dependent. For example the percentage of vegetation pixels is varying significantly between data that was captured in summer or in winter. Other factors are light and weather conditions, soil type or daytime. Therefore we cannot use the same thresholds for different data sets. In order to avoid the problem of defining data dependent thresholds, we introduce an object-wise supervised classification approach. The object-wise classification works exactly in the same way as a pixel-wise classification with the difference that we classify not each pixel but we combine all pixels of an object and classify them together. Again, the training areas for the classification of the objects are derived from the existing database in order to avoid a time consuming manual acquisition.

In a 'normal' classification the greyscale values of each pixel in different multispectral channels and possibly some other pre-processed texture channels are used as input. For the classification of groups of pixels we have to define new characteristics which can be very simple (for example the mean grey value of all pixels of a pixel group in a specific channel) but also very complex like measures that describe the form of an object. This approach is very flexible because it can combine very different measures for describing an object. We can even use the result of a pixel oriented classification and count the percentage of pixels in the pixel group which are classified to a specific object class.

2. OBJECT-WISE CLASSIFICATION

2.1 Input Data

The following tests were carried out with ATKIS data sets. ATKIS is the German national topographic and cartographic database and captures the landscape in the scale 1:25,000 (ADV 1988). In (Walter, 1999) it was shown that a spatial resolution of at least 2m is needed to update data in the scale 1:25,000. The remote sensing data was captured with the DPA system which is an optical airborne digital camera (Haala, N., Stallmann, D., Staetter, C. 1996). The original resolution of 0.5m was resampled to a resolution of 2m.

2.2 Classification Classes

Currently 63 different object classes are collected in ATKIS. There are a lot of object classes which have very similar appearance (for example industrial area, residential area, area of mixed used, special used area) and cannot be distinguished only based on remote sensing data without additional information (even for human operators). Therefore we subdivide all object classes into the five landuse classes *water, forest, settlement, greenland* and *streets*. The landuse class *streets* is only used for the pixel-wise classification. Because of the linear shape, streets consist of many mixed pixels in a resolution of 2m and have to be checked with other techniques (see Walter, 1998). The object-wise classification approach can also be used for streets if the input data has a higher resolution, so that streets are represented as areas.

2.3 Input Channels

Like in a pixel-wise classification we can use all spectral bands as input channels. The difference is, that in the pixel-wise classification each pixel is classified separately, whereas in the object-wise classification all pixels which belong to one object are grouped together. In order to analyse the spectral behaviour of objects, we calculate the mean grey value of each channel for all objects.

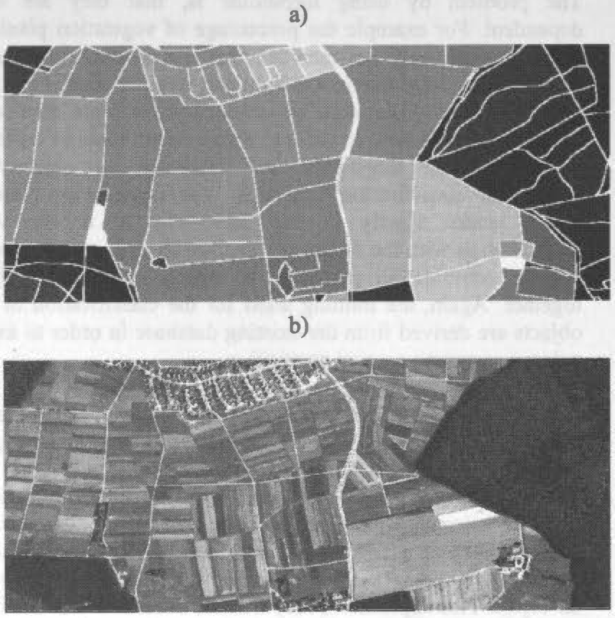


Figure 1: Object-wise vs. pixel-wise classification

Figure 1 shows on an example the original input data (b) and the mean grey value (a) in the green channel of all objects. The result of the pixel grouping is like a smoothing of the data. The spectral behaviour of the objects is equivalent to the typical spectral behaviour of the pixels. For example forest areas are represented in the green channel by dark pixels/objects whereas settlements are represented by bright pixels/objects.

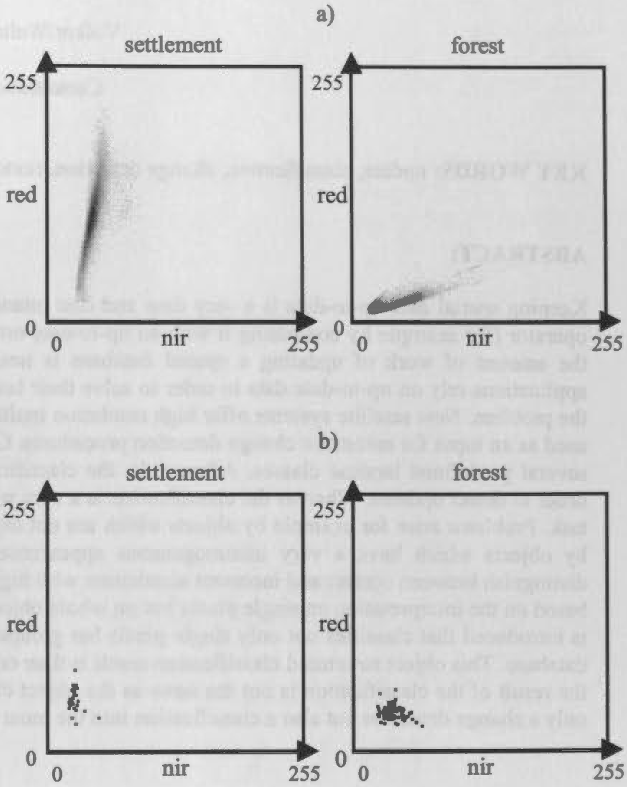


Figure 2: Scatterplot pixels vs. objects

This behaviour can also be seen in figure 2. The scatterplots show the distribution of the grey values of settlement and forest pixels (a) compared with the distribution of the mean grey value of settlement and forest objects (b) in the channels red and nir. It can be seen that the behaviour is similar but the separation of the two classes becomes blurred because of the smoothing. In the object-wise classification all multispectral bands of the DPA camera system (blue, green, red and nir) are used as input channels.

Different land use classes cannot only be distinguished by their spectral behaviour but also because of different textures. Texture operators transform input images in such a way that the texture is coded in the grey values. In our approach we use a texture operator based on co-occurrence matrices which measures the contrast in a 5 * 5 window.

Figure 3 shows the used texture operator on an example. The input image is shown in figure 3 a, the texture image in figure 3 b and the object textures which were calculated again by the mean grey value for all object pixels in figure 3 c. Settlements are represented with dark pixels, greenland with bright pixels and forest with middle grey pixels.

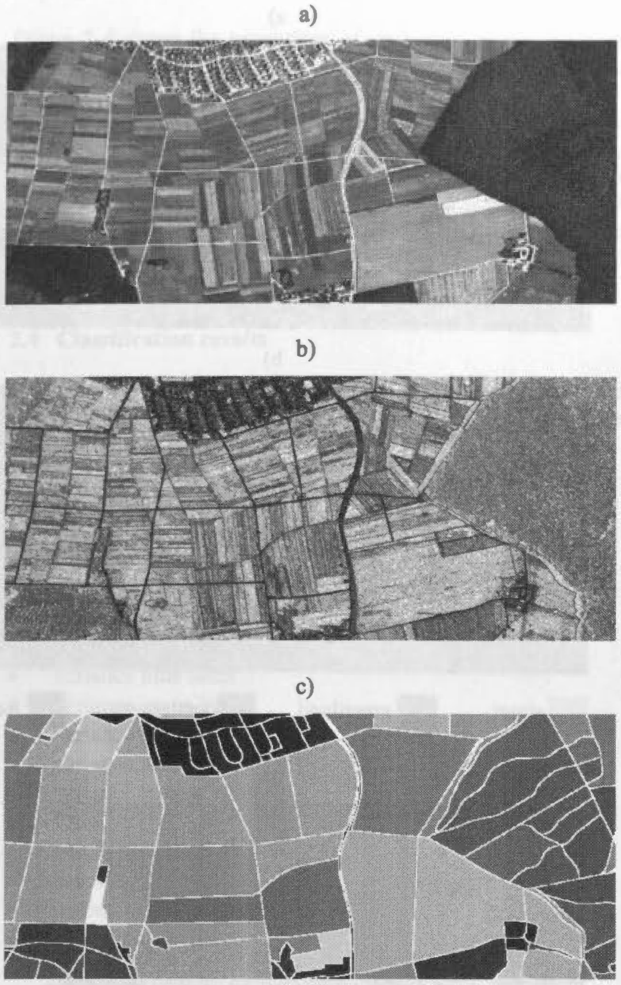


Figure 3: Calculation of the texture for pixels and objects

The variance of the pixels of an object is also a good indicator about the roughness of a texture. Figure 4 shows the calculated variance in the blue band for all objects. Settlement objects have high, greenland objects middle and forest objects low variance values.

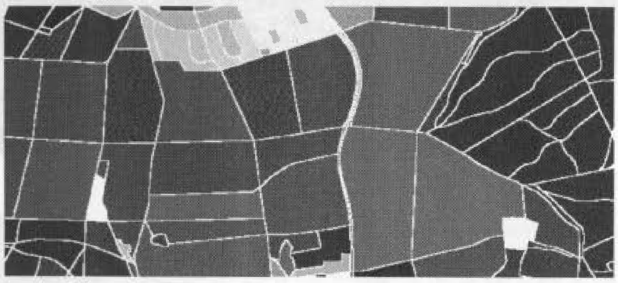


Figure 4: Variance of the objects

Figure 5 shows the behaviour of the variance in the different bands: blue, green, red and nir. The best discrimination between the landuse classes using the variance can be seen in the blue band. In the nir band all landuse classes have a similar distribution which makes a discrimination in this band impossible.

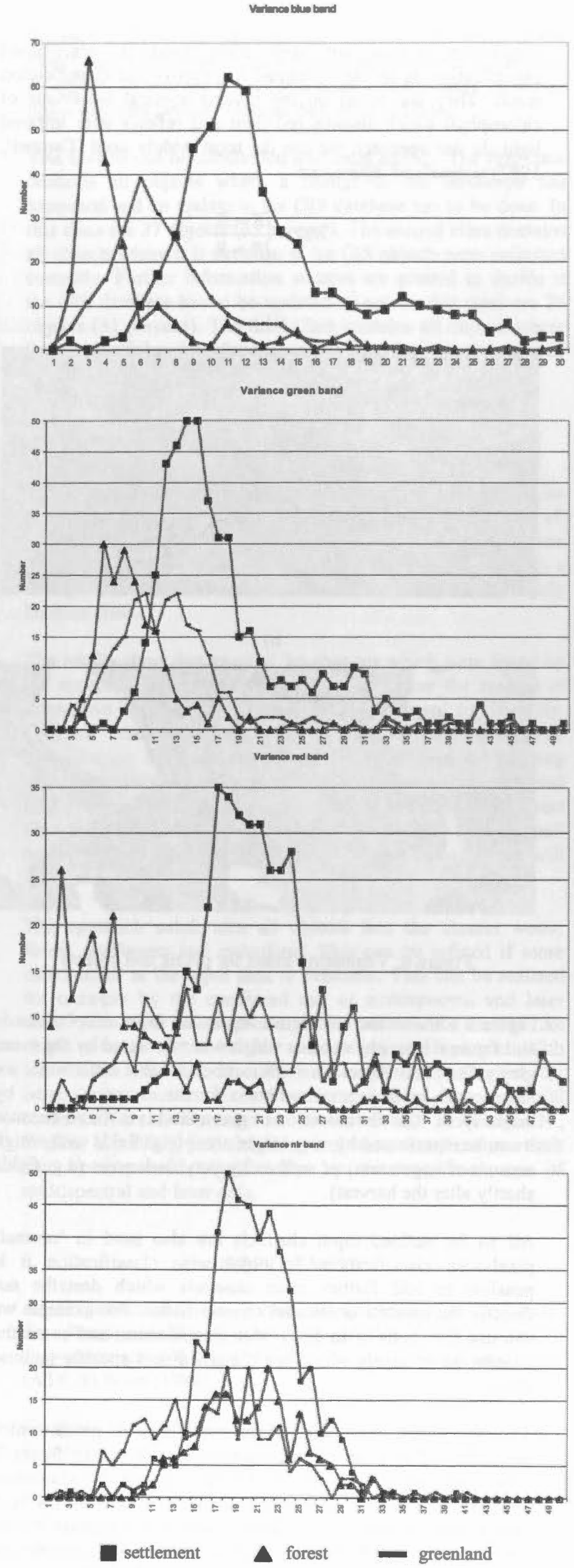


Figure 5: Object variance in different bands

Vegetation indices are very often used in pixel-based classification as an input channel to improve the classification result. They are based on the inverse spectral behaviour of chlorophyll which absorbs red light and reflects near infrared light. In our approach we use the most widely used (Campell, 1987) *normalised difference*:

$$ND = \frac{IR - R}{IR + R}$$

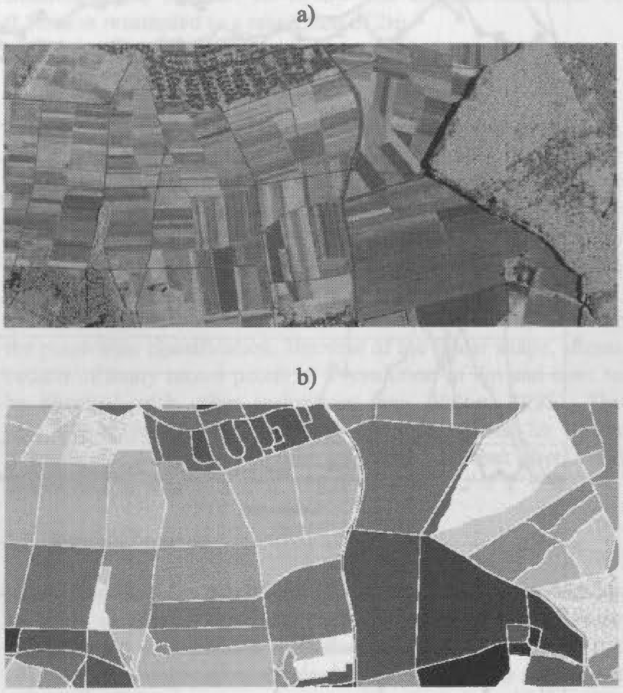


Figure 6: Vegetation index for pixels and objects

Figure 6 a shows the calculated vegetation index on pixel basis and figure 6 b on object basis which was calculated by the mean grey value of all object pixels. It can be seen that settlements are represented by dark areas whereas forests are represented by bright areas. The classification of green land is difficult because it can be represented by very bright areas (e.g. fields with a high amount of vegetation) as well as by very dark areas (e.g. fields shortly after the harvest).

All so far defined input channels are also used in 'normal' pixel-wise classification. In object-wise classification it is possible to add further input channels which describe not directly the spectral or textural characteristics. For example we can use the result of an pixel-wise classification and count the percentage of pixels which are classified to a specific landuse class.

The object-wise evaluation of the percentage of pixels which are classified to a specific landuse class is shown in figure 7. The input image is shown in figure 7 a and the pixel-wise classification result in figure 7 b. Figures 7 c shows for each object the percentage of pixels which are classified to the landuse class forest. White colour represents 100 percent and black colour 0 percent. Forest is a landuse class which can be classified with high accuracy in pixel-based as well as object-based classification.

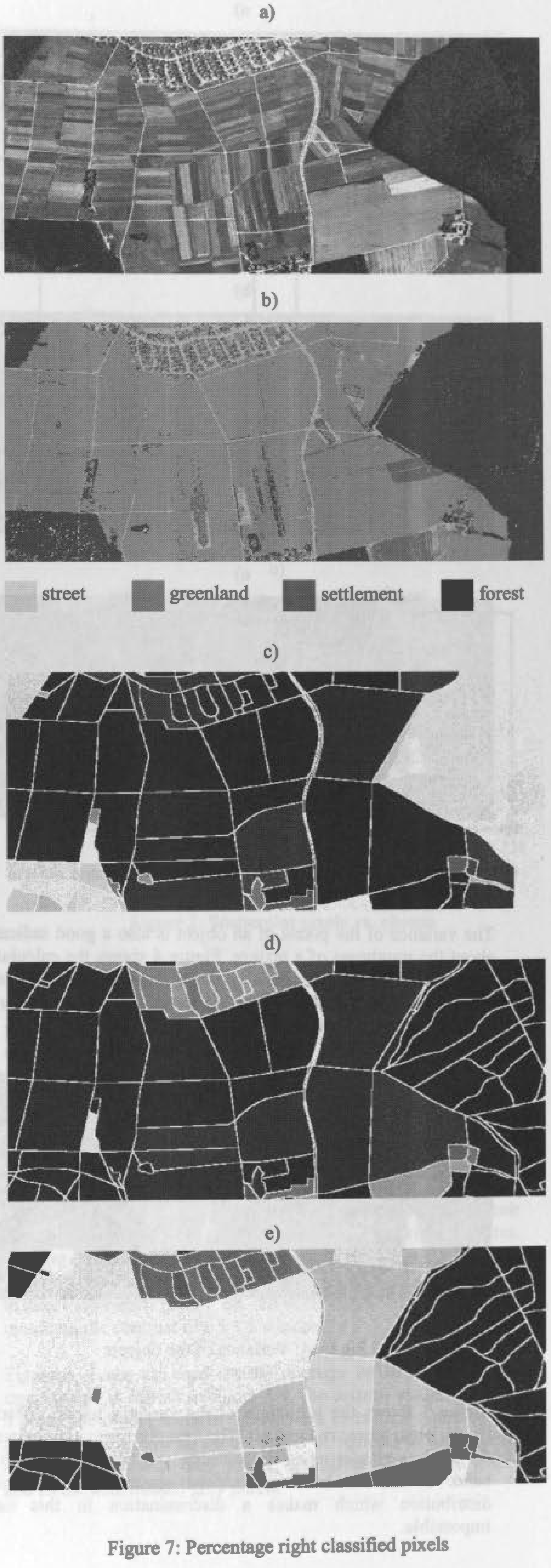


Figure 7: Percentage right classified pixels

Figure 7 d shows the percentage of settlement pixels. Because of the high resolution (2m) of the data, settlements cannot be detected as homogenous areas but they are splitted into different landuse classes depending on what the pixels are actually representing. Therefore settlement objects contain typically only 50 to 70 percent settlement pixels in a resolution of 2m. This can also be seen in figure 7 e which shows the percentage of greenland pixels. Whereas greenland contains up to 100 percent greenland pixels it can be seen that in settlement areas also pixels are classified as greenland.

2.4 Classification results

The approach was tested on two test areas which were acquired at different dates with together 951 objects. The input channels are:

- mean grey value blue band
- mean grey value green band
- mean grey value red band
- mean grey value nir band
- mean grey value vegetation index
- mean grey value texture
- variance blue band
- variance green band
- variance red band
- variance nir band
- variance vegetation index
- variance texture
- percentage forest pixels
- percentage greenland pixels
- percentage settlement pixels
- percentage water pixels

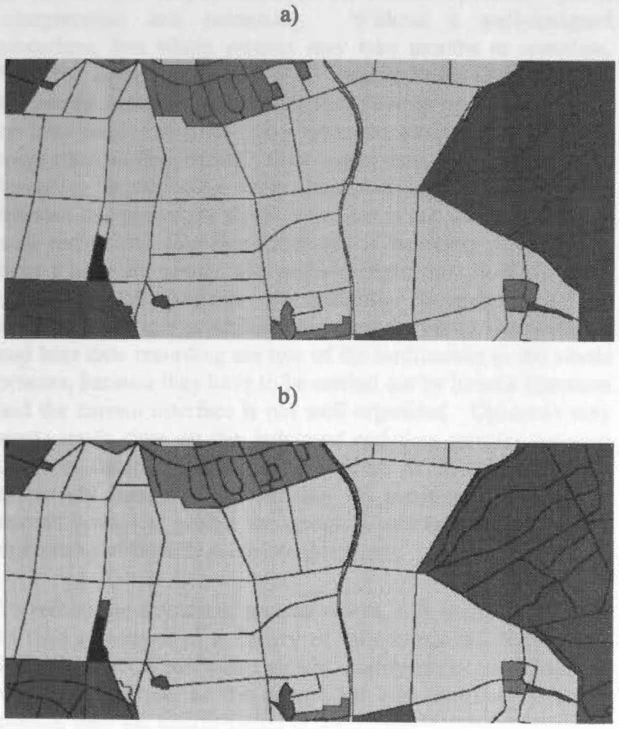


Figure 8: GIS data and result of the classification

Figure 9 a shows the GIS data and figure 9 b the result of the object-wise classification on a part of one test area. Altogether 82 objects (which are 8.6 percent of all objects) were classified into a different landuse class as they were collected.

This objects can be subdivided into three classes. The first class contains all objects where a change in the landscape has happened and an update in the GIS database has to be done. In this class are 37 objects (43 percent). The second class contains all objects where it is not clear if the GIS objects were collected correctly. Further information sources are needed to decide if the GIS database has to be updated or not. In this class are 26 objects (31 percent). The third class contains all objects where the result of the classification is incorrect. In this class are 19 objects (23 percent).

3. DISCUSSION

The object-wise classification needs no tuning parameters like user defined thresholds. It works fully automatic because all information for the classification are "learned" from automatically generated training areas. The result is not only a change detection but also a classification into the most likely landuse class.

The results show that most of the changes which were found by the approach are real changes. That means that the amount of interactive checking of the data can be decreased significantly. On the other hand we have to ask if the object-wise classification finds all changes. Because it uses the existing object geometry, a change in the landscape can only be detected if it affects a large part of an object. If for example a forest object has a size of 5000 m^2 and in that forest object a small settlement area with 200 m^2 is built up, then this approach will fail.

The approach subdivides all objects into the classes *water, forest, settlement* and *greenland*. This can be refined if more information in the input data is available. This can be realised for example by the combined use of multispectral and laser data. With laser data further input channels can be created like slope, average object height, average object slope, etc. With information like that, it could be possible to distinguish for example between residential settlement areas and industrial settlement areas. Furthermore it was shown in (Haala, N., Walter, V., 1999) that also the pixel-wise classification result can be improved significantly by the combined use of multispectral and laser data.

REFERENCES

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) (1998). Amtlich Topographisches-Kartographisches Informationssystem (ATKIS) Bonn (1988).

Campbell, J. B. (1987). *Introduction into remote sensing*. The Guildford Press.

Haala, N., Walter, V. (1999). Classification of urban environments using LIDAR and color aerial imagery in: *Fusion of sensor data, knowledge sources and algorithms for extraction and classification of topographic objects*, 3 - 4 June, Valladoid, Spain.

Hahn, M, Stallmann, D., Staetter, C. (1996). The DPA-Sensor System for Topographic and Thematic Mapping, in: *International Archives of Photogrammetry and Remote Sensing (ISPRS)*, Vol. XXXI, Part B2, 141 - 146, Vienna.

Volker Walter (1998): Automatic classification of remote sensing data for GIS database revision, in: D. Fritsch and M. English (Hrsg.), *"ISPRS Commission IV Symposium on GIS – Between Visions and Applications"*, Stuttgart, Germany", 641 - 648.

Walter, V., Fritsch, D. (2000). Automatic verification of GIS data using high resolution multispectral data in: *International Archives of Photogrammetry and Remote Sensing (ISPRS)*, Vol. XXXII, Part 3/1, 485 – 489.

Walter, V. and Fritsch, D. 2000. Automated revision of GIS databases, in: Ki-Joune Li et. al. (Edts): *Proceedings of the eight ACM symposium on advances in geographic information systems*, Washington, 129 – 134.

Walter, V. (1999). Comparison of the potential of different sensors for an automatic approach for change detection in GIS databases. International Workshop on Integrated Spatial Databases: Digital Images and GIS, Portland, Maine.

Walter, V. 2000. Automatic change detection in GIS databases based on classification of multispectral data, in: *International Archives of Photogrammetry and Remote Sensing (ISPRS)*, Vol. XXXIII, Part B4, Amsterdam 2000, 1138 – 1145.

RICE INVENTORY DATA UPDATE PROCEDURE BASED ON REMOTE SENSED IMAGES: A USER INTERFACE IMPROVEMENT PERSPECTIVE

Hsiung-Peng Liao[1] Jung-Hong Hong[2]

[1] Ph.D. Student, [2] Associate Professor, Department of Surveying Engineering, National Cheng Kung University No.1 University Road, Tainan 701 Taiwan, R.O.C

darkness@sv.ncku.edu.tw junghong@mail.ncku.edu.tw

KEY WORDS: Remote Sensing Images, Rice, Inventory, Spatial and Temporal Database, User Interface

ABSTRACT:

To ensure sufficient and balanced supply of rice production, the Council of Agriculture (COA) must monitor any possible change to the land use and cycle for rice productions. Though COA has been carrying out nation-wide rice inventory procedure in every cropping season for more than 20 years, current COA approach still requires a huge amount of works on photo interpretation, field survey, data edition and statistics report from human operators. By analyzing all procedures required, we proposed an integrated interface concept to effectively reduce required operations and consequently improve the efficiency for creating rice inventory result. The interface allows operators to concentrate on interpretation only and leave all other works to computers. Provided the digital orthophoto can be acquired, the proposed procedure will be much more efficient than the current approach. Though mainly designed for rice inventory data, this mechanism can be applied to all kinds of agricultural crops with a slight modification on the attribute domain values.

1. INTRODUCTION

Rice is the major food to people in Taiwan. To have a better control over the coverage, cycle and production of rice, the Council of Agriculture (COA) has been carrying out nation-wide rice inventory procedures in every cropping season for more than 20 years (Chen, 1989). Current COA rice inventory procedure still largely depends on human operators' photo interpretation and processing. Without a well-designed procedure, this whole process may take months to complete. With the agreement of the WTO (World Trade Organization), the import agricultural products shall have an immediate impact on local market and the rice production control must be much more accurate than before. In the mean time, investment on rice inventory is declining every year and arguments regarding whether this procedure should continue or not were raised. It is well recognized that rice inventory is necessary for COA to have a accurate policy, the problem right now is if it can be completed with minimum cost spending. By analyzing all the steps in the current procedure, it is clear the photo interpretation and later data recording are two of the bottlenecks in the whole process, because they have to be carried out by human operators and the current interface is not well organized. Operators may waste much time on the judgment and data transfer between different medias. The major concern is therefore how to effectively integrate and simplify all necessary steps in the current procedure within the designed interface so that human operators can fluently complete their jobs.

To reduce the amount of manual works, it is certainly possible to take advantage of a variety of remote sensing technology. Even the current procedure should qualify as the use of remote sensing technique, as the source for rice inventory is aerial photos, only the interpretation is completed by human operators. Since vegetation has a strong reaction in the infrared spectrum, the use of infrared image may help operators to reduce interpretation time by quickly excluding non-arable lands. They can be used for detecting the health status of vegetation as well. Unfortunately not only the cost for taking aerial infrared photo twice a year is too expensive for COA to afford, the interpretation still has to be completed by human operators. As for the improvement on the degree of automation, image classification approaches based on the training data and difference among images acquired in multiple spectrums are ultimate choice, the supervised classification technique based on multi-spectral images is applied to accomplish the rice inventory data. The strength of reflection of objects differs in different spectral bands. As mentioned above, we can identify the signal pattern of test objects（like rice）first. Then use the signal pattern to compare to the whole images, such that we can obtain the classification result. With the developing of airborne camera and appropriate classification software, the system can analyze multi-spectral images at the same time. the hyper-spectral can substantially improve the accuracy of classification in recent years.

In the past 20 years, the research on creating rice inventory has evolved from a huge amount of human involvement with field survey, to photo interpretation, supervised classification, and multi-spectral image classification. In order to decrease the required human operation and complete the rice inventory more efficiently, Sabins (1987) proposed an automatic photo interpreted procedures based on rice interpretation knowledge and geo-data. Hwan (1979) first tried to interpret the rice crop status according to the spectral character of rice in Taiwan. Then Liu (1988) analyze the spectral characteristics based on the growth of rice. Some researchers also proposed approach to use remote sensing images to improve the procedure on the estimation of the area of arable lands (Janssen etc., 1990; Dong, 1997; Chen, 1998). In some cases, we must even record the actual cropping status of lands, rather than only the classification result based on pixels (Tseng, 1998). The interpretation results need to be later digitized and stored in digital format (like ArcInfo coverage). Such digital formats can provide a precise description on the scope and area of arable land and easily be embeded in GIS. In order to link with spatial data, we even more record the land ID. The accuracy and the degree of automation of the above approaches may be different, nevertheless, the goals of helping to complete rice inventory in an more efficient way are the same.

Though the above automatic or semi-automatic procedures based on remote sensing images have been improved. Unfortunately, due to the rather small size of arable lands in Taiwan, they either fail to provide 100% correct interpretation result or will require a huge invest to complete, consequently all the proposals are still in experiment stage. Current COA approach is still completed by human operators. We therefore target our research focus on a better control and friendly user interface for so that the efficiency of interpretation can be improved. The mechanism can help users to retrieve and query the cultivation phenomena easily.

In the next section, we'll first discuss the basic property of rice inventory data and the system framework for the proposed integrated system that can handle the creation and management of rice inventory data. Section 3 follows by providing a detailed discussion regarding the user interface design in our system prototype, as well as the analysis on operators' behaviors. Section 4 demonstrates some test results of our designed prototype and section 5 discusses the major contributions of this research.

2. BACKGROUND

2.1 Rice inventory data

A Rice inventory data records if an arable land is used for rice crop in a particular cropping season. There are two basic characteristics for any rice inventory data: coverage and temporal. The characteristic 'coverage' means the boundary of the arable lands used for rice crop must be appropriately determined. The characteristic 'temporal' means the temporality of rice inventory corresponds to a specific crop season. In the rice inventory process, COA is not particularly interested in finding the time of rice seedlings or reaping, what really matter is whether the arable land is cultivated or not. So even if conceptually speaking an 'time interval' may be more appropriate to describe the time-period nature of the crop seasons (Allen, 1983), but it can be abstracted to a time-point level. It is therefore possible to distinguish the temporality of the same arable lands in different crop seasons by inspecting their respective year and cropping season.

Since a rice inventory procedure will investigate the cropping status of all arable lands in a particular season, we can treat its interpretation result as a database. All data in this database corresponds to the same time (i.e., the combination of year and season). For efficient management, COA divides this database into separate map files based on the map coverage of 1/5000 national topographic map framework. We shall define this kind of maps as 'rice inventory file.' It is therefore easy to search and compare results in two different seasons, as they will both refer to the same spatial coverage (denoted by a 8-digit map number). By adding temporal considerations, the filename of every rice inventory files consists of 11 digits (8 for spatial, 2 for year and 1 for season). Every rice inventory file may contain thousands of arable lands represented by 2-D polygons (Frank, 1992). It is therefore easy to use those vector-based commercial GIS softwares to develop our system prototype (e.g., ArcView or MapInfo), as they can accurately describe the polygon boundary, completely record the associated descriptions into their attribute tables, as well as maintain the link relationships between the two.

The basic unit for the management of rice inventory data in our approach is rice inventory file. When opened in our system, it acts just like a 'map layer.' With the grid tessellation characteristic of grid tessellation, spatially speaking, there are only 'disjoint' and 'neighbor' relationships among the coverage of different map layers. Each map layer stores a number of arable lands and their attributes in a selected season. Since all of the arable lands in a map are referring to the same season, though not explicitly stored in their attributes (represented by the filename of the rice inventory file), they have temporal meaning as well. Therefore we shall define a record in a map layer as a 'temporal-arable.' Figure 1 illustrates the concept of the above layer-based model, with each layer respectively describing the crop status in a season. Note the partition and number of polygons are different in the two layers. COA requires the rice inventory data to record only the arable lands that is used for rice crop, so if only parts of an arable land is used, the arable land must be divided into at least two polygons. These polygons will share the same land number, but possibly with different area size. This will surely increase operators' burden in both the interpretation and edition process.

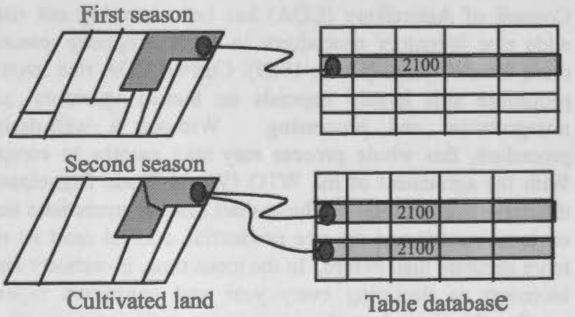


Figure 1 Layer-based rice inventory data

2.2 Related database

Figure 2 shows the framework of the system prototype we proposed in this paper. In spite we only concentrated on the user interface consideration toward the interpretation mechanism in this paper, all temporal-arable is automatically stored in its respective data archives for possible future use. The following discusses the purpose of the three databases:

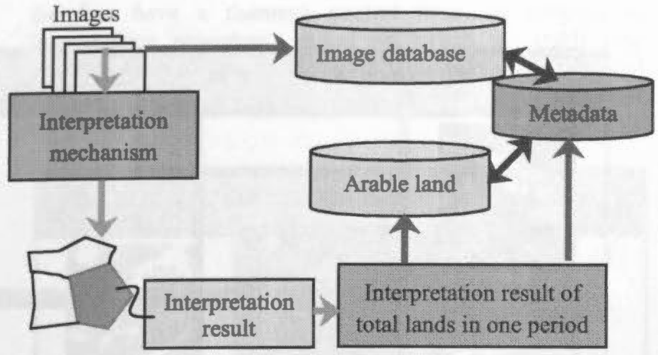


Figure 2 the main architecture of rice investigated and management system

(1) Image database

The major data source for rice inventory data is aerial photos or remote sensing images. No matter the images are digital or not, we need an efficient mechanism for their management. Since the determination of crop status of a selected arable land is based on the interpretation or classification of one or multiple remote sensing images, it is important for systems to be able to provide image information when the temporal-arable is queried. Other than the typical querying and searching functions, the image database must also be able to store a huge volume of image and their associated description information, and in the mean time maintain a link relationship between a temporal-arable and its related remote sensing images.

(2) Historical rice inventory database

A historical rice inventory database must have a temporal management mechanism to manage and analyze rice inventory data in different seasons. Users should be given the flexibility to express their spatial and temporal constraints on attributes or map coverages. System should have a built-in capability for querying any arable lands satisfying a spatial or temporal constraint, as well as providing historical changes for a selected arable land or region. Like image databases, historical rice inventory database constantly grows with the addition of rice inventory data every season, a more powerful querying mechanism that can reduce necessary data volume to its minimum is preferred.

(3) Metadata Database

The metadata database stores the description information about existing images and rice inventory data. It enables the system manager to understand all the data available and quickly retrieve required data.

Under current system framework, operators first choose map coverage to work, the system will automatically search appropriate images for the interpretation or classification process via the data stored in metadata database. Operators interpret the crop status of arable lands and the results of a map coverage are stored as a rice inventory file (with a temporal description). This file is stored in the historical rice inventory database and their associated description information is stored in metadata database as well. The newly input data in metadata database is now available for further search.

3. USER INTERFACE

The current COA approach consists of the following major steps:

(1) Select an map coverage and print a paper map
(2) Find its corresponding aerial photo
(3) Inspect and interpret the crops status of an arable land on the photo
(4) Write down the interpretation result inside the corresponding arable land on the paper map
(5) If necessary, further mark the part of an arable land used for rice crop
(6) Continue step 3 to step 5, until the interpretation of all arable lands on photo or paper map are completed
(7) Change photo or paper map

Based on the content of paper maps, record interpretation results in ArcInfo format manually.

If all the above steps must be completed by human operators, obviously there are some bottlenecks existing in current procedures. The first problem is related to the media used. Operators have to work with a paper map and possibly a number of aerial photos simultaneously. If neither of them are in digital format, operators' vision has to constantly switch back and forth between the two medias throughout the whole process. It gets even worse if a number of images are introduced to provide more reference information. Remember operators have to write down the result on the paper map, so whenever a new arable land is being interpreted, the corresponding relationship will have to be rebuilt visually, i.e., it would be better for operators to remember which arable land is the last one to be interpret. The worst situation is the orientation of photo and maps may not be the same and operators have to align them properly before any examination can begin. The process to transfer recorded results to digital format is the second problem. we have to transform the interpretation result into digital format. Since traditional approach records the result on the paper map, operators have to first select the arable land, then key in its interpretation result one by one. This again requires a huge amount of eye vision change. For those arable lands that are partly cultivated, the transfer has to rely on operators' skill to obtain satisfactory result. The third bottleneck is it is very difficult to control the accomplishment percentage of arable lands on photos. After losing track of what have been done, the work efficiency of operators begins to diminish. The last bottleneck is to maintain the relationships between an arable land and its associated images. Again, if there is only an image involved, to record the filename of an image would probably not bring too much inconvenience to operators. Nevertheless, if a number of images are used as the basis for interpretation, then the working load will be tremendous if every arable land will have to correspond to all of these images. It is not surprising the current approach cannot reach a satisfactory efficiency, as too many unnecessary steps were involved and operators have to do all the works manually.

To improve the current approach, one thing is very clear, the operators' procedure must be simplified with the new techniques of remote sensing and GIS, we therefore proposed the following principles:

- The system must be able to simultaneously display rice inventory map coverages and all the related images in the interface so that users can have a clear and overall visualization impression about the results. This principle is like the automatic alignment of photos and paper maps.

- The system must be able to simultaneously display all related images and the content of all these image windows must be dynamically linked and updated, even if their respective formats are not the same. This is similar to a scenario where three corresponding photos are lying on the table and operators can have four eye visions respectively looking at the same arable land in different map and images.
- The boundary of arable lands can be superimposed onto selected images for operators to realize their corresponding spatial relationships.
- The recording is stored in digital format, such that no more transfer work is necessary.
- Thematic display of the interpretation via an appropriate design of visual variables. Users have a chance to both visualize the interpretation result and control the accomplishment percentage of photo interpretation.
- The use of image compression format so that the retrieval and processing of remote sensing images can be quickly completed.

By introducing metadata database and geometric intersection operations, the first principle can be easily completed. Whenever an arable land or a region is selected, system automatically search images whose coverage 'overlap' the queried area and automatically open image windows to display these images. The presumption is that the time of selected images should refer to the cropping season being processed. For this test, it does not matter what types of images are tested. Figure 3 demonstrates the basic layout of our user interface design. This basic layout consists of 6 windows at the most. The top-left window is the rice inventory map being processed while the bottom-left window is simply an enlargement of the selected area in the top-left window. The other four windows can display four selected images at the most. For regular COA procedure, it is likely only one aerial photo is available. Such a design enables operators to complete their jobs in an integrated user interface without switching eye visions. The bottom-left window provides such tools as zoom in/out and pan, so that operators can select arable lands easily. In our system, whenever an arable land is selected, the corresponding areas in the respective image window are found and the boundary of the selected area is superimposed onto these corresponding areas. To put it simple, system automatically handle the overlap process and all operators need to do is interpreting the pixels within the selected polygon to determine its crop status. Operators' work is therefore simplified to only the selection of arable lands and interpretation. The above discussion shows that loading of rice inventory can be effectively reduced and there is no need for operators to switch their eye visions.

As noted earlier, the left-top window is an "overview window" displaying a rice inventory map. After several experiments , we find the coverage is still too big to work. Every rice inventory map is divided by a 5 x 5 grid, so that the number of processed arable land in each interpretation procedure can be reduced.

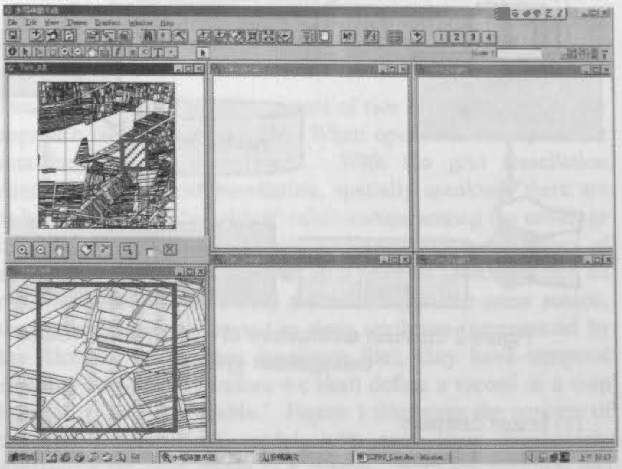


Figure 3 User interface

The selected grid is then loaded into the left-bottom window for further inspection. The rice crop judgment is based on the images displayed, what images to be selected should leave domain experts. For example, the processing of images in different spectrum is similar to the multi-spectrum classification, while images of different growing stages may provide more accurate result by applying rice crop knowledge.

As good as the proposed user interface is, there are two major factors influencing the efficiency. First of all, to obtain good quality result, the images must be digital as well as in high resolution. Since such images easily reaches hundreds of MB, to open a number of images simultaneously is a critical challenge to the hardware. The second factor is that images are preferred to be orthophotos. The original aerial photos do not satisfy this presumption and the generation of orthophotos would require a huge amount of human operation. As a part of research in the integrated system, Professor Tseng, Yi-Hsing (2000) in the Dept. of Surveying Engineering, National Cheng-Kung University is investigating an efficient approach to produce orthophoto based on DTM and aerial triangulation technique. To reduce the amount of image data acquired, we first divide the image based on the above 5 x 5 grid tessellation. Not only the amount of image data reduced, the efficiency of searching images is also improved as these two types of data are now sharing the same subdivision hierarchy. We also introduced ECW image format of ER Mapper (2001) into our system. ECW is a wavelet-based transformation that can effectively reduce the size of a file up to 1/20 of its original size. The selection and reading of images is now much easier and more efficient than before.

After selecting an arable land and visually inspecting the corresponding content in image windows, operator can click the right button of the mouse and evoke a popup menu for the crop status input (Figure 4). The popup menu consists of all possible crop status and all operators need is to select the interpretation result and it will be automatically recorded in the associated attribute table. Note the recording procedure is again automatic and the format would be digital, no media transfer is therefore necessary. Operators can also use multi-select function to select a number of arable lands and assign the interpretation result to all of these selected lands with a single mouse click. As soon as the interpretation result is recorded, the color of the arable land refresh to a refer to its crop status (cultivated or not). Operators therefore have a thorough control over the progress of interpretation procedure. Since we switch the arable land selection from photo approach to rice inventory map, when all the arable lands are colored, the interpretation process is completed.

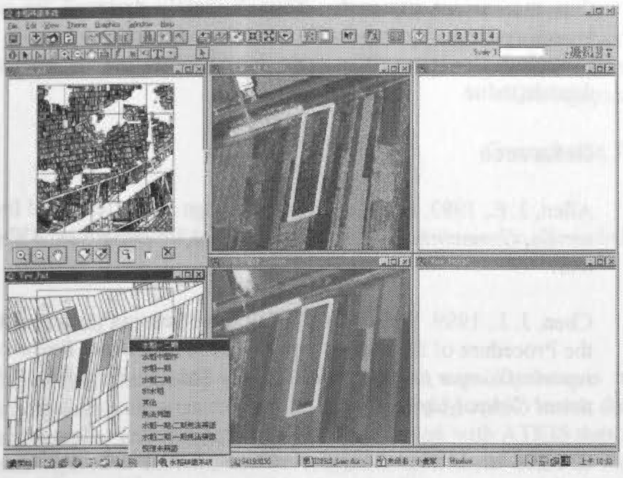


Figure 4 the creation interface of rice interpretation result

The relationship between an arable land and its associated images is stored in the attribute table of rice inventory data (Figure 5). It shows the results of the attribute table. It is possible that arable lands in the same rice inventory map may be interpreted based on different sets of images, so it is better to record associated images independently. This can be completed by a procedure automatically detecting the filenames of the image file in the current image windows. Note this procedure is again automatic.

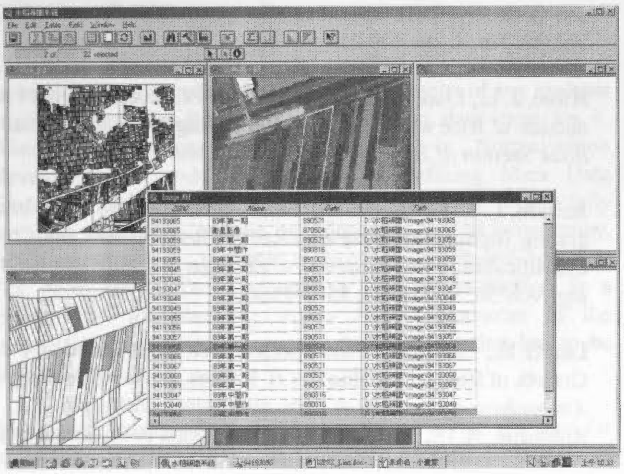


Figure 5 the record of relative image in metadata database

From the above discussion, we demonstrate the new interface and system framework is capable of reducing human operators' loading to as least as possible. Though the acquisition of digital orthophoto is still a possible impediment, the proposed system is nevertheless more powerful and efficient than the current COA approach.

4. TEST

Our mechanism was developed in ESRI ArcView3.2 environment with application programs written in Avenue. All images were processed using software like Imagine and ER Mapper beforehand. In order to display ECW format images in ArcView, we also plug-in a free ECW Viewer in ArcView. The experiment area is located in Shikan, a town in the prefecture of Tainan.

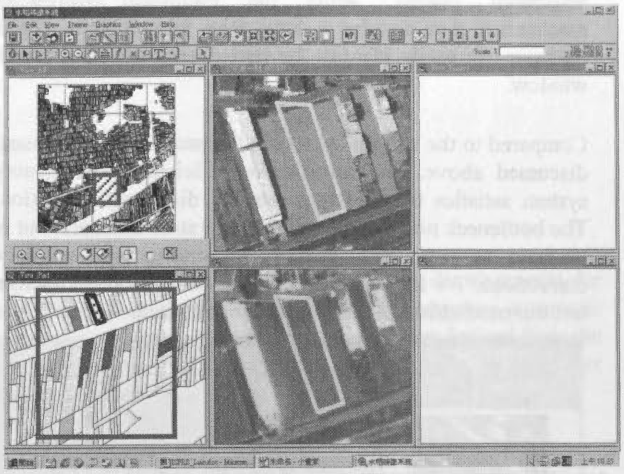


Figure 6 Rice interpretation procedure diagram

Figure 6 illustrates the interpretation process. As mentioned above, the system can automatically locate images to the corresponding arable land. Further more, in the left-bottom window, the system will immediately change the interpreted land's color depending on the interpreted result operators choose (Figure 7).

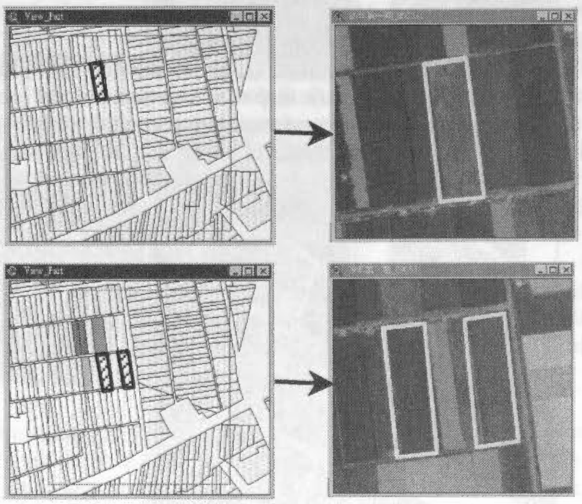


Figure7 Automatically locate images to the corresponding arable land

In order to emphasize the difference between arable lands, the symbol is designed as fill-hue polygon symbol (Figure 7). Human can directly visualize the difference of arable lands (Robinson etc. 1995). Operator can easily control the achieved percentage of schedule, because the arable land of original color means that the arable land has not to be identified. Please exam the interpreted result in figure 8 at the same time. In the left overview and working window, the color is changed based on interpreted result. Actually, when operators interpret the data, the operator via the mechanism already can have a quite control over arable area. Please note the cultivated habit of the middle banding area in figure 8, the habit around this area is almost the same. However the habit and use of the right window is not quite regular. In query mode like figure 9, when operators select one arable land, through the recorded attribute in the table, the system will automatically retrieve the images by querying the metadata database. Should there be a farmer argues about the interpretation result, the system can quickly retrieve its associated images, while the traditional management mechanism may require a long time to complete. Besides, operator can easily identify the period and date label on each window.

Compared to the current approach, the proposed mechanism, as discussed above, can increase the efficiency. The prototype system satisfies the five requirements discussed in section 3. The bottleneck now is probably not operators anymore, but may be how to scan aerial photos, registration and orthophoto corrections. We hope that we can solve this problem soon, and test our mechanism in a large experiment place.

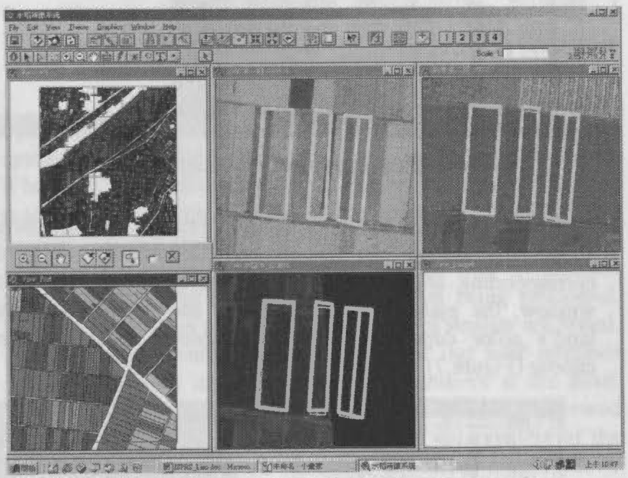


Figure 8 Thematic map of interpretation result

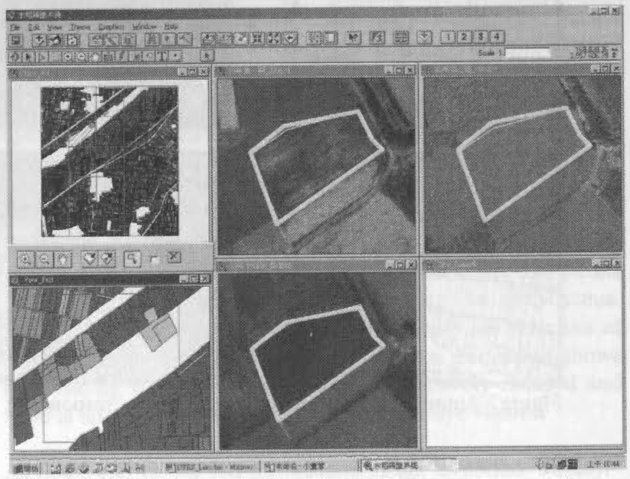


Figure 9 Retrieve relative images

5. CONCLUSION

How to continue its rice inventory procedure at a low cost, while still maintaining its accuracy, is a tough challenge to COA. To effectively reduce possible operators' work, we proposed a user interface of rice inventory procedure, as well as its associated system prototype in this paper. Such an interface provides a simplified and integrated environment to operators. Operators no longer need to switch eye vision between different medias, and the system practically handles most of the work automatically. Provided the digital orthophoto can be acquired, we believe the proposed procedure will be much more efficient than the current approach. Though mainly designed for rice inventory data, this mechanism can be applied to all kinds of agricultural crops with a slight modification on the attribute domain value

Reference

Allen, J. F., 1983. Maintaining Knowledge about Temporal Intervals, *Communications of the ACM*, Vol. 26, No.11, pp.832-843.

Chen, J. J., 1989. With the Aid of Remote Sensing to Improve the Procedure of Estimation the Area of Arable Land, *Research report of Center for Space and Remote Sensing Research at National Central Universit.y*

Chen, Y. H., 1998. Automatic Detecting Rice Fields by Using Multispectral Satellite Images, *Master thesis of Dept. of Surveying Engineering, National Cheng-Kung University*

Dong, M. S., 1997. Integrating Multi-Temporal Remote Sensing Imagery with Cultivating Field Data and Domain Knowledge for a Region-Based Image Interpretation on the Application of Rice-Field Inventory, *Master thesis of Dept. of Surveying Engineering, National Cheng-Kung Universit.y*

ER Mapper, 2001. http://www.ermapper.com

Frank, A. U., 1992. Spatial Concepts, Geometric Data Models and Geometric Data Structures, *Computers and Geosciences*, Vol. 18, No. 4, pp. 409-417.

Hwan, J. L., Liao, W. L., Liao, S. L., 1979. Investigate the Production of Rice with Remote Sensing Images in Chung Hua, *Book Section of Taiwan Forest Bureau*, No. 27

Janssen, L. L. F., Jaarsma, M. N., Linden, E. T. M., 1990. Integrating Topographic Data with Remote Sensing for Land-Cover Classification, *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 11, pp. 1503-1506.

Liu, G. H., 1988. Create the Spectral Characters Based on the Growth of Rice with Rsing SPOT Images, *Master thesis of the Atmosphere department at National Central University*

Robinson, A. H., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J., Guptill, S. C., 1995. *Elements of Cartography*, 6th Edition, John Wiley &Sons, Inc.

Sabins, F. F. Jr., 1987. *Remote Sensing: Principles and Interpretation*, 2nd Edition, W.H. Freeman and Company, New York.

Tseng, Y. H., Hsu, P. H. and Chen, Y. H., 1998. Automatic Detecting Rice Fields by Using Multispectral Satellite Images, Land-parcel Data and Domain Knowledge, *Proceedings of the 19th Asia Conference on Remote Sensing*, pp. R-1-1~R-1-7.

Tseng, Y. H., Chen, S. J., Liao, H. P., Wang, S. D., 2000. Acquisition of Agricultural Geographic Information Using State-of-the art Photogrametric and Remote-Sensing Technologies, Chinese Journal of Agrometeorology, Vol.7, pp.1-10.

QUALITY CONTROL AND UPDATING OF ROAD DATA BY GIS-DRIVEN ROAD EXTRACTION FROM IMAGERY

Felicitas Willrich

Institute for Photogrammetry and GeoInformation (IPI)
University of Hannover, Nienburger Str. 1, 30167 Hannover
willrich@ipi.uni-hannover.de

Commission IV, ICWG II/IV

KEY WORDS: Quality, Updating, GIS, Automation, Knowledge-Base, Imagery, Acquisition, System

ABSTRACT:

Describing the quality of digital geodata in a geodatabase is required for many applications. We present our developments for automated quality control of the German topographic vector data set ATKIS using images. The automation comprises automatic cartographic feature extraction and comparison with ATKIS data, which both are triggered by additional knowledge derived from the existing scene description. To reach an operational solution the system is designed as an automated system which admits user interaction to perform a final check of the fully automatically derived quality description of the data.

1. INTRODUCTION

1.1 Motivation

Digital geodata have increasingly gained in importance for a large number of tasks in planning, documentation and analysis. Today in many application areas digital topographic databases are available und comprise area objects like settlement, agricultural and forestry area as well as linear objects like roads, railroads and water ways.

Describing the quality of digital geodata in a geodatabase is needed for many applications because the results of any analysis highly depend on the quality of the input data used for it. Therefore the Commitee Europeen de la Normalisation developed the model of ISO 19113 defining Meta Data Standards to describe data quality (cf. CEN, 1994). It especially involves the quality criteria completeness, logical consistency, positional accuracy, thematic accuracy and temporal accuracy (cf. Joos 2000). Deriving such a quality description is a prerequisite for revealing errors and inaccuracies in the acquisition, or to discover areas where an updating has to be performed.

In practice among others aerial and satellite imagery is used for acquisition, quality control and updating of geodata. In general these tasks are performed manually by a human operator. As for many application tasks the most expensive part is to supply the basic data an efficient procedure for data acquisition and for quality control is required to ensure that the production process of geodata provides the desired quality with justifiable expenditure. The use of digital imagery has the potential to at least partially automate data acquisition and quality control and thus to speed up and to reduce the costs of data production and maintenance.

For many users geo-databases containing the road network are of especially high interest as one can easily imagine from applications like e.g. in car navigation, route planning and logistics (cf. Fuchs et al., 1998). This is reflected by the high effort in acquisition of a road database which is carried out for car navigation purposes and by the postulated highest up-to-dateness of some few month for the transportation network within the Authoritative Topographic-Cartographic Information System (ATKIS) in Germany.

This paper presents a procedure for automated quality control of roads in the object-based digital landscape model ATKIS DLMBasis being the ATKIS data of the highest resolution approximately equivalent in content to a topographic map 1:25000. The quality description is derived by comparing ATKIS data to imagery which represents reality of the covered scene. The developments result from a pilot R&D project of the University of Hannover initiated and jointly conducted by the Bundesamt für Kartographie und Geodäsie (BKG, Federal Agency for Cartography and Geodesy). The developments are planned to be integrated in the daily production process and thus have to reach an operational solution.

For quality control we compare the vector data to standardised digital black and white orthophotos. Besides the vector data of the DLMBasis these raster data are part of ATKIS and are used because they are area-wide available in Germany.

Automation of the quality control using imagery comprises different subtasks to be solved: a. the extraction of roads and b. the comparison of the extracted roads with the existing data. Our developments especially deal with the exploitation of the scene description given by ATKIS while solving these subtasks. The system is designed to increase the efficiency of the quality control by combining automatic procedures with user interaction in a GIS environment. User interaction still is an indispensable supplement as one cannot expect any fully automatic process to reach the reliability which is required for an operational solution (cf. Lang and Förstner 1996, Gülch 2000). It helps to bridge the remaining deficiencies of the automatic procedures.

1.2 Related Work

Automatic feature extraction from aerial images has been a major activity of international research in photogrammetry and computer vision during the last decades (e.g. Förstner et al. 1999, Baltsavias et al. 2001). Although there is much success in cartographic feature extraction experience has shown that algorithms only give good results if applied to well-defined application areas. All approaches need additional knowledge which can more easily be formulated for restricted situations than for general applications.

GIS data can provide a valuable source of additional knowledge (cf. Vosselman 1996) and can be used to stabilize the image interpretation tasks as algorithms for object extraction can benefit from the information contained in the GIS. This however requires a close and well-defined interaction between image analysis and GIS. Knowledge based systems have proven to be a suitable framework for representing knowledge about the objects and exploiting it during the recognition process.

Quint and Sties (1995) incorporate knowledge derived from maps into a semantic network modeling environment to support image interpretation in urban area. DeGunst and Vosselman (1997) use knowledge-based techniques for the interpretation of road networks. Object-specific knowledge is derived from standards for road construction and the content of digital maps is used for a goal directed segmentation during hypothesis generation. Bordes et al. (1997) perform a database guided road extraction to enhance the vector data accuracy. The underlying road model used for extraction as well as a rough road location is derived from the cartographic database. Wallace et al. (2001) use context information for additional knowledge for linear feature extraction. Within an object-oriented database model different methods for extraction are categorized based on the functionality they provide, which enables the automatic process to choose the best one for a given application. Liedtke et al. (2001) present a system for knowledge-based image interpretation which models structural dependencies by semantic networks. The system is designed to use holistic methods for feature primitive extraction attached to nodes of the network on different semantic levels.

In Walter (1999) automatic feature extraction is applied to quality control and updating of area and line objects in ATKIS by automatically extracting land cover classes by multispectral classification from satellite imagery and comparing it to the corresponding ATKIS objects. Knowledge is derived from the existing GIS for defining training sets for a supervised classification. ATKIS objects that show large differences to the extracted objects with a high probability are presumed to have changed. They are visualized for further interactive analysis by the operator.

The comparison between different data sets is e.g. treated for area objects in Ragia (2000), for linear objects in Wiedemann et al. (1998) and in Straub and Wiedemann (2000).

2. SYSTEM COMPONENTS

The main idea of our developments is to check the quality of the ATKIS DLMBasis by extracting features from black and white orthophotos and comparing the extracted information to the DLM. To increase the efficiency of the quality control, extraction and comparison should be performed fully automatically.

2.1 System Overview

The system is designed to combine fully-automatic analysis with interactive post-processing by a human operator. The fully automatic part attains to reduce the time consuming interaction by a human operator by focussing the interaction on those automatically derived results which are uncertain and unstable.

The system development is embedded in a broader concept of a knowledge-based workstation. Corresponding to the nature of tasks to be solved the system provides functionality from photogrammetry, GIS, and cartography for the acquisition and maintenance of geoinformation. A major goal of this concept is to integrate diverse components performing different subtasks within the framework of a knowledge based system.

Although we presently are focussing on roads the system is designed to handle all object types of ATKIS.

2.2 System Components

The system consists of three major parts: a. the GIS component, b. the process control component and c. the image analysis component (cf. Fig. 1):

The GIS component: The GIS component of the system is based on the GIS ArcInfo 8 and runs with the desktop version under Windows. It is used for automatic pre-processing of the ATKIS data, as an interface to the database and to the image processing system and for interactive post-processing of the automatically derived results. Generally speaking it is the user interface being the workspace for the human operator.

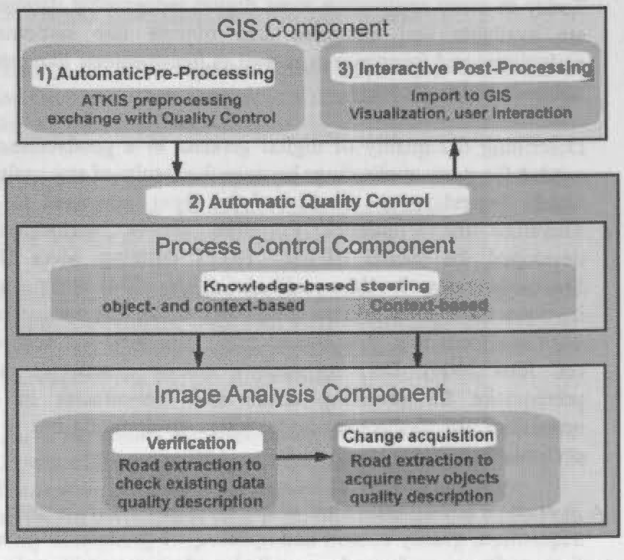


Figure 1: shows the components of the system for quality control

The image analysis component: The image analysis component comprises the automatic cartographic feature extraction modules adapted for the quality check, the comparison of its result to the original vector data and the evaluation of differences between original and extracted roads with the original data leading to quality measures. These tasks are triggered by the GIS data being a valuable source of additional knowledge. The image analysis component is running under Linux and is regarded as being a black box for the operator delivering a preliminary quality description of the data.

The process control component: This part of the system is the link between the GIS component and the image analysis component (cf. Fig. 1). It is responsible for making pre-knowledge from the GIS available to the image analysis component and to transfer it in a suitable way to the object extraction, the comparison and evaluation algorithms. Additionally it is helpful for steering the complete automatic workflow. As the link to the image analysis component is very close the process control component is also running under Linux.

3. PROCEDURE

The procedure for automated quality control is subdivided into three steps: 1) automatic pre-processing, 2) automatic quality control and 3) interactive post-processing (cf. Fig. 1) .

3.1 Automatic Pre-processing

The procedure starts with automatically pre-processing and preparing the GIS data so that it is appropriate for the automatic processes as well as for the interactive analysis by the operator. This pre-processing is performed by the GIS component and compounds e.g. the selection of test area for quality control, the establishment of the link between object geometry and thematic attributes and the supply of an appropriate interface to automatic quality control. Due to practical reasons the working units are image tiles of a certain size, e.g. 2 km × 2 km, or selected image areas which are interactively defined as input to the quality control. For each tile all types of ATKIS objects and their attributes, that are relevant for quality control are requested from the database and are transferred to the image analysis component. At present these ATKIS objects are exported to interchange formats, that can be read by the process control component. In future the transfer will be performed by database queries. The operator simply has to initiate the pre-processing step which then is running in batch.

3.2 Automatic Quality Control

The second step being the core of the system is the automatic quality control which is performed by the process control component and the image analysis component. The quality control starts with deriving knowledge from the GIS for steering the different steps in image analysis. This especially comprises the selection and sequence of algorithms to be executed and the definition of parameter settings for each. The procedure for automatic quality control is described in detail in chapter 4.

At present the knowledge used for process control is coded in rules which are activated by the geodata. The rules presently describe the knowledge which is required for handling roads and the applied object extraction algorithm. In general these rules can be extended to further object classes and additional algorithms. In future the knowledge will be implemented in a more general way in the knowledge based system presented in Liedtke et al (2001) which is to be advanced for quality control purposes within our cooperation with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Hannover University.

The automatic quality control is carried out as a batch process started via the GIS component. The result of the feature extraction steps as well as the quality description of the ATKIS objects are stored in exchange files and are delivered back to the GIS component, where the operator performs the post-editing of those parts of the scene description, which could not be reliably analysed by the automatic process.

3.3 Interactive Post-processing

During the final interactive check of the results the operator focuses on roads for which the automatic quality control indicated an uncertain decision. If necessary the operator has to revise and correct the automatically derived quality description. To ensure that all objects will be handled that are classified as being uncertain, the user can be guided by sequentially presenting him all objects stored in a queue.

For supporting the operator during his interactive intervention the automatically derived quality measures and the underlying extraction results are visualised in an appropriate way together with the orthophotos on the screen. The graphical user interface (GUI) for user intervention provides the following functionalities:

- access to the ATKIS objects and their attributes
- access to the results of the automatic quality control
- tools for the final editing by the human operator
- tools for documentation of the results

It allows the operator to correct or complete the results of the automatic procedure and to classify errors into different error classes like missing objects, inaccurate object geometries and wrong attributes. For further details we refer to Busch and Willrich (2002).

4. AUTOMATIC QUALITY CONTROL

In the following we describe the fully-automatic part in detail. It is to be regarded as being a black box for the human operator delivering a preliminary quality check for focussing the interactive intervention.

The quality control comprises road extraction adapted for the quality check, the comparison of its result to the original vector data and the evaluation of differences between original and extracted roads with the original data leading to quality measures.

4.1 Algorithms for Road Extraction

International research has produced many different algorithms for road extraction each of them being suitable for well-defined extraction tasks. Our concept for checking the quality of roads in general is designed to use different algorithms for road extraction. The selection of the appropriate algorithm in each case is performed by the process control component. We currently apply software developed by C. Wiedemann (cf. Wiedemann et al. 1998, Wiedemann 2001) at the Chair for Photogrammetry and Remote Sensing at the Technical University Munich. We adapted it to our specific tasks especially by exploiting the GIS scene description and embedded it into the knowledge-based framework for deriving a quality description of each handled road object. The underlying road model thereby is partly adapted by parameter settings which are automatically defined or even adapted to the image content by the process control component.

4.2 The use of pre-knowledge for quality control

The knowledge we use for road extraction and for describing the quality of roads is derived from object-specific and context-specific properties whereas the context can be subdivided into global and local context (cf. Baumgartner et al. 1997):

Object-specific properties are e.g. the road geometry as well as road attributes like the road type (highway, single/multi track, road, path), road width, road material (asphalt, concrete). These properties usually are partially represented in the underlying road model of each algorithm and thus characterize the application domain of the algorithm. In many algorithms they can be adapted to the scene by parameter settings. The applied road extraction software e.g. has a parameter describing the width of the road to be extracted. In ATKIS the road width is an attribute of road objects and thus for checking the quality of existing roads this parameter can directly be defined by ATKIS. The road geometry of existing roads can further be used to define regions of interest in which the road extraction is performed.

The global context, e.g. the environment through which a linear feature passes influences the appearance of the road in the images e.g. by probabilities for having disturbances like shadow, fragmentation or low contrast. The expected appearance of the road in general is part of the underlying road model of each algorithm and in many cases can be adapted to the scene by parameter setting as it is done in our case. For extraction and evaluation we use three types of context regions derived from the geo-data: rural, urban and forestry. The appropriate parameters settings for each context class were defined by empirical studies and then are coded in rules.

The local context, e.g. the local neighborhood relations between different objects also influences the appearance of the road in the images by interrelationships like occlusion and shadow, connectivity and parallelism conditions (e.g. buildings cast shadows on roads and buildings in general are connected to roads). It is complex to model the influence of local context and therefore we did not considered it in our approach yet.

4.3 Verification and Change Acquisition

The main principle of our procedure for quality control is to exploit the initial scene description in the geo-database to guide and constrain the underlying road extraction and the comparison with the given vector data in the following way:

- by definition of regions of interest
- by selection of the appropriate algorithm
- by parameter control of the road extraction
- by parameter control for evaluating the results

Depending on the different quality aspects to be checked like geometric or thematic accuracy of existing roads on the one hand and completeness of the road data on the other hand we perform the quality control in two stages: 1. the verification of existing roads and 2. the acquisition of changes in the road data. This partitioning mainly is motivated by the different amount of knowledge which can be exploited. In both cases we first perform road extraction and then compare its result to the roads given in ATKIS.

4.3.1 Verification

The verification checks those objects which are described in the database. It is able to check the positional and the thematic accuracy as well as the commission error (road does not exist in reality). Besides very general knowledge about the roads, in this case specific knowledge can be derived for each individual instances of an ATKIS object and can be used for parameter steering during extraction and quality evaluation in the sense of the hypothesis-verify paradigm.

To introduce the specific knowledge given by object instances the verification is performed object by object by comparing the existing road data with roads extracted from the images. The geometric and thematic description of each existing object is transferred to constraints in the road extraction of each object. The geometric description e.g. is used for defining regions of interest, the thematic attributes are used to select the appropriate algorithm and its parameters.

4.3.2 Acquisition of Changes

The acquisition of changes aims at finding objects which are not contained in the database and thus serves for describing the completeness of the data especially by registering roads which are missing in the database. This task being equivalent to object extraction from the scratch is more difficult than the verification task because we only can introduce general knowledge and have to renounce on specific knowledge derived from existing roads. Solely the global context can be derived from ATKIS and can be used for parameter setting of the road extraction algorithm as it also is done during the verification step. The acquisition of changes is executed subsequently to the verification to at least introduce verified ATKIS objects as reliable road parts and use it as seeds during the network generation (cf. Wiedemann 2001).

4.4 Quality evaluation

The first step in both tasks, verification and acquisition of changes, is the extraction of roads. The extracted roads then are to be compared to the given ATKIS roads to derive a quality description of each single road object as well as of the complete road network. Our focus is on describing the quality in terms of correctness, completeness and accuracy (cf. Wiedemann et al. 1998) of the geometric and thematic description of the objects. We do not cope with the consistency of data and model which can be checked by logical tests as it is already performed in many cases in practice.

Respectively to the road verification and change acquisition tasks the quality evaluation also has to distinguish 1) roads in the database and 2) roads that are newly acquired.

4.4.1 Quality description in road verification

To derive a quality description in road verification the differences between the extracted roads and the GIS data are analysed and evaluated. The quality description is simplified to a so-called traffic-light solution (cf. Fig. 2) indicating three types of quality attributes: *accepted, rejected, undecided.*

We first check if an extracted road matches the corresponding ATKIS object. If the road extraction has been successful the road is denoted *accepted*. If no road matches the ATKIS object the local situation is to be further analysed in a second step by a feedback loop to determine if the error was caused by inaccurate attributes, inaccurate geometry, by application of a unsuitable road model, by difficult contrast conditions or by the road having ceased to exist. We first go back to the underlying line extraction being the basis for extracting roads and test how good these lines fit the ATKIS road. We especially check the coverage of lines which fit the direction and position of the ATKIS road. If the coverage[1] c is larger than a threshold (in our cases set to 60%) the road is denoted as *undecided.* If c is less than this threshold we propose to go to the pixel level to test the local contrast by analysing grey value profiles perpendicular to the road axes. This could give hints to wrong geometry, or to contrast situations which actually can not be handled by the currently employed road model like an unsymmetric line model. Depending on the situations the ATKIS road to be inspected is denoted as *rejected* or in case of the unsuited road model denoted as *undecided.* The quality classes *undecided* is to be

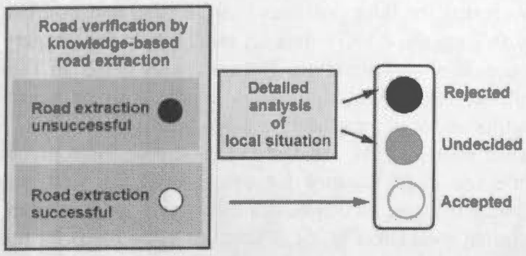

Road verification by knowledge-based road extraction
Road extraction unsuccessful
Road extraction successful
Detailed analysis of local situation
Rejected
Undecided
Accepted

Figure 2: shows the classification of ATKIS roads into the three quality classes *accepted* (white), *rejected* (black) and *undecided* (grey).

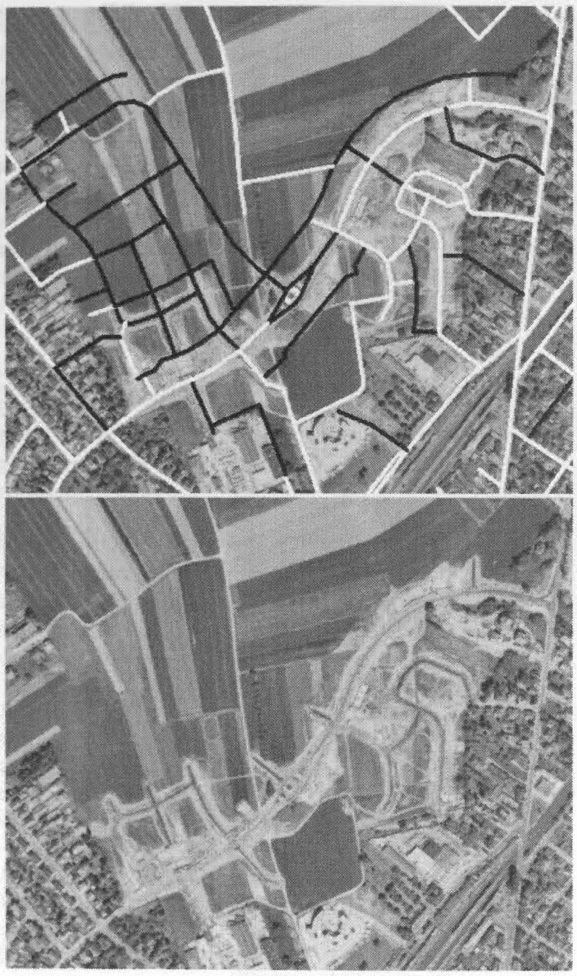


Figure 3: shows an example of the verification result for an updating situation where the geo-data differ from the image content due to road construction. Please note that in this example the imagery is older than the ATKIS data and thus the imagery does not reflect all new roads which already are contained in ATKIS (accepted roads in white, rejected roads in black) (ATKIS®, DLMBasis und Orthophotos; Copyright © Hessisches Landesvermessungsamt 2002).

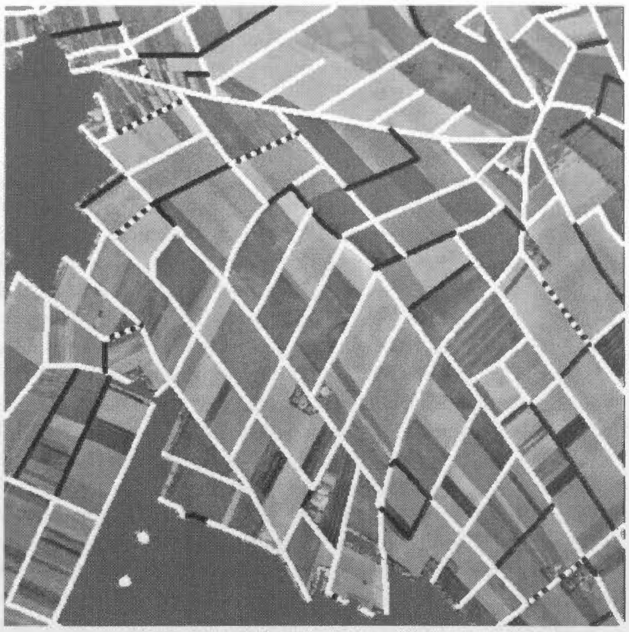


Figure 4: shows an example of the verification result in rural context represented by a classification of the ATKIS roads into the three quality classes *accepted, rejected* and *undecided* as it is transferred to the interactive post-editing. White lines denote accepted, black lines rejected and dashed lines undecided roads. (ATKIS®, DLMBasis und Orthophotos; Copyright © Hessisches Landesvermessungsamt 2002)

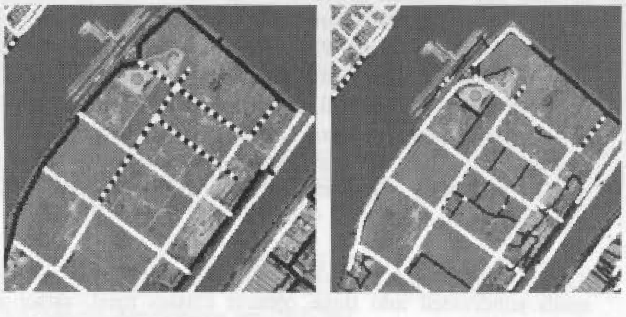


Figure 5: shows the result of the automatic verification (left) and the change acquisition result (thin black lines) overlaid over the verification result (right).

---

[1] $c = \frac{l_{ml}}{l_a}$ with l_{ml} = length of matched lines and l_a = length of ATKIS road

presented to the operator. Fig. 3 and 4 show examples of the automatically derived quality description in road verification.

4.4.2 Quality description in acquisition of changes

In acquisition of changes we are interested in new objects which are not contained in the original dataset. They are derived by subjecting the roads extracted in the complete image to the evaluation scheme proposed by (Wiedemann et al. 1998) by matching the extracted roads to the existing roads in the database and vice versa. Unmatched extractions are denoted as changes and are delivered to the operator. Fig. 5 shows an example of automatically derived changes. As there still are many extracted road elements being wrong interpretations a further classification of these changes is required, e.g. by deriving their internal accuracy using the underlying line extraction to obtain a reliability measure of the changes. The reliability of the road extraction result especially depends on the extraction context and on the underlying low-level extraction used for road network generation. It could be used to derive a traffic-light-solution describing the quality of the data by a qualitative description as it is done during the verification of the existing data. Thus the user interaction can further be reduced to very probable changes.

5. RESULTS AND DISCUSSION

For evaluating the performance of our procedure we tested the verification step with 30 orthophotos covering an area of *10 km x 12 km* near Frankfurt a.M. The black and white images meet the ATKIS orthophoto standards with *0.42 m* resolution on ground. The complete scene is subdivided into three classes of context areas. The 10368 roads in the covered scene roughly split into 43% in rural, 42% in urban and 13% in forestry context. The used road extraction software (cf. chapter 4.1) is designed to provide optimal results for resolutions between *1–2 m* on ground in rural context. Therefore we resampled the given imagery to *1.70 m.* The road data we checked are part of the ATKIS DLMBasis. Compared to the imagery these vector data are nearly up-to-date.

Table 1 shows the achieved verification results of the existing roads subdivided into three context classes rural, urban and forestry. The classes are automatically derived from the given ATKIS objects of type region by grouping those ATKIS regions showing similar appearance in the images. For each class we used empirically determined optimal parameter settings. The table shows the number of roads being either *accepted, rejected* or *undecided* in % of all roads covered by the context area. In rural context the percentage of accepted roads is about 79 %, the percentage of rejected roads were 17 % whereas in 4 % of the roads an uncertain decision of the automatic system was derived. The optimum of accepted roads is estimated to be approximately 95 % as the scene is nearly up-to-date and only some roads especially being tracks are even uncertainly arbitrable by a human specialist.

The reasons for rejecting roads in rural context is mostly caused by applying an unsuitable road model for extraction in local contrast conditions changing from dark to light neighbourhood or vice versa or by roads showing very low contrast to its surroundings. Therefore we propose to further inspect rejected roads by analysis on the pixel level e.g. by analysing grey values of cross sections along the road axis to obtain hints for the reason of rejecting the road which e.g. can be used to select a more suitable road model. The results in urban context only reach a percentage of accepted roads of 57% while 34% were rejected and 8% were undecided roads. The reason for the worse results is that parameter settings had to be set differently to prevent a large number of false positives.

We tested for false positives both in rural and in urban context by shifting the ATKIS data set relatively to the imagery *30 m* in x and *30 m* in y direction. The results are given in Table 2. The optimal result should be 0 % of roads to be accepted. The results show in rural area a false positive rate of 33% and in urban area of 21%. In fact in rural area these numbers are a little too large because for some roads the shift results in a change in road direction still effecting a large overlap with the original road (cf. Fig. 6). Therefore some roads by mistake are assigned to the class *accepted*. Further false positives are caused by roads running through textured context area like e.g. specialised cultivation in rural area or low contrast area. In future we therefore propose to handle specialised cultivation area as an additional context area with different parameter settings.

Context region	*% accepted*	*% rejected*	*% undecided*
rural	79	17	4
urban	57	34	8
forestry	60	34	6

Table 1: shows the achieved verification result subdivided into three context classes by the number of roads in % of all roads covered by the context area in the image being either *accepted, rejected* or *undecided.*

Context region	*% accepted*	*% rejected*	*% undecided*
rural	33	64	3
urban	21	75	4
forestry	18	81	1

Table 2: shows the achieved results with ATKIS shifted relatively to the imagery.

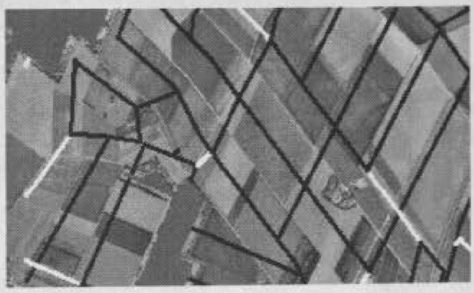


Figure 6: shows an example of the verification where the given ATKIS roads are shifted (accepted roads in white, rejected roads in black)

The results in forestry area are not very satisfying but are plausible as many of the roads even could not be verified by a human operator based on the imagery. Only very wide roads like e.g. highways could be automatically verified in this context area. For verification in urban as well as in forestry area different sensors like e.g. laser scanning data have to be used additionally to the imagery. Thus the 3d information can supplement the 2d information to preclude buildings and trees.

6. CONCLUSIONS AND OUTLOOK

We presented our procedure for automated quality control of the Authoritative Topographic-Cartographic Information System ATKIS in Germany for checking the consistency of the data with reality represented by imagery. To solve this task in an economical way the procedure comprises a fully automatic part which is followed by interactive correction by a human operator. The fully automatic part reduces the subsequent user interaction by focussing on those objects where the results of the automatic process are uncertain and therefore a user interaction is required to reach a reliable results. First results show the feasibility the concept. The verification has to be refined by integrating alternative algorithms for road extraction, by further feed-back during the quality evaluation and by reducing the rate of false positive alarms. The results in change detection have to be refined by evaluating the initially deduced changes and have to be tested on a large test data set to rate its performance.

Acknowledgements: We acknowledge the BKG and the University Hannover for funding this work and the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, for the valuable cooperation at Hannover University.

Furthermore we thank the Hessische Landesvermessungsamt for providing us with the test data. Finally we would like to thank the Chair for Photogrammetry and Remote Sensing at TU Munich, esp. C. Wiedemann for his support in using his road extraction software.

REFERENCES

Baltsavias, E, Grün, A., van Gool, L., 2001, Automatic Extraction of Man-Made Objects from Aerial and Space Images(III), A.A. Balkema Publishers, Lisse / Abingdon / Exton(PA) / Tokio.

Baumgartner A., Eckstein W., Meyer, H. Heipke C., Ebner H., 1997: Context-supported road extraction. In Grün, A., Baltsavias, E.P., Henricsson, O., (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Publishers Basel, 1997, pp. 299-308.

Bordes, G., Giraudon, G., Jamet, O. , 1997 : Road Modeling Based on a Cartographic Database for Aerial Image Interpretation, In. Förstner, W., Plümer, L. Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, Birkhäuser Publishers Basel, 1997, pp. 123-139.

Busch, A., Willrich, F., 2002: Quality Management of ATKIS Data. Proceeding CD of the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21-22 March 2002, Istanbul, Turkey, pp. 30-41.

CEN, 1994: Data description: *Quality. Draft of Definitions Document* No. 15, rev. 4(august); CEN TC 287 WG2.

DeGunst, M. , Vosselman, G., 1997 : A Semantic Road Model for Aerial Image Interpretation, In. Förstner, W., Plümer, L. Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, Birkhäuser Publishers Basel, 1997, pp. 107-122.

Förstner, W., Liedtke, C.-E. and Bückner, J. eds., 1999, Proceedings of the Workshop 'Semantic Modeling for the Acquisition of Topographic Images and Maps SMATI 99,

Gülch, E., 2000: Digital Systems for automatd cartographic feature extraction, In. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, B2, pp. 241-256.

Joos, G., 2000, Zur Qualität von objektstrukturierten Geodaten. In Schriftenreihe des Studienganges Geodäsie und Geoinformation der Universität der Bundeswehr München, Nr. 66.

Lang, F., Förstner, W. 1996: 3D-city modelling with a one-eye stereo system, In. International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, B4, pp. 261-266.

Liedtke C.-E., Bückner J., Pahl M., Stahlhut O., 2001: Knowledge based system for the interpretation of complex scenes, In E. Baltsavias, A. Gruen, L.,van Gool (eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), A.A.Balkema Publishers, Lisse / Abingdon / Exton(PA) / Tokio, pp. 3-12.

Quint, F., Sties, M., 1995: Map-based semantic modeling for the extraction of objects from aerial images, In: Grün, A., Kübler, O., Agouris, P. (eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag, Basel, pp. 307-316.

Ragia, L. 2000: A Quality Model for Spatial, In International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, B4, pp. 855-862.

Straub, B.-M., Wiedemann, C., 2000: Towards the update of geodata by automatic object extraction, In Serpico S.B. (ed.), SPIE Vol. 4170, Image and Signal Processing for Remote Sensing VI. 304-315.

Vosselman, G., 1996: Uncertainty in GIS supported road extraction, In. International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, B3, pp. 909-917.

Walter, V., 1999, "Automated GIS data collection and update". In D. Fritsch and R.H. Spiller, eds., Photogrammetric Week '99, Wichmann, Heidelberg, pp. 267-280.

Wallace, S.J., Hatcher, M.J., Priestnall, G., Morton, R.D., 2001: Research into a framework for automatic linear feature identification and extraction, In E. Baltsavias, A. Gruen, L.,van Gool (eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), A.A.Balkema Publishers, Lisse/Abingdon/Exton(PA)/Tokio, pp. 381-390, 2001

Wiedemann, C., 2001: Extraktion von Straßennetzen aus optischen Satellitenbildern, Dissertation Lehrstuhl für Photogrammetrie und Fernerkundung, TU München.

Wiedemann, C., Heipke, C., Mayer, H., Jamet, O., 1998: Empirical evaluation of automatically extracted road axes, In: Empirical Evaluation Methods in Computer Vision, Ed. Bowyer, K.J., Phillips, P.J. IEEE Computer Society Press, pp. 172-187.

EXTRACTION AND UPDATE OF STREET NETWORKS IN URBAN AREAS FROM HIGH RESOLUTION SATELLITE IMAGES

Renaud Péteri, Thierry Ranchin

Groupe Télédétection & Modélisation, Ecole des Mines de Paris
B.P. 207 • 06904 Sophia Antipolis cedex, France
renaud.peteri@cenerg.cma.fr, thierry.ranchin@cenerg.cma.fr

Commission IV, IC WG II/IV

KEY WORDS: Automatic road extraction, street network update, high resolution image, urban area, assessment criteria

ABSTRACT:

Among possibilities of remote sensing for urban uses, extraction of street networks is an important topic. A good knowledge of the street network is necessary for urban planning, map updating or estimation of atmospheric pollution. Moreover, an assisted street extraction from remotely sensed data can highly help cities which do not have a cartography of their road network.
However, if road extraction from rural or semi-rural area images has led to encouraging results, the urban context, because of its complexity and the numerous artefacts encountered, is still a new and challenging topic.
The context of our current work is urban area and high spatial resolution imagery. In order to help image interpreters, we are concerned about extracting and updating street networks. In this paper, we present a complete methodology adapted to those aims. It includes the definition of a reference object for quantitative evaluation and also the processing algorithm. The extraction process is then applied on a high resolution IKONOS image. Results and future prospects are finally discussed

RÉSUMÉ:

Parmi les apports possibles de la télédétection en milieu urbain, l'extraction automatisée de réseaux de rues est un sujet important. En effet, une bonne connaissance du réseau de rues est nécessaire pour l'aménagement de la ville, la mise à jour de cartes ou l'estimation de la pollution atmosphérique. De plus, une assistance à l'extraction manuelle de rues peut grandement aider les villes ne disposant pas d'une cartographie de leur réseau routier. Cependant, si l'extraction de routes en milieu rural ou peri-urbain a mené à des résultats encourageants, le contexte urbain, du fait de sa complexité et du grand nombre d'artefacts qu'il engendre, est encore un nouveau domaine à explorer. Notre travail se situe dans le domaine de l'imagerie spatiale à haute résolution de scènes urbaines. Dans le but d'aider les photo-interprètes, nous souhaitons extraire et mettre à jour les réseaux de rues.
Dans cet article, nous présentons une méthodologie complète adaptée à ces objectifs. Elle comprend l'établissement d'une référence pour l'évaluation quantitative des résultats, ainsi que l'algorithme de traitement. Ce procédé d'extraction est ensuite appliqué à une image haute résolution du satellite IKONOS. Enfin, les résultats ainsi que les améliorations futures sont discutés.

1. INTRODUCTION

1.1 Context

Among possibilities of remote sensing for urban uses, extraction of street networks is an important topic. A good knowledge of the street network is necessary for urban planning, map updating or estimation of atmospheric pollution. Moreover, an assisted street extraction from remotely sensed data can highly help cities which do not have a cartography of their road network.

Road extraction from remotely sensed images has been the purpose of many works in the image processing field, and because of its complexity, is still a challenging topic. These methods are based on generic tools of image processing, such as mathematical morphology (Destival, 1987), linear filtering (Wang and Howarth, 1987; Wang *et al.*, 1996), or on more specific tools using Markov fields (Merlet and Zerubia, 1996; Stoica *et al.*, 2000), neural networks (Bhattacharya and Parui, 1997), cooperative algorithms (McKeown and Denlinger, 1988), dynamic programming (Jedynak, 1995; Gruen and Li, 1995), or multiresolution (Baumgartner *et al.*, 1996; Couloigner and Ranchin, 2000; Laptev *et al.*, 2000). Promising studies try to take the context of the road into account in order to focus the extraction on the most promising regions (Baumgartner *et al.*, 1999 ; Ruskoné, 1996). Efficiency of all of these methods depends on the use they are targeting. For instance, for operational constraint, semi-automatic approaches are prefered.

In order to overcome all problems raised by an automatic or semi-automatic road extraction, most of those works focused on rural or semi-rural areas, where artefacts are less numerous than in cities. However, the recent possibility to have high resolution satellite images (1 meter or less) has increased the interest in satellite images for urban use.
Figure 1 is a good example of artefacts generated both by the urban context and the high spatial resolution. Vehicles (circled), the bridge and its shadow disturb the road

extraction process. Other artefacts encountered can be trees, tarred areas (parking, airport) or buildings with a radiometry similar to roads and with an important contrast to their environments.

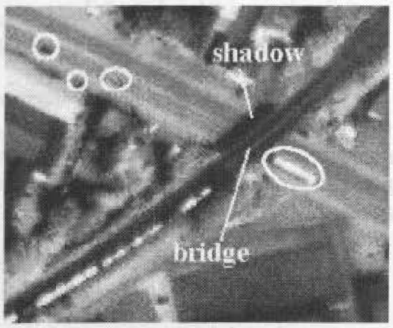


Figure 1. Artefacts encountered in urban context and high resolution images. Copyright 2000 Space Imaging Europe. *Courtesy of GIM - Geographic Information Management*

Baumgartner *and al.* (1996), and Couloigner (1998) have shown the interest of using multiresolution algorithms, especially on high resolution images: small details, such as shadows or vehicles do not disrupt the numerical processing. In high resolution satellite images, a road is represented as a surface element which was not the case with lower resolution where streets were seen as linear elements. This increase of the spatial resolution leads to a more complex scene but also enables a more accurate extraction.

1.2 Objectives

The context of our current work is urban area and high spatial resolution imagery. In order to help image interpreters, we are concerned about street extraction.
In the case where a road database already exists, we also aims at updating street networks. Indeed, with the rapidly changing environment and increasing amount of new information, remotely sensed images provide the perfect media for keeping a geographic information system (GIS) up-to-date. Research on this topic have been carried out in (Agouris *et al.* 2001a), incorporating accuracy information to identify local or global change to the prior information given by the initial GIS.

In the next sections, the proposed methodology is presented and applied on an IKONOS image of the town of Hasselt, Belgium.

2. METHODOLOGY

2.1 Description

In order to extract urban street networks from high resolution satellite images, we propose a methodology, aiming at helping operational users.
For operational purposes, the main constraint is the reliability of the result: false detection or over detection are banned. To fit to this goal and increase the robustness, a semi-automatic approach has been chosen (see Figure 2): the inputs of the algorithm are linear elements coming from a database or manually given by the user. In both cases, it provides an initialization of the graph representing the road network, which vertices are street intersections and ridges are street axis. Models of streets (using roads properties defined by Couloigner 1998) and properties of street network (such as connexity) are introduced to perform the extraction.

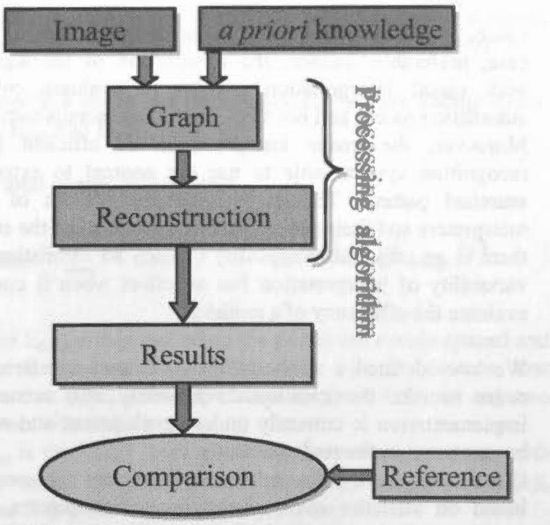


Figure 2. The methodology including topology management and street reconstruction

As for (Airault and Jamet, 1995), our approach is decomposed in two sequential parts.
Firstly, a topologically correct graph of the street network is extracted. This step aims at giving correct spatial connections between streets as well as an approximation of their location. If available, road database is updated.
The next step is the actual street reconstruction. Due to the high resolution of the images, a surface reconstruction has to be performed. This step uses the previous step of graph management as an initialization for the reconstruction.
Finally, results are quantitatively characterized, using mathematical criteria (Heipke *et al.*, 1997; Couloigner and Ranchin, 1998).
An important point of the methodology is to separate the process in a step of topological management and an other of street reconstruction as surface element. First of all, it enables to focus separately on each step which are both of them very challenging. An other advantage of separating the algorithm is that it will enable potential user to check if the road network topology is correct during the process.
In the next sections, we describe more precisely the different parts of the process.

2.2 Evaluation and reference

As an operational image processing system has to be validated and its limits quantified, we aim at obtaining a measure of the quality of the results delivered by our method. This purpose needs the definition of a reference in order to compare it with extracted objects.

Despite the fact that a reference based on ground truth enables an accurate location of the object in the geographical space, a reference based on image interpretation is, in our case, preferable. Indeed, the comparison of the algorithm with visual interpretation enables to evaluate only the automatic process and not the whole image acquisition chain. Moreover, the image interpreter is an efficient pattern recognition system, able to use the context to extract the searched pattern. Though, considering a team of image interpreters and their subjective interpretations of the context, there is an important variability of such an estimation. This variability of interpretation has an effect when it comes to evaluate the efficiency of a method.

We have defined a methodology to extract a reference, in order to take the mentioned variability into account. Its implementation is currently under development and will not be presented in the real case study.
Our choice is an *a priori* reference, in a vector representation, based on statistics over several image interpreters, which leads to an "average" object (Péteri and Ranchin, 2002).
As the average can be not representative of a manual extraction, we also define a reference area which gives an *acceptance zone* for the result of an automatic extraction. An extraction method could then be considered as efficient as an human interpreter if the extracted object is included in this region.

2.2.1 Acceptance zone

Several image interpreters acquire the road contour in a vector representation via polylines. Figure 3 exhibits an acceptance zone, representative of the interpretation discrepancies between different interpreters.

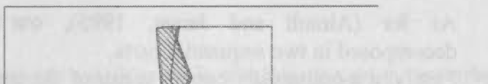


Figure 3. Acceptance zone

One consider the A_i surface inside the object contour acquired by the interpreter *i*. If the contour is open, it will be closed by a fictitious contour (see Figure 3).
The acceptance zone is then defined as:

$$Z_t = \overline{\left(\bigcup_i A_i\right)} \setminus \overline{\left(\bigcap_i A_i\right)} \qquad (1)$$

where A_i is the surface of one interpretation *i*, and the bar above is the closure of the set.
The quantitative evaluation is done by calculating what percentage of the extracted contour is included in this acceptance zone.

2.2.2 Extracted contour reference

We aim here at defining a "reference" object, representative of the different interpretations. For each polyline *i*, each segment *j* is represented by a slope vector $\mathbf{V}_{ij}$ (vectors are written in bold font).

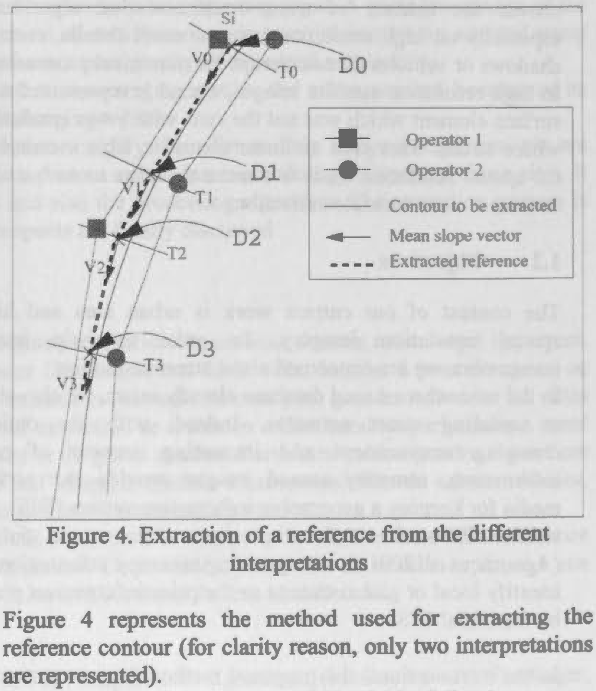


Figure 4. Extraction of a reference from the different interpretations

Figure 4 represents the method used for extracting the reference contour (for clarity reason, only two interpretations are represented).
The extraction of a reference polyline is the following:

- Firstly, the average vector $\mathbf{V}_0$ of the N slope vector is computed : $\vec{V}_0 = \frac{1}{N}\sum_i \vec{V}_{0i}$.

We then consider the transect (To) orthogonal to this vector and located to the extremity of the last polyline met in the $\mathbf{V}_0$ direction (vertice S_i on Figure 4).

An average is computed between points located at the intersection between (To) and the different polylines (rejecting external modes). The point D_0 is obtained and will be a vertice for the reference polyline.

- The transect (To) is "propagated"(in the direction of $\mathbf{V}_0$) to the first vertice met. The mean slope vector $\mathbf{V}_1$ is re-evaluated, as well as the transect (T_1) and the point D_1.

- This procedure is repeated till the final vertices of one of the polylines is met.

Quantitative criteria for comparing the extracted object to this "reference" object will be applied using (Heipke *et al.*, 1997 ; Couloigner and Ranchin, 1998).
For more details, see (Péteri and Ranchin, 2002)

2.3 The processing algorithm

As mentioned in section 2, the processing algorithm is composed in two sequential parts: a graph management step extracting the correct "skeleton" of the network, and a reconstruction step using active contours (snakes) initialized by the previous step.

2.3.1 Topology management of the graph

2.3.1.1 Graph management

At this step, a topologically correct graph of the street network is extracted. It aims at giving correctly spatial connections between streets as well as an approximation of their location. If available, road database will be updated at this step.
Polylines are currently entered by the user (Figure 5), without strong constraints on the segment location and on the number of vertices. The use of road databases is under development.

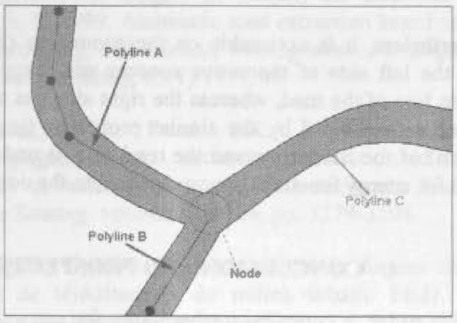


Figure 5. Graph of the road network

2.3.1.2 Initialization of the snake

This step also enables to initialize the contour of the reconstruction process: it is simply obtained by thickening the polyline (propagation of two parallel sides). More sophisticated approach could be to register the propagated contour by evaluating the gradient next to the polyline segment.

2.3.2 The reconstruction step: combining snakes and multiresolution

2.3.2.1 Snake model

Active contours (snakes) are suitable for road extraction, as they allow to overcome certain inherent problems like occlusions, width changes, surface material variations, and the effect of overpasses and intersections (Agouris *et al.*, 2001b). They have been first introduced by (Kass *et al.*, 1988).
As a snake seeks to minimize its overall energy, its shape will converge on the image gradient contours.
We use the greedy algorithm for contour extraction described in (Williams and Shah, 1992). This greedy approach claimed to have a significant speed advantage over the dynamic programming model introduced by (Amini *et al.*, 1990).

Energy in a snake can be represented by three energy terms: E_{con}, E_{curv} and E_{image}.

The total energy is therefore:

$$E_{snake} = \sum_{i} \alpha E_{cont}(i) + \beta E_{curv}(i) + \chi E_{image}(i) \quad (2)$$

where E_{Cont} is minimal when the points are evenly spaced and E_{Curv} is minimal when the snake has a constant first derivative (C^1 continuity, change in slope between any two segments is minimized).
E_{image} is the image force attracting the snake to contours with large image gradients. More precisely, $E_{image}(v) = -|\Delta I(v)|$, where $I(v)$ is the image intensity at point v.

For computation, the greedy algorithm takes an initial snake and iteratively refines the location of its points by looking at a "neighborhood" of pixels surrounding each and selecting the pixel of this neighborhood where the energy is minimized (see Figure 6).

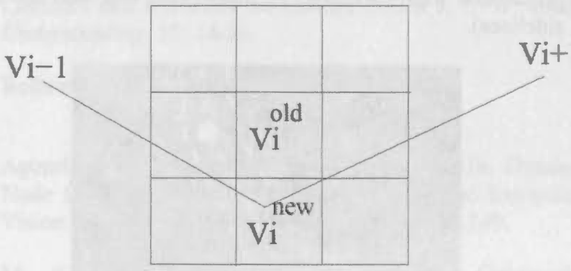


Figure 6. Optimization procedure

2.3.2.2 Combining snakes and multiresolution

As it was mentioned in the first section, artefacts such as vehicles or shadows, often encountered in the urban context may disturb the extraction process.
An approach to prevent the snake from being trapped by little artefacts such as vehicles or reservation is to use multiresolution. Multiresolution was indeed previously successfully used in (Baumgartner *and al.*, 1996), (Couloigner, 1998) and (Péteri *et al.*, 2001).
The algorithm includes this multiresolution approach: the snake is first applied on the coarsest resolution image, then on each intermediate resolution images until it runs on the original resolution.
An important point is to adapt the coefficients of the different energy terms to the image resolution. For instance, at the coarsest resolution, a high value of the image term allows the snake to be attracted from far. Refining the estimation on a finer resolution image is then done by releasing image constraints and increasing the importance of the internal energy.

3. REAL CASE STUDY

Here is presented the result of the extraction algorithm on an image from the IKONOS satellite of the region of Hasselt,

Belgium. The resolution of one pixel is 1 meter in the panchromatic band.
The road to be extracted is in fact a portion of a motorway crossing a railway station. This road presents a slightly curve (see Figure 7).

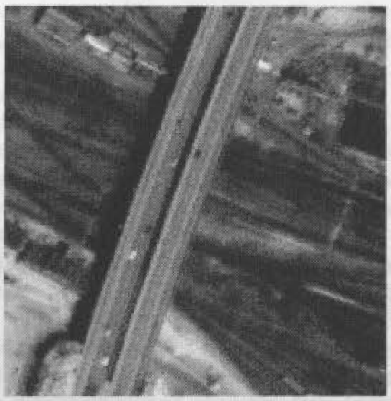


Figure 7. Original image for Hasselt

An operator has drawn a polyline very roughly (the dash and dot centerline on Figure 8). From this polyline, a contour was propagated with a width of 26 meters (the two dash and dot sidelines).

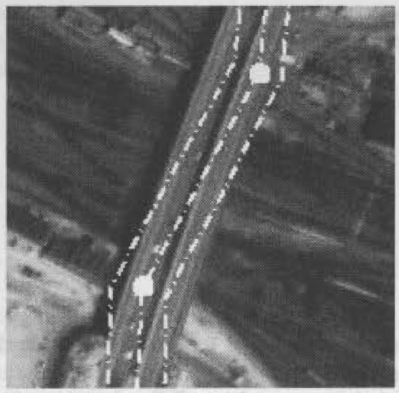


Figure 8. The polyline and the initial contour derived from it

The snake algorithm was applied on the image from this initialization of the contour. In the multiresolution scheme, the snake was first applied at the coarsest resolution of 8 meters. Figure 9 presents the result with the extracted contour (in plain line).
Compared to the rough initialization of the contour, the snake fits quite well to the motorway sides. The multiresolution has prevented the snake from being trapped by little artifacts with gradient of high magnitude (vehicles for instance).

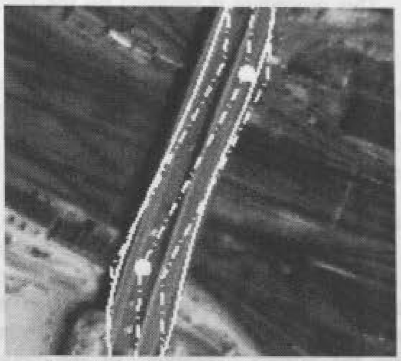


Figure 9. Result of the extraction (in plain line)

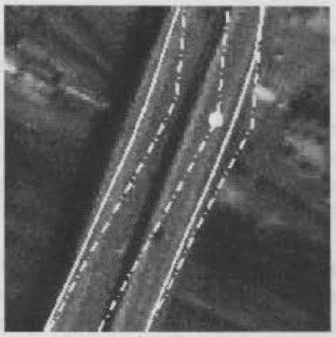


Figure 10. Zoom area

Nevertheless, it is noticeable on the zoom area (Figure 10) that the left side of the active contour was trapped by the inside line of the road, whereas the right side has succeeded. It can be explained by the similar properties (gradient and length) of the inside line and the road side. It underlines the need for energy functionals more adapted to the context.

4. CONCLUSION AND PROSPECTS

In this paper, a complete methodology for extracting streets from high resolution images was presented. To reach the reliability required by operational uses, a semi-automatic approach was chosen.
The first step of this method aims at extracting a topologically correct road network. In the second step, a reconstruction of roads as surface elements is proceed, using snakes combined with multiresolution. Snakes are initialized by the graph obtained at the first step of the algorithm.
The algorithm was then applied on an IKONOS image, leading to promising results. Due to the multiresolution approach, the reconstruction process in not disrupted by the numerous artefacts such as vehicles encountered in the urban context.
A way for improvement is to introduce more contextual information for constraining the reconstruction process to converge to the actual street sides. An important effort is currently held to find for the active contours some energy functionals proper to the urban context.

ACKNOWLEDGEMENT

Renaud Péteri has a fellowship from the Ministry of Defence (DGA/CNRS) for his Ph.D.

We would also like to thank the firm G.I.M. (Geographical Information Mapping) for having provided the IKONOS image of the Hasselt area.

5. REFERENCES

References from Journals:

Agouris, P., Stephanidis, A., Gyftakis, S., 2001a. Differential snakes for change detection in roads segments. *Photogrammetric Engineering & Remote Sensing*, 67(12), pp. 1391-1399.

Airault, S., Jamet, O., 1995. Détection et restitution automatique du réseau routier sur des images aériennes. *Traitement du Signal*, 12(2):189-200.

Amini, A., Weymouth, T., Jain, R., 1990. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):855--867.

Baumgartner, A., Steger, C., Wiedermann, C., Mayer, H., Eckstein, W., Hebner, H., 1996. Verification and multiresolution extraction. *Photogrammetric Engineering and Remote Sensing*, 31(3):53-58.

Baumgartner, A. Steger, C. Mayer, H. Eckstein, W. & Heinrich, E. 1999. Automatic road extraction based on multi-scale, grouping, and context. *Photogrammetric Engineering and Remote Sensing*, 65(7):777-785.

Bhattacharya, U., Parui, S.K., 1997. An improved backpropagation neural network for detection of road-like features in satellite imagery. *International. Journal of Remote Sensing*, volume 18, n°16, pp. 3379-3394.

Couloigner, I. 1998. Reconnaissance de formes dans des images de télédétection du milieu urbain. Ph.D. Thesis, Université de Nice Sophia-Antipolis.

Couloigner, I,. Ranchin, T., 1998. Extraction of urban network from high spatial resolution satellite imagery: assessment of the quality of an automated method, Proceedings of the EARSeL Symposium "Operational Remote Sensing for Sustainable Development", Enschede, The Netherlands, 11-14 May, Balkema, Rotterdam, The Netherlands, pp. 309-314.

Couloigner, I., Ranchin, T., 2000. Mapping of Urban Areas : A Multiresolution Modeling Approach for Semi-Automatic Extraction of Streets. *Photogrammetric Engineering and Remote Sensing*, 66(7):867-874.

Destival, I., 1987. Recherche automatique de réseaux linéaires sur des images SPOT. *Bulletin de la Société Française de Photogrammétrie et de Télédétection*, n° 105, pp. 5-16.

Gruen, A., Li, H., 1995. Road extraction from aerial and satellite images by dynamic programming. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50, 4, pp. 11-20.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of Automatic Road Extraction. *In International Archives of Photogrammetry and Remote Sensing*, volume (32) 3-2W3, pp. 47-56.

Jedynak, B., 1995. Modèles stochastiques et méthodes déterministes pour extraire les routes des images de la terre vue du ciel. Ph.D Thesis of the University of Paris Sud, January, 186 pages.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active Contour Models. *International Journal of Computer Vision*, pp.321-331

Laptev, I., Mayer, H., Lindeberg, T., Eckstein, W., Steger, C., Baumgatner, A., 2000. Automatic extraction of roads from aerial images based on scale space and snakes. *Machine Vision and Applications*, 12:23-31.

Merlet, N., Zérubia, J., 1996. New prospects in line detection by dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18, 4, pp. 426-430.

Wang, D. He, D.C. Wang, L., Morin, D., 1996. L'extraction du réseau routier urbain à partir d'images SPOT HRV. *International. Journal of Remote Sensing*, 17, 4, pp. 827-833.

Williams, D. J., Shah, M., 1992. A Fast Algorithm for Active Contours and Curvature Estimation Donna J. *CVIP: Image Understanding*. 55: 14-26.

References from Other Literature:

Agouris, P., Gyftakis, S., Stefanidis, A., 2001b. Dynamic Node Distribution in Adaptive Snakes for Road Extraction, Vision Interface 2001, Ottawa, Canada, pp. 134-140.

Mc Keown, D.M., Denlinger, J.L., 1988. Cooperative methods for road tracking in aerial imagery. In Proceedings of IEEE Computer Society Conference - Computer vision and pattern recognition, Ann Harbor, Michigan, pp. 662-672.

Péteri, R., Couloigner, I., Ranchin, T., 2001. A multiresolution modelling approach for semi-automatic extraction of streets: application to high resolution images from the IKONOS satellite. Proceedings of the EARSeL/SFPT Symposium "New Solutions for a New Millennium", Marne-la-Vallée, France, pp. 327-332.

Péteri, R., Ranchin, T., 2002. Assessment of object extraction methods from satellite images: reflections and case study on the definition of a reference. Proceedings of the EARSeL Symposium "Geo-information for European-wide Integration", Prague, Czech Republic. *(to be published)*

Stoica, R., Descombes, X., Zérubia, J., 2000. A Markov Point Process for Road Extraction in Remote Sensed Images. INRIA report n°3923, 38 pages.

Wang, J.F., Howarth, P.J., 1987. Automated road network extraction from Landsat TM imagery. *In Proceedings. of Annual ASPRS/ACSM Convention* ,Baltimore, MD, vol. 1, pp. 429-438.

TECHNOLOGICAL ASPECTS OF THE 3-DIMENSIONAL PHOTOGRAMMETRIC UPDATING OF THE ISRAELI NATIONAL GIS

Yuri Raizman[a], Ammatzia Peled[b]

[a] Survey of Israel, Tel Aviv, Israel
[b] University of Haifa, Department of Geography, Haifa, 31905 Israel
peled@geo.haifa.ac.il

IC WG II/IV

KEY WORDS: Updating, 3-D, GIS, Digital Mapping, Digital Photogrammetry

ABSTRACT:

The Survey of Israel is the National mapping Agency responsible for the National GIS. Maintenance, in terms of completion, updating and standards, among other, are the task of the Photogrammetric Division, in charge of six private companies that are producing new data and on the meticulous Quality Control processes, done in house.

This paper will describe the protocols of the 3-Dimensional digital photogrammetric updating of the National GIS spatial database. Mainly, the paper will focus on the principles and guidelines of the updating projects: (1) Quality Control of existing data in terms of aero-triangulation, digital elevation model, consistency and compliance with the newly developed standard for digital mapping in a topologically structured spatial environment; (2) Correction of all the irregularities found according to the above mentioned procedures; and only then (3) Updating of the digital spatial data base according to the changes taken place between the two epochs of mapping (air-photography missions).

In addition, the paper will focus on the administrative maintenance in terms of preserving the original Feature-Unique-ID and Source-Code that are kept, in the Israeli National GIS, for each feature. This according to the standard adopted by the Survey on October 1991, already.

Also, the paper will discuss the new experiments with incremental updating which is envisaged as the major tool of updating and maintenance of a national core spatial database.

PHOTOGRAMMETRIC MANAGEMENT GIS SYSTEM

Ammatzia Peled[a], Basheer Haj-Yehia[b]

[a] University of Haifa, Department of Geography, Haifa, 31905 Israel
peled@geo.haifa.ac.il

[b] Pelled GIS Mapping Ltd.,Haifa, 34995 Israel
basheer@rjb-3d.com

IC WG II/IV

KEY WORDS: Photogrammetry, Management, GIS, Air Photo Archiving, Mission Planning

ABSTRACT:

In the past, most of the photogrammetric processes, at the survey of Israel (SOI), were implemented manually. Only very few tasks were computer-assisted, in partial and a very limited format. The computerization at SOI included an alphabetical database of the photogrammetric data, such as mission, camera, triangulation, parameters. These parameters, in the past, were saved in separated and "private" files. These files served different users, in several departments, in terms of following-up only. All these data were kept without any geographical relations of the aero-photographs. This was due to the enormous amount of the aero-photographs data and the complexity of these data. This very complexity of past years lead to a decision that developing a spatial database and a photogrammetric management GIS system, is needed.

The photogrammetric management GIS system, discussed in this paper, was developed in order to integrate all the separate files, maintained by the different departments, and the Israeli National GIS coverages into one management system.

The paper describes, also, the major milestones of the newly developed system. The first stage included the collection and analyses of the different photogrammetric processes, which were implemented in different departments. These processes include mission planning, aero-photographs shop, etc. The second stage was to collect the digital and non-digital photogrammetric data at the SOI and the development of a GIS database, which integrate and combine different data types from different users. The third stage was the development of modular and user-friendly algorithms which enable: (a) automatic mission planning; (b) finding archive aerophotographs by geographic, date and other parameters; (c) automatic production of orthophoto mosaics (from single, stereo or multi photographs); (d) automatic production of standardized "orthophotomap"; and (e) updating the database in real time, including quality control to disable entering error data to the photogrametric database.

The newly developed system proved to be an efficient and most productive tool. Planned for the use of two sections only, the system is serving now all the sections of the division of photogrammetry. Further development is planned for year 2002 in terms of replacing old archiving techniques, pricing, customer service, and much more.

UPDATING AND SPECIFICATION CHANGES - STATISTICAL ESTIMATION OF WORKLOAD

Dorthe Holme and Peter Højholt

National Survey and Cadastre
Rentemestevej 8
2400 København NV
dh@kms.dk and phs@kms.dk

Commission IV, IC WG II/IV

KEY WORDS: Top10dk, Updating methods, workload, specification changes

ABSTRACT:

The vector dataset Top10dk is updated incrementally and new features are added as single objects and objects that are no longer valid are made historic. Single objects can be revised by choosing one of two methods depending on what is considered most suitable in the specific situation.

This paper will look into the usage of the two methods and conclude how the two methods are applied in the production of revision data for Top10dk. The paper analyzes statistical material based on the last three years production of revisions. Based on these statistics estimates of the extend of future updates are made.

Changes within the specification of Top10dk cause a changed workload within the production of revision data. The changed workload is assessed as a correlation between a specification change and the number of revision objects needed to implement the specification change. The experienced gained from the present investigation suggests that a statistical production monitoring system would benefit future workload estimates.

1. INTRODUCTION

Top10dk is the national topographic vector-database covering the entire Denmark in scale 1:10000. All data produced and updated adhere to strict topological rules. The initial establishment of Top10dk was finalized December 2001.

Top10dk can be separated in three main parts: the topographic base data, the place names and points of interest and the digital terrain model. This article will cover only the part related to the topographic base data.

2. SPECIFICATION CHANGES

The specification of Top10dk has changed over the years from 1995 when the first edition was written and up till 1999 when the sixth edition was published. The first specification had 40 feature types and the last one 51 feature types to map.

The specification has changed to accommodate comments and needs from both the users and private producers of Top10dk. Besides adding new features one of the main improvements of the specification for Top10dk has been the topology table, where the relationship between the different features is established.

3. UPDATING OF TOP10DK

The first updating on Top10dk was made based on aerial photos from 1998 in the region of Frederiksborg. From 1999 the full-scale production of updates for Top10dk was started. The area in an annual production covers approximately 20 percent of the country. In September 2001 42 % of the country have been updated.

The production of updates is organized as EU tenders and an annual production is typically split into 5 different contracts. Each contract is divided into two or three sub-areas with an app. size of 80000 ha.

The contract holder supplies KMS with data as photogrammetric models for control and data later data is sampled into 10 x 10 km blocks.

4. METHODOLOGY USED IN UPDATING

When updating Top10dk there are two basic methodologies to choose between. The first method is the object method where an object is either completely deleted or added. Changes are registered as a deletion of one object followed by an insertion of a new object. The object method can be applied on all feature types irrespective whether it is a point, line or polygon object.

The partial method can delete, change or add only part of an object. The method can only be applied to features of line or polygon types.

5. STATISTICS FOR THE USE OF THE PARTIAL REVISION METHOD COMPARED TO THE OBJECT REVISION METHOD.

5 different 10 x10 km blocks were selected for closer statistical investigation. The selected areas covered a range of specification changes and data from different data producers. The areas had been updated at varying intervals from 3 to 5 years. In the statistical analysis this was taken into account by representing the figures as changes pr. year.

It was found that in all areas the operators had used both the object method and the partial method for all types of revisions. Experienced operators across different companies seem to make similar choices of method.

But there were significant differences depending on object feature type.

For revisions of buildings both method was used to an almost equal extent. It is assumed that the low number of points in each building makes the methods evenly applicable.

For built-up areas operators almost always selected revision by the partial method. In these objects only small changes in few places along a (long) perimeter makes this method the fastest.

For small lakes and other similar labile boundaries operators most often chose the object method, reflecting the fact that almost the entire perimeter of the objects had to be redone.

6. CHANGES IN WORK LOAD CAUSED BY SPECIFICATION CHANGES.

Three different kinds of specification changes are identified:

- new features
- changes in the registration of existing features
- recoding part of a feature class to a new feature class

The new feature class that caused the most additional revision data was road boundaries. It increased the amount of revision data with approximately 15-20 %, and the amount of objects in the database by 10%.

The most significant example of changes in the registration of existing features was changes made in the registration of built-up areas. This change was present in two of the selected areas. The interpretation of statistical data for these cases will be included in the final paper.

Recoding parts of a feature class to a new feature class has caused no significant changes in the amount of revision data.

7. STATISTICS FOR CHANGES NOT RELATED TO CHANGES IN THE SPECIFICATIONS

These changes fall into two groups:

changes caused by real world changes

changes caused by errors in the initial registration.

Because of improvements in the error detection tools and routines, the level of errors detected in the production process has increased dramatically during the 6 years period covered by the selected data,.

At present it has not been possible to separate data for the two kinds of changes. Depending on the object feature change levels varies from 1% to 15 % pr. year, and it is at the moment difficult to reach a general conclusion.

8. CONCLUSIONS

Both the object method and the partial method of updating are used and it is justified to maintain both as parallel choices.

Including new major features in the dataset can increase the amount of revised objects substantially while most minor new features have almost no significance.

To estimate the future level of pr. year general updates it will be necessary to consider a detailed analysis that would eliminate the distortion causes by errors in the initial registration, or wait new productions to finish.

The experienced gained from the present investigation further suggests that a statistical production monitoring system would benefit future workload estimates.

INVESTIGATION ON LAND DATABASE UPDATING BASED ON HIGH RESOLUTION AIRBORNE IMAGES

Jie YIN[2] Guomin ZHU[1] Jiguo WANG[1] Chengming LI[2]

(1. Yiwu Design & Science Research Institute of Surveying and Mapping,

2. Chinese Academy of Surveying and Mapping)

cmli@casm.ac.cn

Abstract

China has the biggest population all over the world. The food is always a problem faced by Chinese government. Therefore, to protect the land, which can grow the grain on, from decreasing is very important. However, with the development of economy, great deal of lands have been occupied by the factories and urban building. Development will make the land less, the food demand from the land is not decreasing even increasing, how to balance the relations between development and increasing food demands is a controversial question. In China two measures were used to overcome the question, one is to increase the production per mu, the second is to protect the land from reducing. Monitoring the land changes and updating the existed land database is the base of the second ways.

There exist many methods to update the land database. Field surveying is suitable to a small area, remote sensing -based method is suitable to the large scope. For a middle scope such as 1000-10000 square kilometers, field surveying is not economic; the resolution of TM or SPOT image is not enough to monitor the changes of land. Maybe airborne- based method is rather suitable.

In this paper a method based on airborne was proposed to monitor the land changes and update the land database. According to the airborne image, there are two cases. One is that two period of time color airborne image have existed, the second is that one period of time image is black airborne image and another is color image. For different cases, the method has little difference. They are used in Yiwu city and was testified operable and useful to manage land effectively.

Keywords: Land database, the changes of land, update the land database, airborne- based method, remote sensing -based method

1. Introduction

Land resource is one kind of important resource, because it is related to the food of people. With the quickly deducing of land, which can grow grain, the measures must be taken to protect it from erosion. It demands the land management division know the detail information about the land and land changes. In 1988 the detail information on land was investigated all over the China, thirteen years later, it has changed a lot. The land resource database needs to be updated. If the line graph was used as base map in traditional method, in fact the line graph was been produced ten years ago in most parts of China, especially in west China. The precise of the land resource gotten according to this kind of base map is not enough to update land database. Furthermore, these base maps only were updated every ten years or fifteen years. Obviously, the above idea is not suitable to be used to update the land database at present.

Of course the total field surveying can get the high precise land detail information, but it will last several years and also will spend much money. This can't meet the demands of land management, can't match the speed of economical and modern technology development. So this line also is abandoned.

Airborne and remote sensing Image-based method was used in Chinese land resource updating, remote sensing image such as TM/SPOT is suitable for large scope even the total China. Only is the airborne image suitable for the middle and small urban territory.

Following the introduction, the existing work was reviewed, part three proposed the basic line of land database update, and several key steps were described in detail. Part four gave an example. At last, the concluded remarks were drawn.

2. Existed research works

From the Internet, many papers and web pages can be queried. After studying it closely, it is not difficult to know that image-based land database update was rather matured in abroad. However, if these technologies were extended to use in China, there will occur some problems, because the actualities of the land resource and land resource using are rather complex. Furthermore, the difference to use modern technology between rich territory and poor territory is much great in China.

Under the support of national department of land resource of China, our academy, with other two institutes, began the land resource of China investigation in 1998, and it will last about twelve years. This work is based on TM/SPOT image. But this method is not suitable to the middle and small areas.

3. Basic Line and Methods

Several years ago, the line graph map was used as the base map in traditional methods. It lagged the progress of scientific technology. At present, the remote sensing image TM/SPOT or airborne image will substitute for the traditional line graph as base map. The total process will be divided into the following steps:

3.1 Ortho-image

The projection mode of the airborne image is a central projection. It needs to be transferred into the ortho-projection mode. The process includes air-photo scanning, selecting control points in topographical maps, air-triangulation, image mosaic, colors and brightness adjustment. After the total digital photogrammetry, the digital elevation model (DEM) and ortho-image can be gotten. These are the base maps of land detail information investigation. Sometimes, the airborne image is black, we had better merge it with TM image, make it change into color and high resolutions image. It will be helpful to extract land resource information.

3.2 Image interpretation and field checking

On the computer the image processing software such as Erdas, PCI and ERMAP were used for drawing the land boundaries, statistic the area, and justify which kind of land it is with the help of experienced RS workers. 10-20 percent results of these interpretations will be testified by field investigation or checking.

3.3 Digitization of modifying information in field checking

After field checking, part results will modified in paper base map. These changes information will return to the experienced RS workers, they will put the changed land information into the computer.

3.4 Updating the land detail investigation database

The software such as arc/info is used to overlap old land resource information and the new investigation information. By the software functions developed by us find the changes automatically, and label it in bright color. If operation workers are sure it is right, the changed land resources information will go into database and update the old information automatically. If they are sure it is right, the changed information will be ignored. Through this process, the detail land investigation database, which is built in 1988, will be updated.

4. An example

The Yiwu city was selected as the test territory. The total area of Yiwu testing territory is about 1100 square kilometer. It has land database built in 1995, and the color airborne image, which were taken in1999. This test was finished in2000. The workflow is showed in Figure 1.

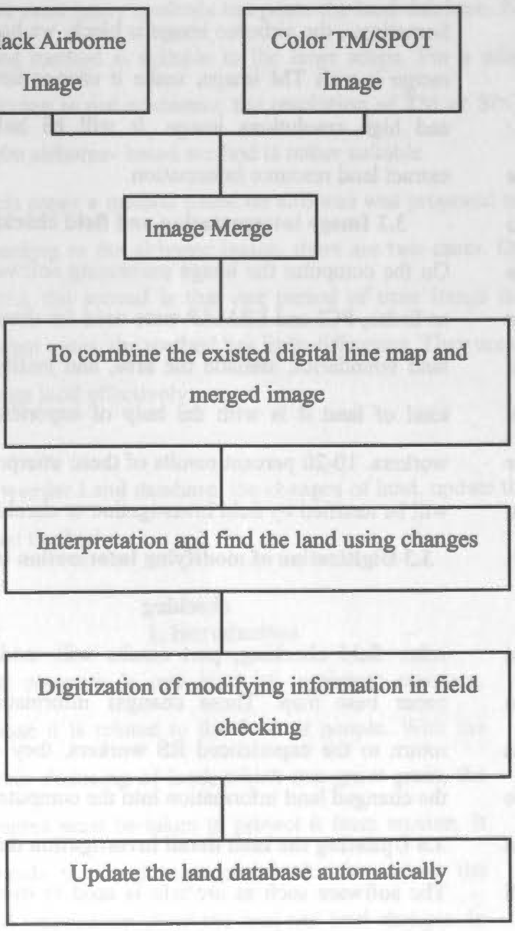

Black Airborne Image
Color TM/SPOT Image
Image Merge
To combine the existed digital line map and merged image
Interpretation and find the land using changes
Digitization of modifying information in field checking
Update the land database automatically

Fig. 1 Workflow of Updating land using changes

At last after finishing updating, the results of land using changes were gotten and listed in table 1.

No.	Before changes	After changes	Area(mu)
1	infield	Build field	11004
2	Non-infield	Build field	718
3	Infield	pothole	8.2

5. Concluded remarks

In this paper after reviewing the traditional methods in China to use for updating land resource database and actualities in abroad, an airborne image-based method was proposed. And the detail steps were also introduced. By the test territory experiment, this kind of method to update land database is effectively. The precise is rather high, it can meet the needs of land management and decision support of government.

Reference

1. Chengming LI et al, 2000, Scheme of Yiwu urban Geographical Information System. Research Report.

2. Hakima Kadri-Dahmani, 2001, Updating Data in GIS: Towards a more generic approach. Proceeding 3 of The 21th International Cartographic Conference. August 6-10,2001, in Beijing. Pp. 1463-1470.

3. Uitermark. H, Oosterom .P, Mars. P, Molenaar. M, 1998, Propagating Updates Corresonding Objects in a Multi-source Environment. Proceeding 8th International Symposium on Spatial Data Handling, pp. 202-213, 1998.

4. H. Samet, Neighbor finding in images represented by octrees ,Computer Vision[J] ,Graphics, and Image processing, 1989, 46, 367~386

5. Egenhofer M and Herring J. Categorizing binary topological relationships between regions, lines and points in geographic databases. In: A Framework for the Definition of Topological Relationships and an Approach to Spatial Reasoning within this Framework, edited by Egenhofer M, Herring J, et.al. Santa Barbara, CA: 1991. 1-28.

DIFFERENTIAL OBJECT EXTRACTION METHODS FOR AUTOMATED GIS UPDATES

P. Agouris*, A. Stefanidis, S. Gyftakis, G. Mountrakis

Dept. of Spatial Information Engineering, National Center for Geographic Information and Analysis,
University of Maine, 348 Boardman Hall, Orono, ME 04469-5711, USA
{peggy; tony; sotiris; giorgos}@spatial.maine.edu

WG IC II/IV

KEY WORDS: Change Detection, Object Extraction, GIS Updates, Snakes, Template Matching

ABSTRACT:

In modern geospatial applications object extraction becomes increasingly part of larger cycles of GIS updates. In such update cycles, the objective is to compare new information to older one, and to identify changes that occurred in the meantime. In this paper we present an image-based GIS updating framework and corresponding image analysis algorithms developed by our group to automate GIS updates. More specifically, we present an overview of our work on two different types of objects to be extracted and monitored, roads and buildings, and the corresponding algorithms, namely differential snakes and differential template matching. Our approach to both problems is characterized by the comparison of a new image to pre-existing information through an automated image analysis tool. This is equivalent to comparing an object as it is represented in an image to the same object represented in a GIS at a prior instance, setting up a novel matching problem, whereby an object is compared to itself to identify how it has changed. For both roads and buildings our algorithms make use of accuracy estimates accompanying the pre-existing information, to ensure the meaningful update of geospatial databases. Making use of these accuracy measures we differentiate change detection and versioning. In this paper we present an overview of theoretical issues behind our algorithms and experimental results from the developed software solutions.

1. INTRODUCTION

Object extraction from digital imagery is a key operation for modern geospatial applications. The evolution of the current state-of-the-art among digital image analysis and computer vision activities on the subject of automated object extraction from digital aerial and satellite imagery may be found in Suetens et al. (1992), Gruen et al. (1997), and Lukes (1998). Attempting to summarize the current state-of-the-art in this area we could point out that currently existing solutions are semi-automatic, with a human operator providing manually approximations of the location and shape of the object to be extracted. Then, an automated algorithm uses these approximations to precisely delineate the object's outline, or centerline as might be the case in roads.

In modern geospatial applications object extraction becomes increasingly part of larger cycles of GIS updates. In such update cycles, the objective is to compare new information to older one, and to identify changes that occurred in the meantime. In this paper we present an image-based GIS updating framework and corresponding image analysis algorithms developed by our group to automate GIS updates.

More specifically, we present an overview of our work on two different types of objects to be extracted and monitored: roads and buildings (and similar structures). Our approach to both problems is characterized by the comparison of a new image to pre-existing information through an automated image analysis tool. This is equivalent to comparing an object as it is represented in an image captured at instance *T* to the same object represented in a GIS at *T-dt*. This sets up an interesting matching problem, whereby an object is compared to itself, to identify how it has changed. The algorithms we developed to support this type of differential image analysis make use of accuracy estimates accompanying the pre-existing information, to ensure the meaningful update of geospatial databases. This allows the differentiation of change detection from versioning, an important distinction to improve information flow within modern geospatial databases.

In this paper we present an overview of the theoretical issues behind the algorithms we developed to monitor roads and buildings. Regarding *roads*, we have extended the model of deformable contour models (snakes) to function in a differential mode, producing the concept of differential snakes. Regarding *buildings*, we have developed differential application of template matching to compare the content of an image to the record of a building in an older GIS database and identify changes in it.

The paper is organized as follows. In Section 2 we present our view of image-based change detection and versioning. In Section 3 we present an overview of our work on differential snakes for the updating of road segments. In Section 4 we present an overview of our work on differential template matching to update building outlines. Experimental results form our algorithms are presented in Section 5, and concluding remarks are presented in Section 6.

---

* Corresponding author.

2. IMAGE-BASED CHANGE DETECTION

Image-based change detection processes typically proceed in two steps. First, a complex algorithm (e.g. snakes, template matching) is used to identify objects in a new image. Then, the GIS database is updated by comparing the newly extracted outline to the prior recorded information. If the new outline is different, we commonly proceed by using it to replace the older information and thus update the database. In this context, information is treated as deterministic in nature: any difference between the two outlines is considered as change. This often results in storing multiple slightly different representations of an object, even though this object has actually remained unchanged.

We aim to remedy this problem by integrating object extraction and change detection in a single process. It is meant to function within an integrated geospatial environment, whereby image analysis proceeds by having access to pre-existing information for the processed area. We assume a process where a new image is analyzed to determine changes in and update the existing GIS of a specific area. Within this environment pre-existing GIS information provides us with shape information for geospatial objects (e.g. roads, buildings) and accuracy estimates for this information. This prior information may have been produced by prior image analysis processes (exploiting older imagery), or by any of the other established methods to collect GIS information (e.g. traditional surveying processes).

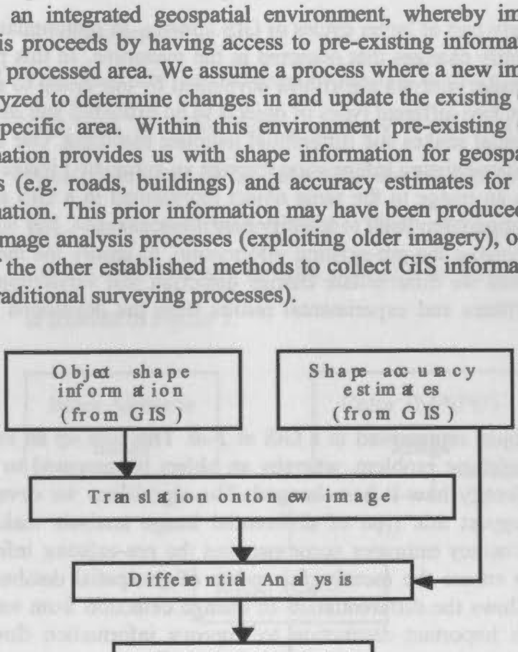

Object shape information (from GIS)
Shape accuracy estimates (from GIS)
Translation onto new image
Differential Analysis
Change detection
Versioning
GIS Updating

Figure 1: GIS updating process

A schematic overview of our process is described in Fig. 1. As mentioned above we make use of pre-existing object outline information. The older outline is projected onto the new image, using standard orientation parameters and relevant transformations. Once projected onto the new image, we proceed with *differential image analysis*, whereby the older outline together with its accuracy information becomes the input for a GIS update cycle. This update cycle comprises two processes:

- a mandatory *change detection* process, whereby we identify any parts of an object that have changed since its last record, and
- an optional *versioning* process, whereby we replace an otherwise unchanged object (or object segment) by a higher accuracy version of it, if the current imagery so permits.

The results are used to update the GIS information for this object by modifying its outline and/or updating its uncertainty estimates.

Following common practice in computer vision, we have developed two distinct differential analysis algorithms for smooth curvilinear (e.g. roads, rivers) and regularly shaped (e.g. buildings) objects. Details of these two algorithms are presented in the following two sections.

3. DIFFERENTIAL SNAKES FOR ROAD UPDATES

Deformable contour models (a.k.a. snakes) have been developed in the computer vision community as object extraction tools (Cass et al., 1987). In its numerical solution the snake is represented by a polygonal line, defined by nodes and segments connecting these nodes. The geometric and radiometric relations of these nodes are expressed as energy functions, and object extraction becomes an optimization problem (see e.g. (Williams & Shah, 1982) for an appropriate optimisation algorithm).

In a traditional snake model the total energy of each snake point is expressed as:

$$E_{snake} = \alpha \cdot E_{cont} + \beta \cdot E_{curv} + \gamma E_{edge} \qquad (1)$$

where : E_{cont} , E_{curv} are energy terms expressing first and second order continuity constraints (internal forces); E_{edge} is an energy term expressing edge strength (external force); and α, β, γ are (relative) positive weights of each energy term. For brevity we avoid further analysis of the snakes model here. The reader is referred to (Cass et al., 1987) and sub-sequent publications of these researchers for further details on the formulation of these parameters.

To support change detection we expand the traditional snake model to perform a comparison of the current image content to the prior outline projected on it (and its uncertainty measures) instead of standard object extraction. We use the term *differential snakes model* to refer to this model, as it is used to identify differences. In this differential model, the snake solution is constrained not only by the radiometric and geometric terms of Eq. 1, but also by the pre-existing information. We accomplish that by expanding Eq. 1 to introduce an additional energy term E_{unc} and a corresponding (relative) weight δ:

$$E_{snake} = \alpha \cdot E_{cont} + \beta \cdot E_{curv} + \gamma \cdot E_{edge} + \gamma \cdot E_{edge} + \delta \cdot E_{unc} \qquad (2)$$

The additional energy term E_{unc} describes the discrepancy between the current snake solution and the pre-existing information. It has an effect similar to that of a spring attracting the current snake solution towards its prior record. Change is detected if and only if the gray value content of the new image supports the notion that the object has moved beyond the range of attraction of its older record, breaking its spring-like effect. If the new image content is only suggesting a small move within the limits of the older information's attracting force we do not detect change. In this case, the spring-like force will keep the snake in its earlier location, avoiding the extraction of yet another version that does not differ statistically from its older version.

3.1 Modeling Uncertainty

The above presented differential snakes model is making use of expressions of the uncertainty with which information can be extracted from an image. To model this type of uncertainty we use a method based on fuzzy logic.

Uncertainty information resides implicitly in the values of snake energy E_t and the rate of energy change (DE_t) along a road segment. The analysis of these values at various locations along the snake contour describes how well the extracted contour approximates an ideal road model (as it is expressed by Eq. 1) at these points. More specifically, we use snake energy information to generate uncertainty values (U) using fuzzy linguistic rules. One can select various sets of linguistic values to express the range of energy, uncertainty, and energy rate values. In our approach we use: E_t={low, medium, high}, DE_t={low, medium, high}, and U={low, medium, high}, with each membership function following a Gaussian shape. A fuzzy rule base is the generated, expressing the interrelationship between energy, energy gradients and uncertainty. A sample from this fuzzy rule base is:

- *If E_t is LOW and DE_t is LOW, then U is LOW*

Through this analysis we can assign uncertainty estimates to all points along an extracted outline. The uncertainty coefficients (in the range 0-1) are saved together with the coordinates of the points. By multiplying an uncertainty coefficient by a global accuracy measure we obtain pixel accuracy measures for points along a snake. The global accuracy measures are expressions (in pixel units) of the expected accuracy in extracting a linear object from a specific image.

The additional energy term of Eq. 2, describing the effect of prior information (and its corresponding measures of uncertainty), is expressed conceptually as:

$$E_{unc} = f\,[Unc(v0_i),d] \qquad (3)$$

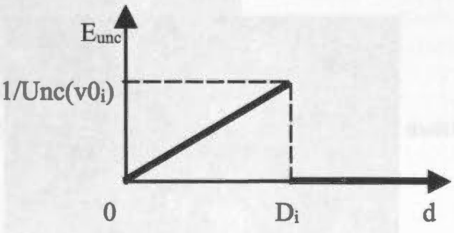


Figure 2: Uncertainty energy (E_{unc}) diagram

This additional energy term at a specific point is a function of the uncertainty with which we know the corresponding prior information ($Unc(v0_i)$), and the distance (d) of the current point from the older outline. Drawing from physics we define this energy term to be proportional to the distance between the current solution and the prior outline, inversely proportional to the corresponding uncertainty of the older outline, and acting within a threshold D_i. Beyond the threshold, this energy component is cancelled. In mathematical terms, this energy function is given by the set of equations:

$$\begin{aligned} E_{unc} &= [1 / D_i \cdot Unc(vO_i)] \cdot d \quad (\text{if } d < D_i) \\ E_{unc} &= 0 \qquad\qquad (\text{if } d > D_i) \end{aligned} \qquad (4)$$

where $Unc(v0_i)$ and D_i are as defined above, and d is the distance between current point (v_i) and original outline. The diagram of this function is shown in Fig. 2.
Based on this modeling of uncertainty energy, when prior information is known with high uncertainty (low accuracy), its equivalent force is small but operates over a rather large interval. When it is known with low uncertainty (high accuracy), its equivalent force is strong but operates over a short interval.

The threshold D_i is a parameter that defines the neighborhood over which the prior information is actively affecting the current object extraction process. Its statistical meaning is equivalent to the selection of confidence intervals in statistical analysis (e.g. selecting a global interval D_o for a complete curve to be equal to 3 times the standard deviation of our solution). Accordingly, the local evaluation D_i of this parameter can be defined at each location along an outline as:

$$D_i = D_o \cdot Unc(vO_i) \qquad (5)$$

where $Unc(v0_i)$ = local uncertainty value, in the range (0,1), and D_0 = global threshold. A more detailed description of the differential snakes model may be found in (Agouris et al., 2001b).

4. DIFFERENTIAL TEMPLATE MATCHING

While the differential snakes method presented in section 3 above is suitable for roads and other curvilinear smooth objects, it does not support operations on buildings and other similar objects. The outlines of such box-like objects violate the smoothness criteria of Eq. 1, forcing the extracted outlines to overshoot corners. Accordingly, we have developed a technique that makes use of least squares template matching to detect changes in buildings.

Our change detection method employs least squares matching for the detection and tracking of edges in digital images. Using prior information we generate a window depicting an edge pattern, and introduce it as a reference template. This reference template will be matched to digital image patches in the vicinity of actual edge segments. The concept behind the method is simple yet effective: by matching the edge template window to an image window, we can identify edge locations in the image as conjugate to the a priori known template edge positions [Gruen & Agouris, 1994].

Assuming *f(x,y)* to be the reference edge template and *g(x,y)* to be the actual image patch, a matching correspondence is established between through least squares matching, observation equations can be formed relating the gray values of corresponding pixels, and they are linearized as:

$$f(x,y)-e(x,y) = g^o(x,y)+[fg^o(x,y)/fx]\,dx+[fg^o(x,y)/fy]\,dy \qquad (6)$$

The derivatives of the image function in this equation express the rate of change of gray values along the *x* and *y* directions, evaluated at the pixels of the patch. Depending on the type of edge, the geometric relationship describing the two windows may be as complex as an affine transformation, or as simple as a simple shift and/or rotation. Regardless of the choice of geometric transformation, the resulting observation equations are grouped in matrix form as:

$$-e = AX - l\,;\; P \tag{7}$$

In this system, l is the observation vector, containing gray value differences of conjugate pixels. The vector of unknowns x comprises the shift at the x direction, while A is the corresponding design matrix containing the derivatives of the observation equations with respect to the parameters, and P is the weight matrix. A standard east squares solution allows the determination of the unknown parameters as:

$$X = (A^TPA)^{-1}A^TPl \tag{8}$$

While the above formulas reflect a standard template matching method, our problem introduces certain challenges. Indeed, comparing templates of the same object in various time instances and often captured by different cameras introduces large amounts of noise. In order to optimize the performance of our template matching method, we have to minimize the effect of radiometric variations among the two images (e.g. due to noise, differences in general histogram properties, or even different resolutions). Towards this goal we have developed a technique that analyses the content of matching windows to identify edges in them and assign higher weights to these locations (Agouris et al., 2000). This allows the solution of Eq. 8 to focus mostly on the information conveyed by outlines and thus be less susceptible to changes in common radiometric variations.

The variance-covariance matrix Σ_X of the solution of Eq. 8 expresses the error in locating the object's outline, and can be used to differentiate between change detection and versioning in a manner similar to the process outlined in the previous segment. A more detailed description of our differential template matching approach for change detection in buildings may be found in (Agouris et al., 2000).

5. EXPERIMENTS

The differential image analysis methods presented in this paper has been implemented in a PC, using Matlab. Fig. 3 and 4 present the application of our differential snakes technique to update road segments. In Fig. 3 we can see a prior record of a road's outline projected onto th enew image. In this specific application the outline was extracted from an older image using a snake solution, and points 1-5 are nodes from that snake solution. Of course, prior information may have been extracted with any of available techniques, and then points along the outline may be selected using any type of sampling technique. The circles around the nodes of Fig. 3 are a visualization of the corresponding accuracy measures. It is easy to see that nodes 1 and 2 are available with higher accuracy than nodes 3-5.

Fig. 4 shows the result of change detection (top) and versioning (bottom) for the data of Fig. 3. By comparing these figures it can be easily seen how the new road segment (as captured in the image) lies within the threshold of nodes 3-5, but clearly beyond the threshold of nodes 1 and 2. Accordingly, during change detection nodes 1 and 2 move to the correct location and this change is recorded (dashed line in Fig. 4 top). Subsequently, nodes 3-5 are moved during versioning, to improve the accuracy of the recorded outline. The result of versioning is marked as dashed line in Fig 4 bottom. Together, change detection and versioning comprise a complete updating process for this road segment.

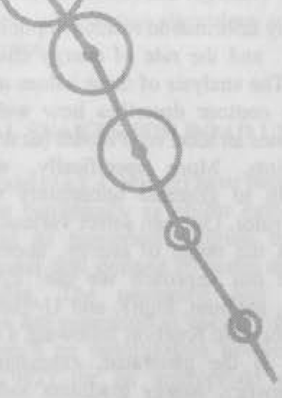


Figure 3: Pre-existing information (outline and accuracy information) projected onto a new image window.

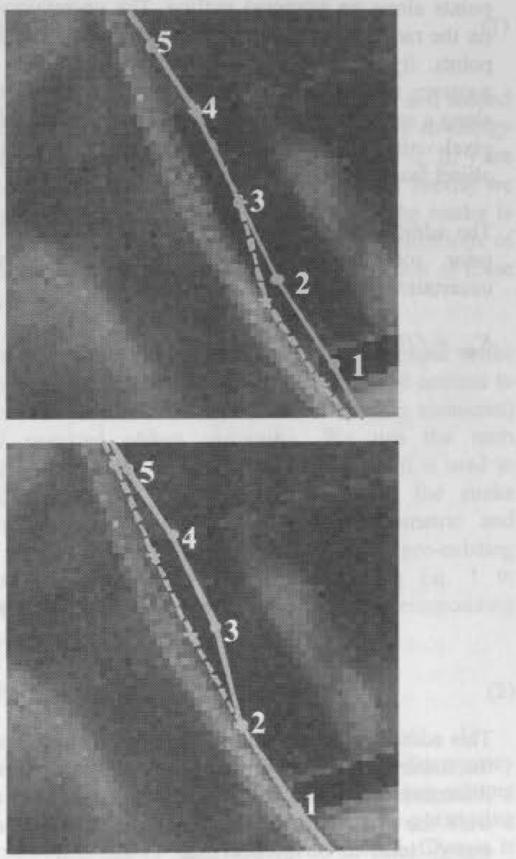


Figure 4: The results of change detection (top) and versioning (bottom).

In our applications, the average time required to perform change detection for a typical road segment was 2.1 sec, while versioning required 0.6 sec, for a complete cycle requirement of less than 3 sec per road segment (using a 1.1GHz processor). When proceeding with a node redistribution process to best describe the geometric complexity of the extracted outline

(Agouris et al., 2001a) we can expect an additional 0.4 sec in computational time per segment.

Fig. 5 shows an application of differential template matching to update a building outline. An older outline has been projected onto a new image. The lower part of the figure shows a zoomed window of the shaded area of the top. We can see that the older outline (thick black line) did not include the new addition (at the bottom left of the figure). Our algorithm proceeded by checking points along the outline as indicated by the check marks and crosses in Fig. 5. A check mark indicates a successful match (where established that nothing changed), while a cross indicates a failure (and therefore marks change). We can see crosses marking correctly the part of the older outline where the new wing was added. The spacing of the points to be checked along the outline depends on the resolution of change that we are after. It also affects in an obvious manner the time required to perform this differential analysis process. For the set-up of Fig.5 the time required to complete the process was approximately 1 minute.

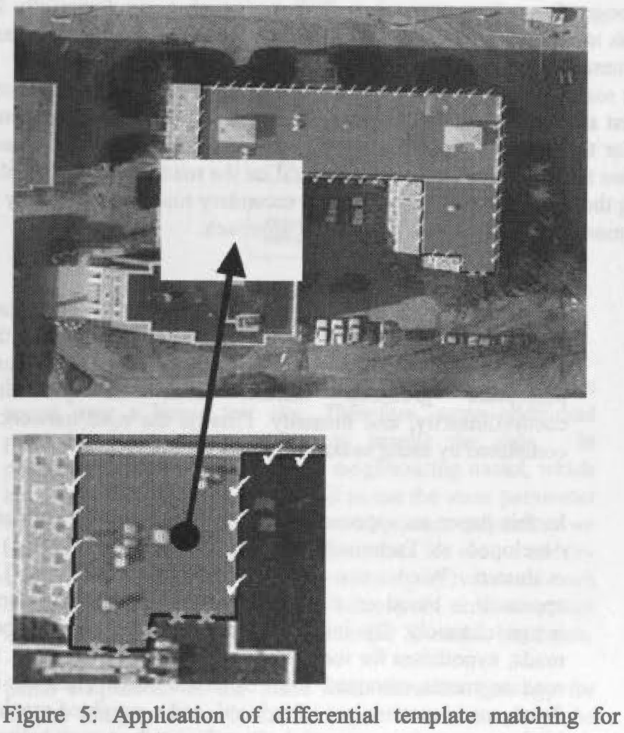


Figure 5: Application of differential template matching for building updating.

6. CONCLUSIONS

The accelerated rate of change of modern environments is bringing forward the need for frequent updates of GIS databases. Digital imagery is highly suitable for this task, as advances in sensors and platforms have expanded its availability to unprecedented levels of spatial and temporal coverage. In this paper we presented an image-based GIS updating framework and corresponding image analysis algorithms developed by our group to automate GIS updates. They can handle both smooth curvilinear features (like roads) and regularly-shaped ones (like buildings). Our approach is characterized by the comparison of a new image to pre-existing information through an automated image analysis tool. By making use of pre-existing information in the form of outlines and accuracy estimates we establish a fully automated updating process. This allows us to update the record of a road segment in less than 5 seconds, and to examine a complete building outline in approximately 1 minute. We are also able to differentiate between change detection and versioning, eliminating the confusion caused by the recording of numerous slightly different records of an object that has actually remained unchanged. We are currently working on integrating our techniques with work on scene similarity metrics to support the updates of abstract GIS views, and on extending the application of the techniques presented here to handle moving objects.

7. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation through awards CAREER IIS-9702233, DGI-9983445, and ITR-0121269. by the National Imagery and Mapping Agency through NURI grant number NMA202-98-1-1113, and by BAE Systems.

8. REFERENCES

Agouris P., K. Beard, G. Mountrakis, & A. Stefanidis, 2000. Capturing and Modeling Geographic Object Change: A SpatioTemporal Gazetteer Framework, *Photogrammetric Engineering & Remote Sensing*, 66(10), pp. 1241-1250.

Agouris P., S. Gyftakis & A. Stefanidis, 2001a. Dynamic Node Distribution in Adaptive Snakes for Road Extraction, *Vision Interface '01*, Ottawa, Canada, pp. 134-140.

Agouris P., A. Stefanidis & S. Gyftakis, 2001b. Differential Snakes for Change Detection in Road Segments, *Photogrammetric Engineering and Remote Sensing*, 67(12), pp. 1391-1399.

Gruen A., E. Baltsavias, and O. Henricsson (eds.), 1997. *Automatic Extraction of Man-Made Objects from Aerial and Space Images II*, Birkhaeuser Verlag, Basel, 364p.

Kass M., A. Witkin, & D. Terzopoulos, 1987. Snakes: Active contour models, in *Proceedings of the 1st Int. Conf. on Computer Vision*, London, pp. 259-268.

Lukes G. (ed.), 1998. *Proceedings of DARPA Image Understanding Workshop*, Morgan Kaufmann, New York, NY

Suetens P., P. Fua & A.J. Hanson, 1992. Computational strategies for object recognition, *ACM Computing Surveys*, 24(1): 5-61.

Williams D. J. and M. Shah, 1992. A Fast Algorithm for Active Contours and Curvature Estimation, *CVGIP: Image Understanding*, 55, pp. 14-26.

EVALUATION OF AUTOMATIC ROAD EXTRACTION RESULTS FROM SAR IMAGERY

B. Wessel[a, *], C. Wiedemann[a], O. Hellwich[b], W.-C. Arndt[c]

[a] Chair for Photogrammetry and Remote Sensing, Technische Universität München, 80290 Munich, Germany - (birgit.wessel, christian.wiedemann)@bv.tu-muenchen.de
[b] Photogrammetry and Cartography, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany - hellwich@fpk.tu-berlin.de
[c] Infoterra GmbH, 88039 Friedrichshafen, Germany,wolf.christian.arndt@astrium-space.com

Commission IV, IC WG II/VI

KEY WORDS: road extraction, SAR imagery, evaluation, road classes

ABSTRACT:

Automatic and semi-automatic methods for the extraction of topographic information are a present research topic. Especially for topographic mapping and navigation applications road extraction is important, but still not in practice. Due to the speckle effect and the limited width of roads with respect to the ground resolution automatic road extraction from SAR imagery is rather difficult.
In this paper we want to investigate the potential of an automatic road extraction approach of the Technische Universität München, Germany. We apply the road extraction approach to two large test sites and evaluate the achieved results to two kinds of reference data: topographic map data and manually plotted vector data. For this purpose, the reference data are separated into three classes: highways, main roads, and secondary roads. The comparison shows that the extraction results depend on the road classes. Presently, for main roads quite satisfying results can be achieved. Regarding the completeness, highways and secondary roads cannot totally be extracted automatically, due to their width, visibility or a missing model for highways in the present approach.

1. INTRODUCTION

Automatic road extraction has been a research topic since several years, from optical images as well as from SAR images. An early approach for detection of roads in low-resolution aerial imagery comes from (Fischler et. al., 1981). In a first step, two kinds of detectors based on local criteria are used and the responses are combined. Then, in a more globally step, the road network is extracted by either a graph search or dynamic programming. This approach was also applied to SAR images by e.g. (Samadani and Vesecky, 1990). For high-resolution imagery (McKeown and Denlinger, 1988) set up a road model for their road tracking algorithm. (Bazohar, and Cooper, 1996) used this approach for an automatic road extraction by defining Markov random fields (MRF). The roads were detected from it by a local maximum a posteriori probability (MAP) estimation.

Automatic extraction of linear features from SAR images especially taking into account the statistical properties of speckled SAR images is done by (Tupin et. al., 1998) and (Kartartzis et. al., 2001). (Tupin et. al., 1998) perform a local detection of linear structures based on two SAR specialized line detectors. The results are fused and the candidates for road segments are organized as a graph. The completion of the network is realized by a MRF. With a priori knowledge about the roads available by the MRF a maximum a posteriori probability (MAP) criterion is identifying the best graph. (Kartartzis et. al., 2001) improve the approach from (Tupin et. al., 1998) and integrated the morphology method of (Chanussot and Lampert, 1998) for selecting road regions for an automatic extraction of roads from airborne SAR images. (Jeon et. al., 2002) apply road detection to space borne SAR images. Roads were detected as curvilinear structures and grouped to segments using a generic algorithm (GA), which is a global optimization method. The GA uses perceptual grouping factors, such as proximity, cocurvilinearity, and intensity. Finally, the road network is completed by using snakes.

In this paper an approach for automatic extraction of roads developed at Technische Universität München (TUM) is evaluated (Wiedemann et. al., 1998, 1999). The TUM approach is based on the extraction of lines from different image channels. By introducing explicit knowledge about roads, hypotheses for road segments are generated. Then, the road segments extracted from different image channels are fused, road junctions are introduced, and a weighted graph of road segments is constructed. In order to close gaps between road segments, weighted links are added to the graph. Finally, a road network is extracted connecting seed points by optimal paths through the weighted graph.

In this paper we want to investigate the potential of the TUM approach for automatic extraction of roads from airborne SAR imagery. The road extraction strategy was developed for optical imagery, so we present some modifications to adapt the algorithm to SAR imagery. We carry out extensive experiments on two larger test sites of about 110 km^2 in total and evaluate the achieved results by comparing the road extraction results to reference data regarding different road classes. Thus, we distinguish between the content of a topographic map and what a skilled operator is able to detect in the data.

---

* Corresponding author

2. EXTRACTION STRATEGY

The extraction strategy is composed of different steps, which are shown in Figure 1. In the following a short description of each step is given, with emphasis on the adaptations made for SAR and for large data sets.

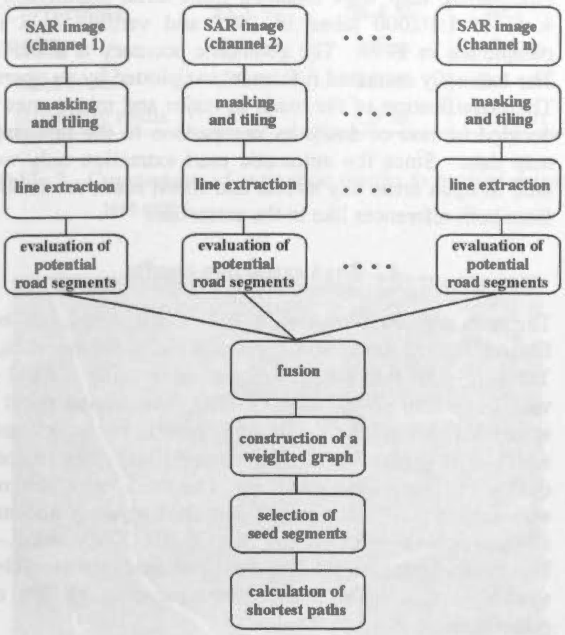

SAR image (channel 1)
SAR image (channel 2)
....
SAR image (channel n)
masking and tiling
masking and tiling
....
masking and tiling
line extraction
line extraction
line extraction
evaluation of potential road segments
evaluation of potential road segments
....
evaluation of potential road segments
fusion
construction of a weighted graph
selection of seed segments
calculation of shortest paths

Figure 1. Road extraction workflow

Pre-processing: For the first time, the TUM approach is tested over a larger test site. Therefore, some additional preparing steps are necessary to handle the data. In particular, test site 1 consists of 3 neighbouring tracks, which have to be radiometrically adjusted to use the same parameter set for the whole test site. To assure a constant radiometry we correct the near far range illumination loss for each track (by scaling the grey values with the mean grey value of each column and rescale it to 256 grey values). In order to reduce speckle, we use the multi look, geocoded X- and L-band data.

Same adaptations of the TUM road extraction strategy for optical imagery towards SAR imagery were necessary. As roads in SAR imagery appear as dark lines, it is possible to facilitate their extraction by using directly the intensity of each pixel, instead of only local contrast information. We established this for the time being by a threshold. So, a restriction to the relevant areas can be achieved. The advantage of this restriction is that the parameter settings for the line extraction can be soften and nevertheless, fewer false alarms are extracted.

In the test scenery exist many residential areas and forests. In these areas, the proposed road model does not fit and many false alarms are to be expected. Since the computation time increases with the number of potential road segments we exclude the regions of no interest from the extraction. A mask containing cities and forests is generated by X- and fully polarimetric L-band data based on the intensity values, ratios, and neighbourhoods using the eCognition software of Definiens (User Guide eCognition). Both masks, the threshold and the city/forest mask, are united and introduced into the line extraction.

For subsequent line extraction the imagery was tiled to control the computational effort.

Line extraction: Line extraction is performed in multiple images of different radiometric and/or geometric resolutions separately using the approach described in (Steger, 1998) which is based on differential geometry. A few, semantically meaningful parameters have to be chosen: The maximum width of the lines to be extracted and two threshold values according to the local radiometric contrast between lines and their surroundings. The result of the line extraction is a set of pixel and junction points for each image in sub-pixel precision. The extraction is not complete and contains false alarms, i.e., some roads are not extracted and some extracted lines are not roads.

Potential road segments: In the next step, the lines are evaluated, proportional to their fitting to a regional model of roads. This incorporates the assumption that roads mostly are composed of long and straight segments having constant width and reflectance. Linear fuzzy functions are used to transform these properties into fuzzy values. An overall fuzzy value for each line is derived by aggregation of the specific values.

Fusion of different image channels: After evaluating the road segments, more global characteristics of roads are considered in terms of the functionality and topology of roads. According to the nature of roads, they form a network wherein all road segments are topologically linked to each other. Thus, the different sub-parts of the test site have to be fused to connect them to each other. Also, the potential road segments of all channels are fused.

Construction of a weighted graph: A weighted graph is constructed from the potential road segments of all channels. Costs for each potential road segment are calculated by dividing the length of the road segment by its overall fuzzy value. These costs are assigned to the respective edges of the graph. The weighted graph contains gaps because, in general, not all roads were detected by the line extraction. Therefore, each gap is evaluated based on the collinearity, absolute and relative gap length (compared to the adjacent lines).

Selection of seed segments: For the network generation various seed points have to be selected. The segments with relatively high weights are selected. Therefore, the extraction depends strongly on the parameters from the fuzzy values of the evaluation of the road segments.

Calculation of shortest path: Each pair of seed points is connected by calculating the optimal path through the graph using the Dijkstra algorithm. The main disadvantage of that procedure is that if there are two gaps longer than the maximum gap length, e.g. caused by low contrast, the part between these gaps cannot be added to the road network because no connection with the seed points can be established through the graph.

3. EVALUATION STRATEGY

The evaluation of the automatically obtained results is done by a comparison to reference data. Here, vector data of a topographic map and manually plotted road axes are used as reference data. A brief description of the evaluation procedure is given below. More details can be found in (Wiedemann, et. al., 1998).

The comparison is carried out by matching the extracted data to the reference data using the so-called "buffer method", in which every proportion of the network within a given distance (buffer width) from the other is considered as matched. Two questions are thought to be answered by means to the defined quality measures: (1) How complete is the extracted road network, and (2) How correct is the extracted road network. The completeness indicates the percentage of the actually present road network, which could be extracted, whereas the correctness is related to the probability of an extracted linear piece to be indeed a road.

Completeness is defined as the percentage of the reference data, which lies within the buffer around the extracted data:

$$comleteness = \frac{length\ of\ matched\ reference}{length\ of\ reference} \qquad (1)$$

Correctness represents the percentage of the correctly extracted road data, i.e., the percentage of the extracted data lying within the buffer around the reference network:

$$correctness = \frac{length\ of\ matched\ extraction}{length\ of\ extraction} \qquad (2)$$

In addition, also the geometric accuracy of the correct extraction is assessed. It is expressed as RMS difference between the extracted and the reference data.

4. RESULTS AND DISCUSSION

4.1 Test site and data description

The test site covers an area of approximately 110 km^2. Because the automatic road extraction works well, only in open areas, the test site consists of two rural sceneries in Germany: test site 1, Ehingen, south Germany, covers about 7.5 km × 11 km and test site 2, Erfurt, east Germany, represents land coverage of about 40 km^2. The test data set consists of airborne X-band and fully polarimetric L-band data taken by the experimental SAR (E-SAR) of the German Aerospace Center (DLR) within the Prosmart II project. The ground resolution is about 2 m, respectively 3 m with pixel spacing of 1 m. The missions took place in September 2000 and March 2001. Each test site is composed of three tracks.

The SAR imagery contains highways, main roads, and different types of secondary roads. The distinction between different secondary roads is more formal, so they were grouped to one class of secondary roads. Figure 2 shows the composed, geocoded, multi look SAR image (X-band, HH-polarization) of test site 1 with the arranged topographic map reference data. The main roads comprise two wide lanes. They are displayed with a thick white line. The thin white lines are two-lane secondary roads. Highways are only contained in test site 2, they consist of four or more lanes separated by a crash barrier.

For the evaluation of the road extraction results we take into account on the one hand all roads which are really existent in the scenery and on the other hand only those roads which are visible in the SAR imagery, i.e., two classes of reference data are used: a digital topographic map and a reference extracted manually from the SAR imagery. The vector data of the topographic map were obtained from aerial photography in scale of 1:10,000 taken in 1995 and verified in a field completion in 1998. The geometric accuracy is about 3 m. The manually extracted reference was plotted by an operator. The classification of the roads to major and minor ones was decided in case of doubt by comparison to the topographic map data. Since the automatic road extraction only works well in open areas city streets and forest roads are excluded from both references like in the extraction.

4.2 Road extraction results

The tests are carried out on X-HH, X-VV, and L-VV band. Extractions are made on single- and multi-channel data, see Table 1. The first result was generated using X-band data with horizontal polarization (X-HH). The second result was achieved based on the X-HH data, as well, but in this case in addition to wider lines, also narrow lines were extracted during the line extraction step. The third extraction result was achieved with the aid of a multi-frequency and multi-channel data set by extracting from X-HH, X-VV, and L-VV. By visual inspection, we found out that roads are most clearly visible in the L-VV channel, compared to all the other polarimetric L-band channels.

	number of channels	bands	line width
Test 1	1	X-HH	wide
Test 2	2	X-HH X-HH	wide narrow
Test 3	3	X-HH X-VV L-VV	wide wide wide

Table 1. Test overview

The three tests have been carried out on each of the two test sites. Two smaller subsets of test site 1 are chosen to optimize the parameter settings. Test site 2 was processed, using the same set of parameters as for test site 1. The results were evaluated using the topographic map data as well as the manually plotted reference data. With regard to the main roads the manually plotted reference corresponds mostly with the topographic map data. In contrast, secondary roads could not totally be gathered during the manual extraction of the reference data. The evaluation was carried out for each category of roads separately. The resulting quality measures are summarized in Table 2- 4.

First analyzing the overall results of the two test sites it is important to note that the extraction with two different line widths yields by far the best results regarding the completeness and the correctness. Here, especially the secondary roads are relatively complete, which is a consequence of the additional extraction of narrow lines in test 2. But nevertheless, automatic road extraction often fails to extract secondary roads. This is, due to the course visibility of secondary roads in comparison to main roads.

	Manual extraction		
	Test 1	**Test 2**	**Test 3**
Completeness	**68.2 %**	**76.4 %**	**61.1 %**
main roads	92.7 %	89.9 %	83.8 %
secondary roads	59.6 %	71.6 %	53.2 %
Correctness	**55.2 %**	**56.4 %**	**46.2 %**
RMS	**2.1 m**	**2.0 m**	**2.2 m**
main roads	2.2 m	2.2 m	2.2 m%
secondary roads	2.1 m	1.9 m	2.2 m

Table 2. Comparison of extraction results to manual extraction of test site 1

	Topographic map		
	Test 1	**Test 2**	**Test 3**
Completeness	**55.4 %**	**65.8 %**	**50.9 %**
main roads	90.9 %	88.1 %	82.5 %
secondary roads	45.5 %	59.6 %	42.1 %
Correctness	**53.8 %**	**57.9 %**	**46.0 %**
RMS	**3.0 m**	**2.9 m**	**3.0 m**
main roads	3.5 m	3.4 m	3.8 m
secondary roads	2.7 m	2.6 m	2.7 m

Table 3. Comparison of extraction results to topographic map data of test site 1

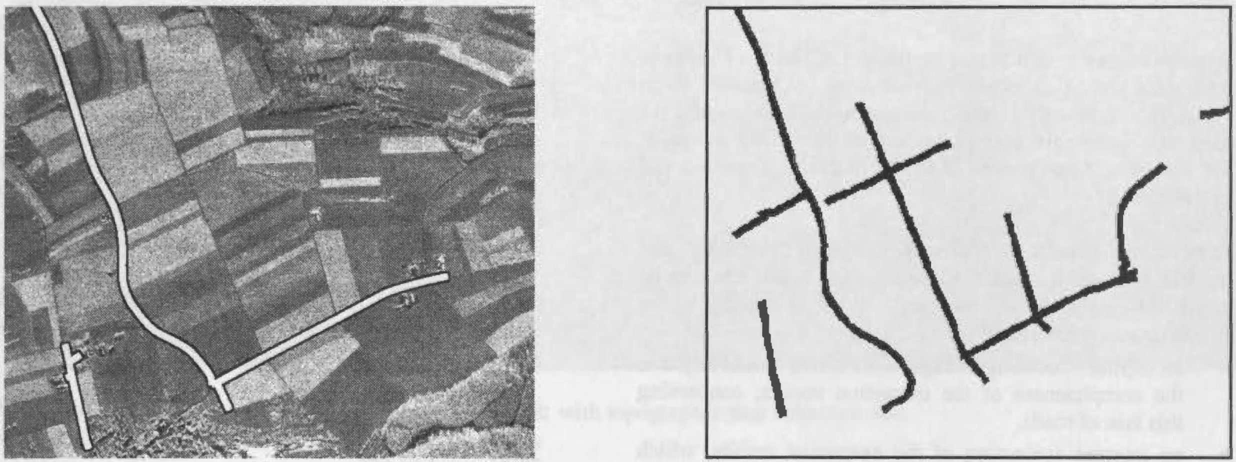

(a) (b)

Figure 2. Extracted ways (b), which are not contained in the reference road network (a)

The extraction results for test site 1 are displayed in Figure 3. The correctness of all extraction results tells that only half of the extracted road segments are correct. Most of the false alarms are other dark linear structures like shadows of the borders of forests and hedges between fields' structures or, in some cases, ways.

By incorporating ways into the reference data the correctness increases in case of test 2 from 56,9 % to 79,0 %. In Figure 4 a scene is shown where ways are detected, which do not belong to the road network. This phenomen presents the absence of a classification of automatic extracted results. Ways could be excluded from the extraction result by integrating a distinction between, e.g., paved and unpaved ways.

The proportion of found main roads is relatively high, contrary to highways. Though highways in the test site 2 are clearly visible, the result of the automatic road extraction is very incomplete for this road category. This is because especially crash barriers, traffic signs, bridges, and low contrast impede the extraction. To improve the results, highways have to be explicitly modeled, which is not the case in the current implementation of the road extraction. A first step could be the modeling of the crude crash barrier and its surrounding.

	Manual extraction		
	Test 1	**Test 2**	**Test 3**
Completeness	**45.6 %**	**59.4 %**	**50.5 %**
highways	35.4 %	68.3 %	52. 9 %
main roads	88.8 %	94.1 %	92.5 %
secondary roads	39.5 %	46.1 %	38.5 %
Correctness	**42.8 %**	**56.9 %**	**42.7 %**
RMS	**2.1 m**	**1.9 m**	**2.2 m**
highways	2.5 m	1.7 m	2.3 m
main roads	2.0 m	2.0 m	2.1 m
secondary roads	2.0 m	2.0 m	2.1 m

Table 4. Comparison of extraction results to manual extraction of test site 2

5. CONCLUSION

Automatic road extraction was performed on two large test sites. The extraction results were evaluated based on a comparison with reference data. Quality measures for the completeness, the correctness and the geometrical accuracy were calculated. For this purpose, the reference data were separated into three classes: highways, main roads, and secondary roads, analogous to the German topographic map standard in scale of 1:25,000. The completeness and the geometric accuracy were calculated for each of these classes separately. All these comparisons were carried out using two kinds of reference data: vector data of a digital topographic map and road axes extracted by a human operator. Thus, we take into account which roads are really existent in the scenery and which an operator is able to detect from the imagery. The comparison of the extraction results achieved with three different test set-ups, shows that it is useful to adapt the parameters of the line extraction relatively exact to the roads to be extracted. Nevertheless, secondary roads could not totally be extracted automatically. But also some of the secondary roads are missing in the manually extracted reference, because they are not visible in the SAR imagery.

A promising approach would be the use of fully polarimetric SAR data since they could provide more information about the surface than only single-polarization data. Especially for modeling the single bounce surface characteristic of roads, the odd bounce component of the Pauli decomposition could be valuable.

In total the results are related to width, visibility, and modeling of each class. Presently, for main roads quite satisfying results can be achieved. What is missing in the current implementation of the road extraction is

- an explicit modeling of highways, which would improve the completeness of the extraction results, concerning this lass of roads,
- an internal evaluation of the extraction results, which would lead to more correct results, and
- a classification of the extraction results into different road classes, such that, e.g., unpaved ways could be excluded from the extraction results.

ACKNOWLEDGEMENTS

The authors thank Infoterra GmbH, Friedrichshafen, Germany and the German Aerospace Center (DLR) for providing the SAR data, Definiens AG, Munich, Germany for generating the mask for cities and forests, and the LVA Baden-Württemberg, Germany, to make available the topographic map data.

REFERENCES

Bazohar, M. and Cooper, D., 1996. Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. In: *IEEE Trans. Pattern Anal. Machine Intell.*, 18(7), pp. 707-721.

Chanussot, J. and Lambert, P., 1998. An application of mathematical morphology to road network extraction on SAR images. In: *Proc. Int. Symp. Mathematical Morphology*, Amsterdam, The Netherlands, pp. 399-406.

Fischler, A., Tenenbaum, J., and Wolf, H., 1981. Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. Computer Graphics and Image Processing, 15, pp. 201-223.

Infoterra, 2002. Prosmart II project http://www.infoterra-global.com/prosmart2.html (accessed 15 Mar. 2002)

Jeon, B., Jang, J., and Hong, K., 2002. Road detection in spaceborne SAR images using a genetic algorithm. *IEEE Trans. Geosci. Remote Sensing*, 40(1), pp. 22-29.

Kartartzis, A., Sahli, H., Pizurica, v., and Cornelis, J., 2001. A model-based approach to the automatic extraction of linear features form airborne images. *IEEE Trans. Geosci. Remote Sensing*, 39(9), pp. 2073-2079.

McKeown, D. M. and Denlinger, J. L., 1988. Cooperative methods for road tracking in aerial imagery. In: *Proc. IEEE Comput. Vision Pattern Recognit.*, Ann Arbor, MI, pp. 662-672.

Samadani, R. and Vesecky, J. F., 1990. Finding curvilinear features in speckled images. In: *IEEE Trans. Geosci. Remote Sensing*, 28(4), pp. 669-673.

Steger, C., 1998. An unbiased detector of curvilinear structures. In: *IEEE Trans. Pattern Anal. Machine Intell.*, 20(2), pp. 549-556.

Tupin, F., Maître, H., Mangin, J. F., Nicolas, J. M., and Pechersky, E., 1998. Detection of linear features in SAR images: Application to road network extraction. In: *IEEE Trans. Geosci. Remote Sensing*, 36(2), pp. 434-453.

User Guide eCognition, http://www.definiens.com (accessed 15 Mar. 2002)

Wiedemann, C. and Ebner, H., 2000. Automatic completion and evaluation of road networks. In: *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, The Netherlands, Vol XXXIII, Part B3, pp. 979-986.

Wiedemann, C. and Hinz, S., 1999. Automatic extraction and evaluation of road networks from satellite imagery. In: *International Archives of Photogrammetry and Remote Sensing*, Munich, Germany, Vol XXXII, Part 3-2W5, pp. 95-100.

Wiedemann, C., Heipke, C., Mayer, H., and Jamet, O., 1998. Empirical evaluation of automatically extracted road axes. In: K. J. Bowyer and P.J. Philips (eds), *Empirical Evaluation Methods in Computer Vision*, IEEE Computer Society Press, Los Alamitos, California, pp. 172-187.

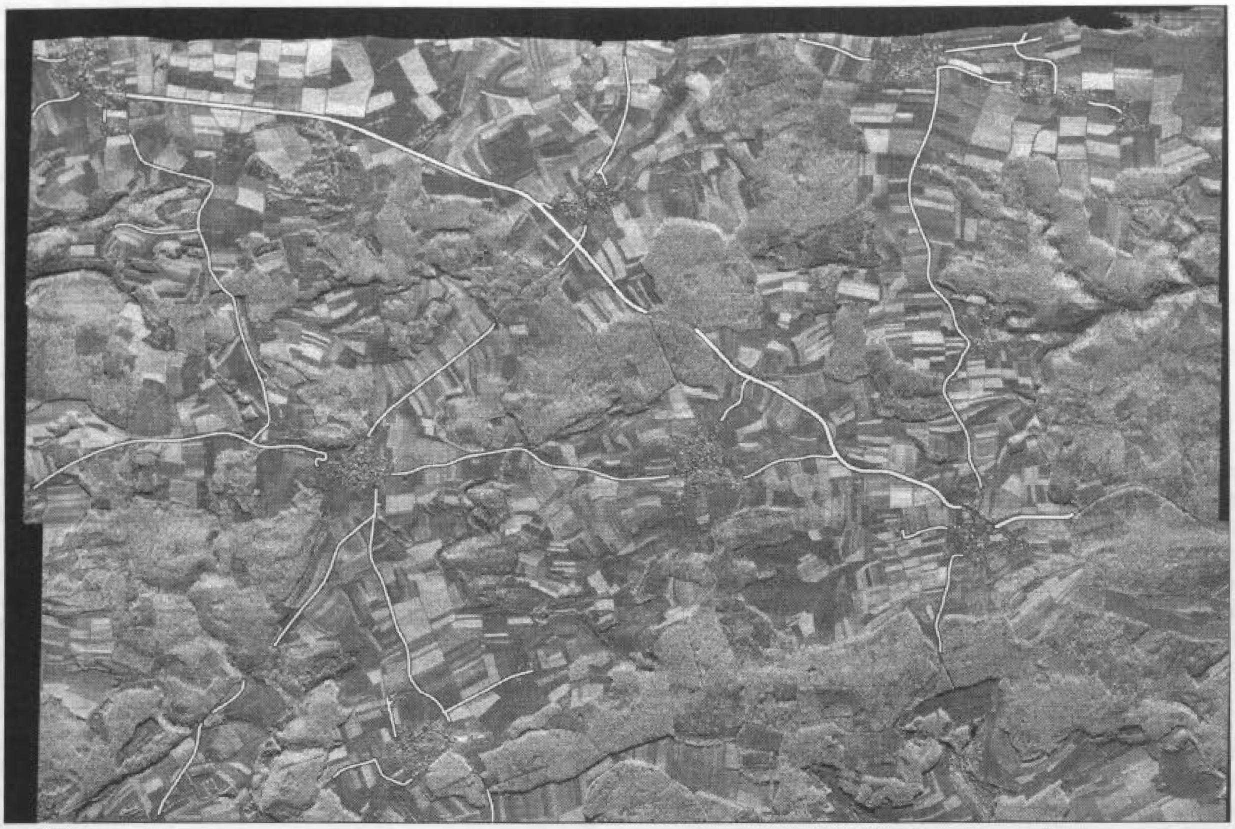


Figure 2. Test site 1 with topographic map reference data

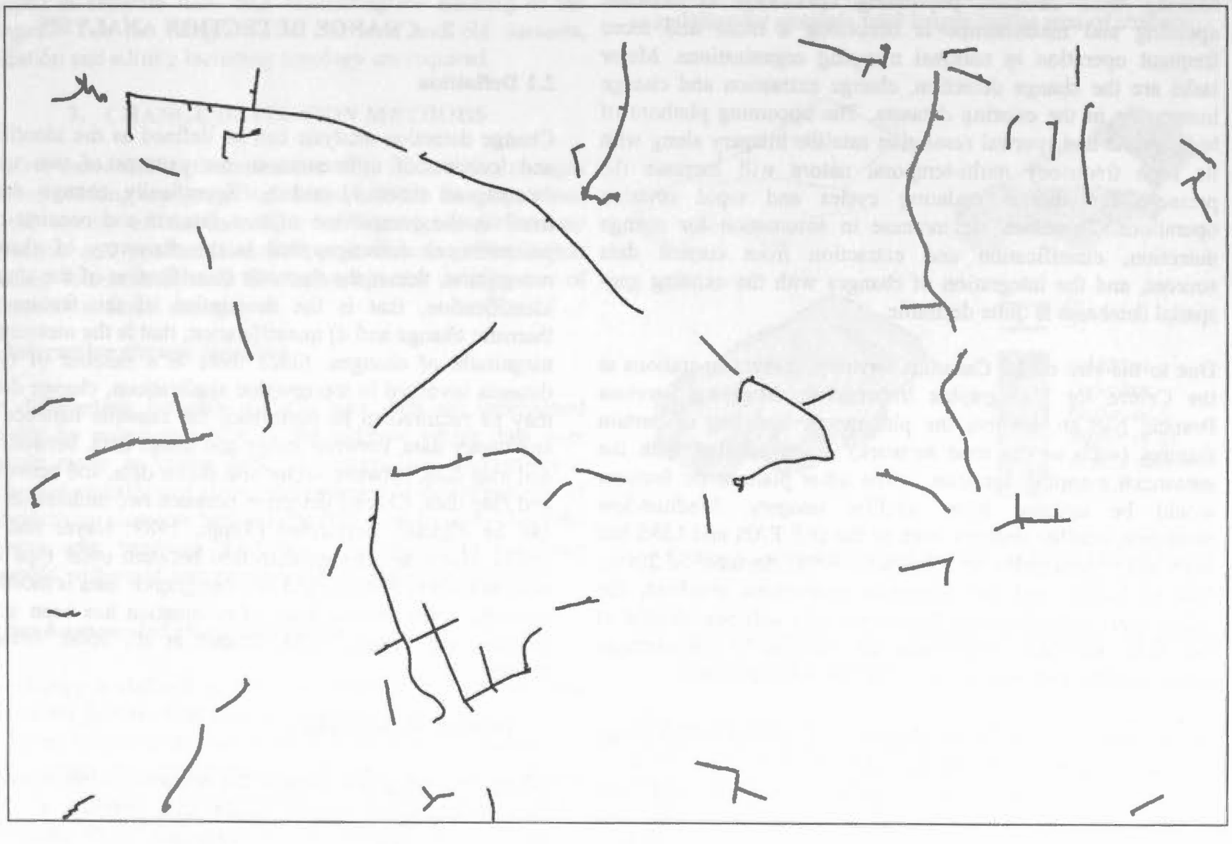


Figure 3. Road extraction result of test 2 (extraction with two line widths)

CHANGE DETECTION METHODS FOR THE REVISION OF TOPOGRAPHIC DATABASES

C. Armenakis*, I. Cyr , E. Papanikolaou

Center for Topographic Information, Geomatics Canada, Natural Resources Canada,
615 Booth Str., Ottawa, Ontario, Canada K1A 0E9
(armenaki, iscyr)@NRCan.gc.ca ; lpapanikoloaou@yahoo.com

Commission IV, IC WG II/IV

KEY WORDS: Change detection, feature extraction, revision, Landsat 7 ETM+, raster maps

ABSTRACT:

Database updating and maintenance is becoming a demanding operation in national mapping organisations. Major tasks are the change detection and extraction and its integration in the existing data sets. Change detection is performed between a number of data sets involved in topographic applications. The implementation of more automation coupled with the availability of heterogeneous data requires the investigation, adaptation and evaluation of new approaches and techniques. The demand for rapid mapping operations is continuously increasing. Prior to change detection, extraction of homogeneous features from both the multi-temporal datasets is necessary. Thresholding and texture measures were used to evaluate the potential of rapid extraction of topographic elements from scanned monochrome maps, while the extraction of features from satellite imagery involved initially image and theme enhancement by applying various image fusion and spectral transformations, followed by image classification and thresholding. Interactive and semi-automated change detection methods are proposed and implemented based on non-intersection of old and new features and the generation of buffer zones. These were used for the determination of planimetric changes. The various approaches and methodology developed and implemented along with examples, initial results and limitations are presented and discussed. The tests showed that the approaches were more or less feature/theme dependent, while at the same time they can augment and significantly enhance the conventional topographic methods.

1. INTRODUCTION

Moving from database populating operations to database updating and maintenance is becoming a more and more frequent operation in national mapping organisations. Major tasks are the change detection, change extraction and change integration in the existing datasets. The upcoming plethora of high spatial and spectral resolution satellite imagery along with its high frequency multi-temporal nature will increase the pressure for shorter updating cycles and rapid revision operations. Therefore the increase in automation for change detection, classification and extraction from current data sources, and the integration of changes with the existing geo-spatial databases is quite desirable.

Due to the size of the Canadian territory, current operations at the Centre for Topographic Information, Mapping Services Branch, NRCan, involve the planimetric updating of certain features (such as the road network) in cooperation with the provincial mapping agencies, while other planimetric features would be updated from satellite imagery. Medium/low resolution satellite imagery such as the IRS PAN and LISS has been used (Armenakis and Savopol, 1998; Armenakis, 2000). Due to higher cost and copyright restrictions involved, the image layer of the national framework data will now consist of the 15m and 30m resolutions of Landsat 7 ortho-images (panchromatic and multispectral) for the entire country.

In this paper we will address the change detection methodology and give examples involving different data types such as scanned maps, vectors from existing database, and new imagery. Detected changes will be used to update the national topographic vector database. The various approaches, methodology including the potential of digital image processing techniques and automated procedures as well as the initial results obtained will be presented and discussed.

2. CHANGE DETECTION ANALYSIS

2.1 Definition

Change detection analysis can be defined as the identification and location of differences in the patterns of two temporal datasets at times t_1 and t_2. Specifically change detection involves the comparison of two datasets and consists of four processes, a) detection, that is the discovery of change, b) recognition, that is the thematic classification of the change, c) identification, that is the description of the feature of the thematic change and d) quantification, that is the measure of the magnitude of changes. Since there is a number of types of datasets involved in topographic applications, change detection may be required to be performed for example between image and vector data, between image and image data, between image and map data, between vector and vector data, and between map and map data. Change detection between two multidated images can be digitally performed (Singh, 1989; Hayes and Sader, 2001). However, change detection between other type of data such as current imagery and old topographic data is mostly done visually, while certain level of automation has been achieved (Heipke and Straub, 1999; Ohlhof et al.; 2000; Armenakis, 2000).

2.2 Domain of comparison

When comparing two datasets the domain of comparison has to be defined. For example, vector data provides a classified abstract representation of the landscape, while imagery is an unclassified continuous but resolution-dependent generalized representation of the landscape. If comparing data of the same nature (i.e., two homogeneous satellite images) the change detection can be determined at the *data level* domain (comparison of pre-processed data), while when comparing heterogeneous datasets the change detection is performed at the *decision level* domain (comparison of analysed data), as some data processing (e.g., image classification, vectorization of detected features) needs to be done prior to change detection. For example, the latter level is the one used in this study for the analysis and comparison of topographic data and Landsat 7 ETM+ orthoimage due to their heterogeneity. For the generation of homogeneous features the analysis involves the decomposition of the data to basic feature elements (linear or polygonal) using several image extraction techniques based on the various topographic themes,

2.3 Type of changes

Changes of features occur, when either the spatial and/or the attribute characteristics of an existing feature have changed. A change can therefore be in the form of:

a) *Spatial modification*, when the changes in the geometry of an existing feature and its corresponding new feature exceed the accuracy buffer limits and/or when the topology changes; and

b) *Attribute modification*, when the thematic values of an existing feature change.

The changes in the data patterns are then determined based on the following definitions:

Confirmation of the existing data, when neither their spatial nor the attribute elements of an existing feature have changed.

Addition of features, when a new feature is added or an existing feature is modified.

Deletion of feature when a feature or part of it is eliminated.

The information about deletions and additions is kept and time stamped in separate files, thus facilitating the tracking of the changes. For the integration of the new and old datasets, validation and editing including topology are required.

3. CHANGE DETECTION METHODS

We have tested and implemented two ways of detecting changes between existing topographic data and satellite imagery, based on the results of the extraction of homogeneous features: interactive and semi-automated change detection. Planimetric accuracy tolerance is used for the acceptance or rejection of differences as actual changes.

3.1 Interactive change detection

The differentials between old and new data are determined visually by superimposition of the old/new vector data over new/old raster or vector data. The interactive comparison is performed visually and allows the operator to detect patterns of non-matching elements between the two datasets and collect the changes with heads-up digitizing. The procedure has been enhanced with rapid feature extraction to assist the operator.

3.2 Semi-automated change detection

The change is defined as the non-intersection of the old and new vector features between two temporal spatial states (S_1 and S_2). If the output of the non-intersection contains data from S_1, then changes are considered as deletions. If it contains data from S_2, changes are considered as additions. The changes consist of additions and deletions. A deletion is the difference between the old state and the common elements between the two states, while an addition is the difference between the new state and the common elements between the two states. The total changes and the additions and deletions are then expressed as follows:

$$C_{ha} = A_{dd} + D_{el} \tag{1}$$

$$A_{dd} = S_2 - S_1 \cap S_2 \tag{2}$$

$$D_{el} = S_1 - S_1 \cap S_2 \tag{3}$$

where, C_{ha} = total spatial change
A_{dd} = additions
D_{el} = deletions
S_1 = spatial state at time 1
S_2 = spatial state at time 2

A detected change could be caused by differences in positional accuracies between the two datasets. The significance of change can be expressed based on accuracy tolerances and minimum sizes. To account for accuracy tolerances, appropriate spatial buffers are generated around both features during the change detection operation, while the minimum sizes satisfying the specifications are handled using appropriate spatial filters. The buffering and filtering operations are used to keep only the significant changes.

Whatever vector segments are outside the buffer zones are considered as changes: a) if the new features from the S_2 data are outside the buffer applied to S_1 features, changes are considered as the actual additions; and b) if the old features from the S_1 data are outside the buffer of the S_2 features, changes are considered as the actual deletions (Fig. 1). This approach allows for quantification of changes per theme by calculating for example total length and/or area of change.

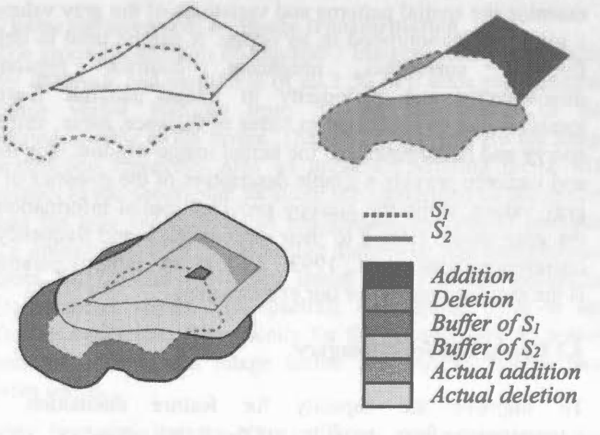


Figure 1: Actual change detection: addition and deletion.

4. GENERATION OF HOMOGENEOUS FEATURES FOR COMPARISON

Due to the heterogeneity of the multi-temporal data types, the change detection in our tests is performed at the decision level. As the existing data is in vector form, initially homogeneous vector features are extracted from the two datasets prior to change detection.

4.1 Extraction from scanned maps

The extraction process from scanned topographic maps involves the distinction of certain map features belonging to one theme (e.g., water bodies) from other map features and from the map background followed by the extraction of the exact geometrical shape of the feature. Automatic extraction of features from scanned topographic has been studied as a viable alternative to manual digitization and to reduce manual editing from raster to vector operations (Nebiker and Carosio, 1995; Carosio and Stengele, 1993). Fully automated vectorization of the raster map at this phase would be rather difficult as it would require the development and implementation of pattern recognition techniques for symbol identification and extraction of various cartographically represented elements, which presently cannot automatically be distinguished without human intervention. Semi-automated approach for feature extraction is the computer-assisted vectorization (e.g., on-screen line follower) of the rasterized map elements. Fast vectorization of map elements can also be done by thresholding and texture analysis followed by raster-to-vector conversion without any topology.

Considering the raster map as a grayscale image, it is possible to apply the thresholding technique using the map histogram. This method uses the fact that the map lines, which delineate or represent most of the cartographic information (including symbols and toponymy), appear much darker than contours and grid lines. Therefore, the thresholding permits the distinction, in a raster format, of relevant topographic information such as the lakes, rivers, wetlands, wooded areas, eskers, roads, etc., from contours and grid lines. The most appropriate threshold value (value delineating these two general classes) has to be determined by the operator since this value may vary according to the printing and scanning specifics.

Texture measures (Haralick and Shapiro, 1992) are used to examine the spatial patterns and variations of the gray values in a pixel's neighbourhood in an image. It can be used to define fineness, coarseness, roughness, contrast, regularity, directionality and periodicity in image patterns. Texture measures can be expressed in terms of variance, mean, entropy, energy and homogeneity of the kernel image window. The mean and variance provide a simple description of the statistics of the gray values, while the entropy provides spatial information of the gray values related to their directionality and frequency of occurrence (Sonka et al., 1993). As such, the entropy parameter is the most promising for our experiments.

4.2 Extraction from imagery

To improve the capacity for feature distinction and interpretation from satellite multi-spectral imagery, various image enhancement techniques (radiometric, spatial and spectral) are performed prior to any feature extraction. Specifically for feature recognition, image fusion (e.g., IHS, PCA, IMG-FUSE), and image transformations (e.g., Tasseled Cap, NDVI) are applied. Image fusion allows for meeting the geometric requirements due to the use of the higher resolution band, while it increases the interpretability of the features due to lower resolution spectral bands. Image transformations can enhance specific characteristics of the surface cover (e.g., vegetation variations, wetness, brightness).

Feature extraction is then performed either by image classification (unsupervised, supervised) or by band-based thresholding. Spatial filtering is applied to satisfy the minimum size specifications for the extracted areal features. An automatic vectorization of the extracted polygonal boundaries follows.

5. CHANGE DETECTION APPLICATIONS

The decisions concerning the domain of comparison of the datasets and the selection of the most appropriate methods in data processing, data extraction and change detection procedures depending on a number of factors. Factors such as the type and time of acquisition of the two datasets, the topographic themes to analyse and the applicable level of automation, are influencing the decisions taken for the whole process of change detection. The following two examples of change detection demonstrate the required flexibility of the process for data extraction and change detection as well as the applicability of the proposed methodology.

5.1 Case 1: Interactive change detection for water bodies between scanned maps and Landsat 7 imagery

Datasets

Both S_1 and S_2 datasets cover a taiga zone located in the North West Territories, Canada, where S_1 (data at time t_1) is monochrome paper maps (1:50 000 scale) edited in 1979 and scanned at 600 dpi resolution; and S_2 (data at time t_2) is Landsat-7 ETM+ image acquired on June 14, 2000 and consists of 7 bands (one panchromatic at 15m-resolution and six multispectral at 30m-resolution).

Vector extraction from scanned map

The direct superimposition of the two raster data (map and imagery) is not the easiest way for detecting interactively cartographic changes. To facilitate this visual process, vectors were extracted automatically from the map representing certain topographic features. Two methods are used to extract the different vector features from the scanned map: thresholding and the entropy measure of texture (Armenakis et al., 2001).

The thresholding method is used to extract most of the cartographic information (including symbols and toponymy) in a single layer (no distinction in the extracted features). The most appropriate threshold value is determined using the histogram by evaluating the natural spectral break between two general classes: (1) feature boundary lines including toponymy and (2) contours lines including the map grid. The results of the map thresholding are then filtered using median and sieve filters to remove noise and all polygons that are smaller than a given minimum size, measured in pixels area. The level of filtering must be chosen adequately to both keep small or isolated feature map lines and remove enough grid lines and contours that may reduce the feature visibility. Vector features are obtained through an R=>V conversion (interactive or automatic) of the thresholded results. From this automatic fast vectorization, vectors of raster elements that represent the old information or our prior knowledge of the topographic features are obtained. This output vectors could be used either in an interactive change detection process with the imagery as the background layer or in a feature extraction process from the imagery by selecting appropriate training sites prior to the image classification. There is a limited use of these output vectors because of the fact that they are not topologically structured and include map elements such as symbols and text. An example of these vectors superimposed with a Landsat 7 band is presented in Figure 2a.

Considering that water bodies are represented differently in the 1:50 000 scale monochrome maps (polygonal screened gray patterns) than the rest of the map (linear patterns), a method using the entropy measure has been implemented to automate the extraction of lakes from the scanned maps. In the entropy calculation, a distance of 1 was used between two neighbouring pixels and a directional invariant outcome. The degree of detection of small objects depends on the kernel size of co-occurrence matrix of the entropy measure (11x11 pixels kernel was used). Larger size kernel allows for more data to compute a significant entropy measure but reduces the ability to detect small objects. To extract automatically the lakes boundaries, the entropy filter was applied to the raster map, followed by thresholding of the resulting image between lakes and no lakes, then by a median filter to eliminate noise, then a filtering to remove lakes that do not respect the minimum size requirement, and finally by vectorizing the resulted binary image. This automated vector extraction is limited to this specific feature. An example of these vectors superimposed with a Landsat 7 band is also presented in Figure 2b.

Vector extraction from imagery

Considering the latitude (around 64°N, taiga zone in Northern Canada) and time of the year in the Landsat-7 acquisition, most of the large lakes (predominant cartographic feature) are still frozen. Within the smaller lakes, most of the covering ice was melted. The spectral difference between frozen and non-frozen lakes with spectral bands ETM1 to ETM4 and even with ETM8 (panchromatic) was clearly visible. Spectral bands ETM5 and ETM7 did not show this distinction. Since we are not interested in this spectral difference, the latter bands were selected for the extraction of waterbodies from Landsat 7. A simple region thresholding of band ETM5, considering values (gray levels) from 0 to 25, allowed for the extraction of the lakes (Figure 2c). In all the fusion experiments, few feature enhancements were obtained because of the nature of each spectral band and the time of acquisition of the image (frozen water, wet soils, few vegetation, etc.). In this particular case, only lakes could be extracted by a simple thresholding (no fusion and supervised classification were necessary).

Interactive change detection

After extracting vectors from the old 1:50 000 scale monochrome map and choosing the most appropriate Landsat 7 image band(s) as background, the change detection process between these two datasets could be done quickly by the interactive superimposition method. Figure 2 gives a visual comparison between lakes vectors from map thresholding (2a) or texture analysis (2b) and from band ETM5 of Landsat-7 imagery (2c). By using the extracted lake vectors from the imagery, the interactive change detection could also be done with the raster map as background (Fig. 2c).

Regarding the type of changes based on the contextual information from the Landsat-7 imagery, confirmations and geometric modifications (additions or deletions) could be done with higher confidence than complete feature additions or deletions (that imply a certain level of feature identification that is difficult to get from Landsat-7 only). These two types of changes have to be determined with caution in Landat-7 image due to its resolution. For example, a map feature will not have necessarily to be removed or added if it is not visible and properly identifiable in the image.

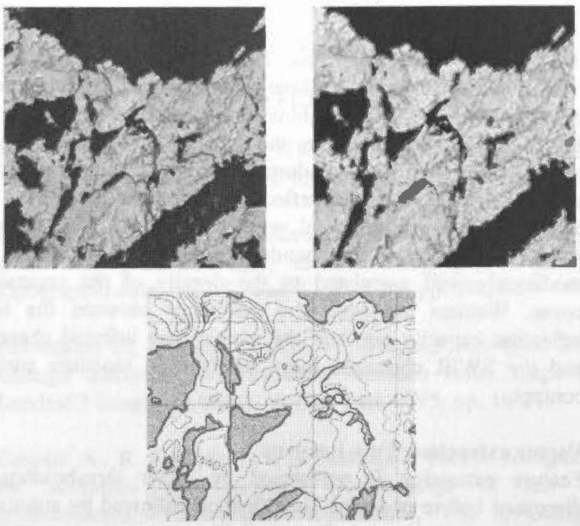


Figure 2 – Interactive change detection with Landsat-7 band ETM5 as background a) map lake vectors obtained by thresholding; b) map lake vectors obtained by texture analysis; c) lake vectors from Landsat-7 band ETM5 with the raster map as background.

5.2 Case 2: Semi-automated change detection for vegetation, wetlands and water bodies

Datasets

Both S_1 and S_2 cover a zone located in mixed boreal forest of Northern Alberta, Canada, where S_1 (data at time t_1): Vectors obtained from the National Topographic Database (NTDB, 1:50 000 scale) dated 1979 with 30m of planimetric accuracy. The data is ready for use as they are in vector form, structured and organized by themes; and S_2 (data at time t_2): Landsat-7 ETM+ image acquired on May 29, 2000 consisting of 7 bands as before.

Theme enhancement by image transformation

The ground resolution of Landsat-7 imagery limits its capacity for detection and extraction of all 1:50 000 scale map features. Some vector features from S_1 are not visible or not distinguishable in the image. The time of the year for the image acquisition represents also an important limiting factor with respect to the geographic location, the topography of the terrain, and the type and number of topographic features to consider. To improve the capacity of feature distinction and interpretation from the Landsat-7, various image enhancement techniques (radiometric, spatial and spectral) are applied prior to any feature extraction. Specifically for the enhancement of feature recognition various image fusion and transformations have been tested.

To fuse Landsat-7 ETM+ bands (panchromatic and multispectral), the IMGFUSE routine of PCI package that is based on cross-correlation between high and low-resolution bands (computed band by band) was used. The advantage of this fusion method is that it preserves the radiometric properties of the original multispectral channels (Cheng et al., 2000), while common fusion methods such as IHS and PCA usually distort them (Chavez et al., 1991).

As the vegetation cover seems to be an important factor in modulating the information content of the various Landsat-7 bands, a spectral band transformation expressly designed for the enhancement of the vegetation cover density and condition, called the "Tasselled Cap" (Mather, 1987), was applied. The six multispectral bands of Landsat-7, except the thermal ones, are used in order to compute three parameters called brightness, greenness and wetness. Brightness is a weighted sum of all six ETM channels and expresses the total reflection capacity of a surface cover. Small areas dominated by dispersed vegetation appear brighter (high total reflection). Greenness expresses the difference between the total reflectance in the near infrared bands and in the visible bands and has been shown to be moderately well correlated to the density of the vegetation cover. Wetness expresses the difference between the total reflection capacity between the visible-near infrared channels and the SWIR channels, more sensitive to moisture surface content.

Vector extraction from imagery
Feature extraction is performed by either thresholding as discussed before or image classification followed by automatic vectorization if required. For certain features such as vegetation and wetlands better results obtained by classification. The three bands of the fused image (PCI-IMGFUSE routine) corresponding to ETM3, ETM4 and ETM5, the three Tasseled Cap parameters (brightness, greenness and wetness) and the NDVI result were used as input data. Noting that field ground truth and verification are not available, the existing topographic vectors were used as 'prior knowledge' to provide cues and guidance in the classification process from imagery. This knowledge is used as contextual information for selecting training sites for pixel-based classification. The training sites were selected considering our prior knowledge of the topographic features: vectors from 1:50 000 scale scanned map and from 1981 fire polygons of the Alberta Sustainable Resource Development, Forest Protection. For features like vegetation and wetlands, the selection of training sites is somehow difficult due to the variation in spectral signatures. These variations could reflect the presence of sub-classes in the defined topographic feature (e.g., cleared cuts, burned areas and regenerated forests are included in wooded areas, while all types and conditions of swamps/marshes are included in wetlands). To take into account these spectral variations even though there are not reflected in the map content, different training sites have to be selected to create several sub-classes for the same feature. After the classification process (by maximum likelihood) the sub-classes were aggregated to adapt the detected classes to general categories of map elements (vegetation, wetlands, water bodies). In this case, six sub-classes were extracted (water bodies, marshes, swamps, recent cuts, burned areas, wooded areas) and aggregated to obtain three final classes (water bodies, wetlands and vegetation). Based on our training sites, the Kappa coefficient was 0.95. In comparison, without image fusion, the classification leads to Kappa coefficient of 0.87. Following supervised image classification, classes detected and properly identified are extracted and then filtered to consider the minimal size requirements of the map. Then, the boundaries of the classified regions were subsequently vectorized. Feature extraction from imagery was limited to polygonal features.

For certain features such as water bodies better results were obtained by a simple region thresholding of the panchromatic band ETM8, considering values (gray levels) from 0 to 25. In all the fusion experiments, few feature improvements were obtained. The use of the single higher-resolution band permits the extraction of better-defined boundaries for the water bodies.

Semi-automated change detection
Figure 3 presents the steps for the semi-automated change detection process between old lake vectors and new lake vectors extracted from band ETM8 of Landsat-7. The non-intersection method is applied directly on the original vectors (3a) to obtain addition and deletion zones representing the change (3b). Then a 30m-buffer is generated around both original vectors to consider the planimetric accuracy tolerance (3c). The non-intersection method between the deletions and the new-buffered vectors determine the actual deletions while the non-intersection method between the additions and the old buffered vectors determine the actual additions (3d).

6. CONCLUDING REMARKS

Rapid change detection and data updating is becoming a significant requirement for mapping organizations and for detecting and monitoring changes of the landscape. The requirements for implementing automated procedures coupled with the availability of heterogeneous data are pushing towards the investigation, adaptation and evaluation of new approaches and techniques for detection and integration of change. The data sets to be compared at two or more time periods are usually in various forms. The domain of comparison is determined based on the level of data semantics and processing. This implies that generation of homogeneous features is required for each temporal datasets.

In this study emphasis was placed in automating the various tasks involved in feature extraction and change detection. The initial datasets at time 1 consisted of monochrome topographic raster maps and database vectors, while the new data set was Landsat 7 ETM+ imagery. The detection of planimetric changes was performed on vector data extracted from both data sets. The entropy measure of the image texture and image thresholding were used for the extraction of certain features from the raster maps. For the image data, image and theme enhancements were performed by applying various image fusion (e.g., PCI-IMGFUSE) and spectral enhancement (e.g., Tasselled cap, NDVI). Image thresholding and image classification were applied for the extraction of features from the imagery.

Having defined and process the data sets for the epoch comparison, change detection was performed both visually and in an automated mode. The proposed and implemented automated technique is based on the non-intersection of old and new features, where a) additions are determined as the difference between the new features and the common elements of the two epochs; and b) deletions are difference between the old features and the common elements of the two epochs. To account for accuracy tolerances spatial buffer zones were incorporated as well. The tests showed that feature extraction operations were more or less feature/theme dependent, while the proposed semi-automated change detection technique can be implemented in the operational environment. They also showed that visual change detection approaches can be improved when enhanced data sets are used, while the incorporation of image processing techniques supports the effort to automate the process. Research work needs to continue to provide improved methods and tools for rapid and flexible extraction of homogeneous features, while further work is required to refine, test and implement the change detection methods. As multitemporal imagery will be coming more available change detection from multi-temporal images has to be investigated for topographic applications.

7. ACKNOWLEDGEMENT

The authors wish to acknowledge and thank François Leduc, Yannick Cabana and Elizabeth Leblanc for their contributions

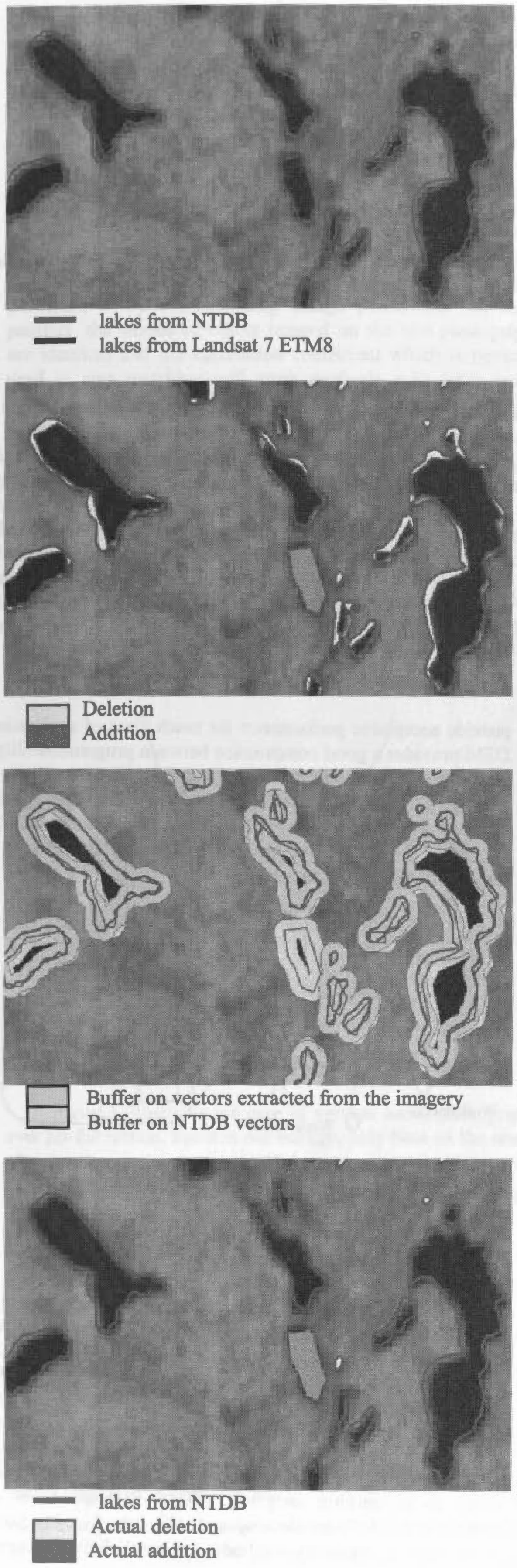


Figure 3: Change detection between old and new lake vectors.

in the developed approaches and their implementation.

8. REFERENCES

Armenakis C., F. Savopol, 1998. Mapping potential of the IRS-1C PAN satellite imagery. IAPRS, Vol 32, Part 4, pp. 23-26.

Armenakis, C. 2000. Differential approach for map revision from new multi-resolution satellite imagery and existing topographic data. IAPRS, Vol. 33, Part B4, pp. 99-104.

Armenakis C., F. Leduc, I. Cyr, F. Savopol,F. Cavayas, 2001. Change detection and extraction between raster maps and Landsat 7 imagery. IAPRS, Vol 34, Part 4/W5, pp. 10-21.

Carosio A., R. Stengele (1993). Automatic pattern recognition for economic map revision, Proceedings of the 16th Internatnational Cartographic Conference, Vol 1, Cologne, Germany

Chavez, P.S., S.C. Sides, J.A. Anderson, 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. Photogrammetric Engineering & Remote Sensing, Vol. 57, No. 3, pp. 295-303.

Cheng P., Th. Toutin, V. Tom, 2000. Orthorectification and data fusion of Landsat 7 data. ASPRS 2000, Washington, DC May 22-26 , 2000.

Haralick R.M., L.G. Shapiro (1992). Computer and Robotic Vision, Addison-Wesley Publishing Company Inc.

Heipke C., D.-M. Straub, 1999. Towards the automatic GIS update of vegetation areas from satellite imagery using digital landscape model and prior information. IAPRS, Vol. 32, Part 2W5, pp. 167-174.

Hayes D.J., S.A. Sader, 2001. Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series. Photogrammetric Engineering & Remote Sensing, Vol. 67, No. 9, pp. 1067-1075.

Mather, P. M., 1987. Computer processing of remotely-sensed images. Published by Biddles Ltd., Guildford, Surrey, Great Britain, 352 p.

Nebiker S., A. Carosio (1995). Automatic extraction and structuring of objects from scanned topographic maps - An alternative to the extraction from aerial and space images? in Automatic Extraction of Man-Made Objects from Aerial and Space Images, Ed. A. Gruen, O. Kuebler, P. Agouris, Birkhauser Verlag Basel, pp. 287-296.

Ohlhof T., T. Emge, W. Reinhardt, K. Leukert, C. Heipke, K. Pakzad, 2000. Generation and update of VMAP data using satellite and airborne imagery. CD-ROM IAPRS, Vol. XXXIII, Part B4, Amsterdam, pp.762-768.

Singh, A., 1989. Digital change detection techniques using remotely-sensed data. International Journal of Remote Sensing, Vol. 10, No. 6, pp. 989-1003.

Sonka, M., V. Hlavac, R. Boyle, 1993. Image Processing, Analysis and Machine Vision. Chapman & Hall, London, Great Britain, 555 p.

FAST PARALLEL IMAGE MATCHING ALGORITHM ON CLUSTER

Zhenge.Qiu [a], *, ZengBo.Qian [a], Zhihui Gong [a], Qing Xu [a]
[a] Zhengzhou Institute of Surveying and Mapping, 66# Longhai Middle Road Zhengzhou China,
Qiuzhenge@sina.com

Commission IV

KEY WORDS: Parallel Algorithm, Cluster, DSM (Distributed Shared Memory), Image Matching, DEM (Digital Elevation Model), No-uniform Image sampling

ABSTRACT:

This paper introduces our experiences with developing fast Parallel stereo matching algorithm on Cluster. The key technique of generating DEMs or DSMs' from remote-sensing images is stereo matching, and it's one of the most time consuming algorithms, so that many efforts have been carried out to develop fast algorithms. Many difficulty problems continue to challenge researchers in this field; they are occlusion, large parallax range, and radiant distortion. To solve these problems need more complicated image-matching algorithm that is more time consuming. Recent advance in high-speed networks, rapid improvement in microprocessor design, and availability of highly performing clustering software implementations enables cost-effective high-performance parallel computing on clusters very attractive. The cluster is a very useful platform on which to develop fast parallel image matching algorithm dealing with the difficult problems above-mentioned. A novel operator of image matching based on no-uniform image resample is used in parallel image matching algorithm, this operator reduces the influence of geometric distortion on those matching operators which are in effect under equal parallax assumption, for example, correlation operator, and expands the scope of them to be in effect.

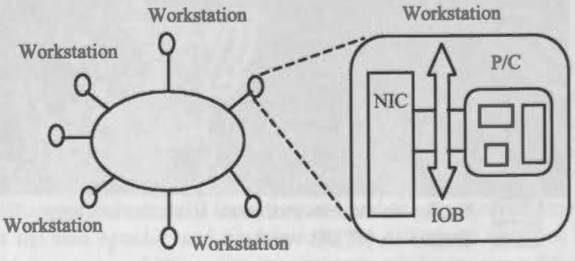
1. INTRODUCTION

The key technique of generating DEMs or DSMs' from remote-sensing images is stereo matching, and it's one of the most time consuming algorithms, so that many efforts have been carried out to develop fast algorithms of image matching. Many difficulty problems continue to challenge researchers in computer vision and digital photogrammetry and other scientific field. The main problems are occlusion, large parallax range, and radiant distortion. To solve these problems need more complicated image-matching algorithm that is more time consuming.

The current trend in high performance computing is clustering and distributed computing. In clusters, powerful low cost workstations and/or PCs are linked through fast communication interfaces to achieve high performance parallel computing as depicted in figure1. Recent increases in communication speeds, microprocessor clocks, protocol efficiencies coupled with the availability of high performance public domain software including operating system, compiler tools, and message passing libraries, make cluster based computing appealing in terms of both high-performance computing and cost-effectiveness. Parallel computing on clustered systems is a viable and attractive proposition due to the high communication speeds of modern networks.

MPI (Message Passing Interface) approach is considered one of the most mature methods currently used in parallel programming mainly due to the relative simplicity of using the method by writing a set of library functions or an API (Application Program Interface) callable from C, C++ or Fortran programs.

DSM (Distributed Shared Memory) is another mature methods currently used in parallel programming. DSM systems provide the illusion of shared memory on top of standard message passing hardware at very low implementation cost, but provide acceptable performance for much class of applications. DSM provides a good compromise between programmability of shared memory multiprocessors and hardware simplicity of message passing multicomputers. Many software DSM systems, such as Ivy [8], Midway [1], Munin [2], TreadMarks [7], Cashmere [10], and JIAJIA[3] have been implemented on the top of message passing hardware.



P/C (Microprocessor and Cache) NIC (Network Interface Circuitry)
IOB (I/O Bus)

Figure1. Cluster of PC

This paper introduces our experiences with developing fast Parallel stereo matching algorithm on the home-based software DSM JIAJIA [3] and discusses techniques of parallelizing a sequential image matching program to run on software DSM.

Satisfactory speedups are achieved for all of these applications on a cluster of eight Pentium II PCs connected by a 100Mbps switched Ethernet. The rest of this paper is organized as follows. Section 2 discusses the image-matching algorithm based on no-uniform sampling, JIAJIA Software DSM as a useful parallel software-developing tool is introduced in Section 3. Section 4 discusses methods for parallelizing sequential

---

image matching programs. The conclusion of this paper is drawn in Section 5.

2. IMAGE MATCHING ALGORITHM BASED ON NON-UNIFORM SAMPLING

The geometric distortion is the main problem of image matching; the no uniform sampling can effectively reduce the influence of geometric distortion, which, in essence, is the relief displacement of pairs of stereoscopic pictures. Figure 2 illustrates the case of vertical aerial photographs over flat terrain, in which all points in their planimetrically correct positions, every corresponding image points has the same parallax, the shapes of object imaged on the two photographs are identical and the correlation coefficient which is typically used in area matching will work perfectly with large image correlation windows. As shown in figure 2, $G(X,Y,Z)$ depicts the reflectance map of terrain and $g_1(x_l,y_l)$, $g_2(x_r,y_r)$ are the corresponding images on the pairs of stereoscopic pictures and their relationship can be defined by a constant d.

$$g_1(x_l,y_l)=g_2(x_l+d,y_l) \tag{1}$$

In this case, in the matching windows every corresponding points has a unique parallax, large window can be used to avoid ambiguous and at the right matching position the correlation will reach maximum.

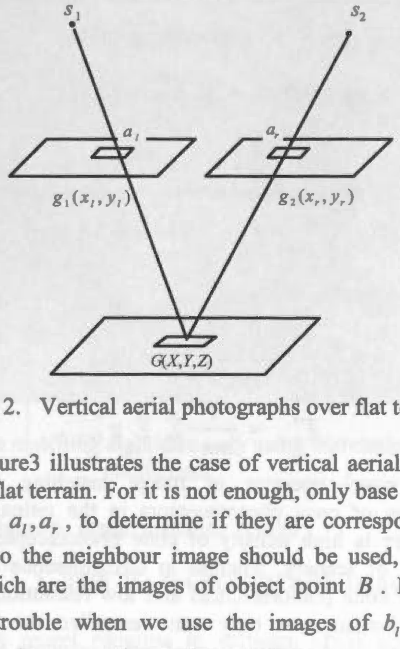


Figure 2. Vertical aerial photographs over flat terrain

Figure3 illustrates the case of vertical aerial photographs over no-flat terrain. For it is not enough, only base on the image of points a_l, a_r, to determine if they are corresponding points or not, so the neighbour image should be used, for example b_l, b_r which are the images of object point B. However, we have a trouble when we use the images of b_l, b_r, for the parallax of a_l, a_r and parallax of b_l, b_r are not identical, in other word, there is a geometric distortion between the image on left photo and that on right photo.

The geometric distortion between the image on left photo and that on right photo, in essence, is the problem of relief displacement due to height difference. As shown in figure3, object point B and point B_0 has the same planimetrically positions and different elevation, the difference between the displacement of b_l (on left image) and that of b_r (on right image) is geometric distortion, where b_l, b_r are imaged from B ; b_{l_0}, b_{r_0} are imaged from B_0 . Let δh_l be the relief displacement between b_l, b_{l_0} on left image, it can be expressed as

$$\delta h_l = \frac{\Delta h \cdot rb_l}{H} \tag{2}$$

Where rb_l is the radial image distance of b_l , Δh is the height of B above the reference object plane, H is the flying height.

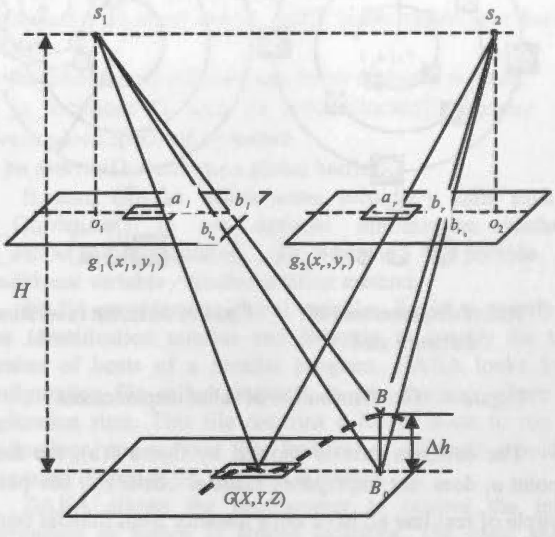


Figure 3. Vertical aerial photographs over no-flat terrain

On the right photo, b_r imaged from object point B also has a relief displacement δh_r between b_r, b_{r_0} which can be described as

$$\delta h_r = \frac{\Delta h \cdot rb_r}{H} \tag{3}$$

Where rb_r is the radial image distance of b_r , Δh is the height of B above the reference object plane, H is the flying height.

The geometric distortion between the image on left photo and that on right photo can be depicted as:

$$\Delta_{lr} = |\delta h_l - \delta h_r| \tag{4}$$

Equation 4 shows that if $|\delta h_l|, |\delta h_r|$ can be reduced, Δ_{lr} can be reduced dramatically.

From equation 2 and equation 3 we can obviously find such relationship that relief displacement is in direct ratio to Δh, rb_l or rb_r, and in inverse ratio to H. It is reasonable to suppose that terrain relief has a continuous change. For studying the change trend of relief displacement, it is also reasonable to suppose that possible maximum value of Δh is in direct ratio to distance of object B from A, this relationship can be depicted by figure 4. Figure 4(a) describes the distribution of relief displacement on left photo, and shows the case that a_l superposes fiducial center o_l, these circles represent the image points having same distance from fiducial center o_l or a_l, under the hypotheses above mentioned, these points also have the same possible maximum value of Δh , namely

$rb_l = c_1, \Delta h = c_2$, and the possible maximum value of Δh is in direct ratio to rb_l. According to equation 2 and equation 3, we can bring out that at each circle, every image points have a relief displacement of same possible maximum value that is in direct ratio to rb_l.

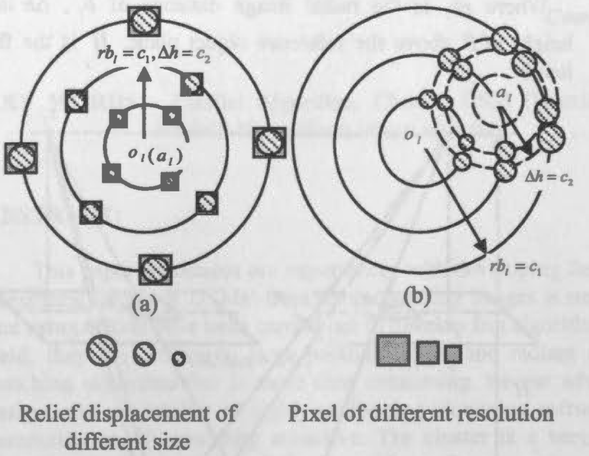

$rb_l = c_1, \Delta h = c_2$
$o_l(a_l)$
(a)
a_l
o_l
$\Delta h = c_2$
$rb_l = c_1$
(b)
Relief displacement of different size
Pixel of different resolution

Figure 4. The distribution of relief displacement

The common case is showed by figure 4(b), the focusing point a_l does not superposes fiducial center o_l, the points at circle of real line all have same distance from fiducial center o_l, namely $rb_l = c_1$, the points at circle of broken line all have same distance from focusing point a_l, according to previous assumption it can be accepted that the their possible maximum value of Δh is also same, namely $\Delta h = c_2$, thus at the cross points of circle of real line and circle of broken line, all the value $rb_l \cdot \Delta h$ are same. Although in this case the distribution of relief displacement does not have symmetry characteristic as the case shown in figure 4(a), they are similar and the distribution of the relief displacement at circle of broken line also has the property of being in direct ratio to rb_l (namely the distance from focusing point a_l), the dissimilarity is that at the point near fiducial center o_l the value is smaller, on the contrary, at the point far away from fiducial center o_l the value is larger.

If we use no-uniform sampling as depicted in figure 4(a), namely in this kind of sampling method, the pixels at different circle has different size, and the size is in direct ratio to semi diameter of the circle (in the case depicted in figure 4(b) the size is in direct ratio to semi diameter of the circle of broken line). In discrete representation, if this kind of pixels of different size is used as metric unit, the numerical value of relief displacement will observably reduce. For example at certain circle, the value of relief displacement: $\delta h_l = 0.1mm$ and the pixel size of no uniform sampling here: $\Delta pixel = 0.2mm$, so the integer value of relief displacement is zero (0 $\Delta pixel$). Another advantage of no-uniform sampling is that the resolution at inner circle is higher than that at outer circle, in this way these image information near the matching point a_l (focusing point) is enlarged and those image information away from the matching point a_l is compressed. So if using same size of match window more information can be used in matching, more detail information in central part of match window is used, more information in low resolution in marginal part of match window is also used and the relief displacement of them is markedly reduced. This analysis of the function of no uniform sampling is also exact to right photo of stereo pair.

In this way every corresponding points' parallax in match window will be approximately equal at exact matched position, so this method can reduces the influence of geometric distortion on those matching operators which are in effect under equal parallax assumption, for example correlation operator, and expands the scope of them to be in effect.

The image matching strategy depicted in figure 5 used here are:

(1) Re-sampling the image at the focusing point on left and right photos ,and get two new images;

(2) Calculate the correlation coefficient of the two new images;

(3) Determine if they are matched images.

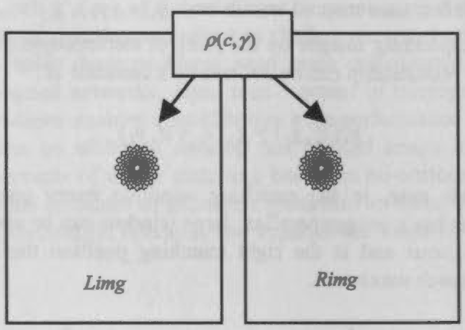


Figure 5. Image matching strategy

The correlation coefficient of discrete digital images can be calculated as follow:

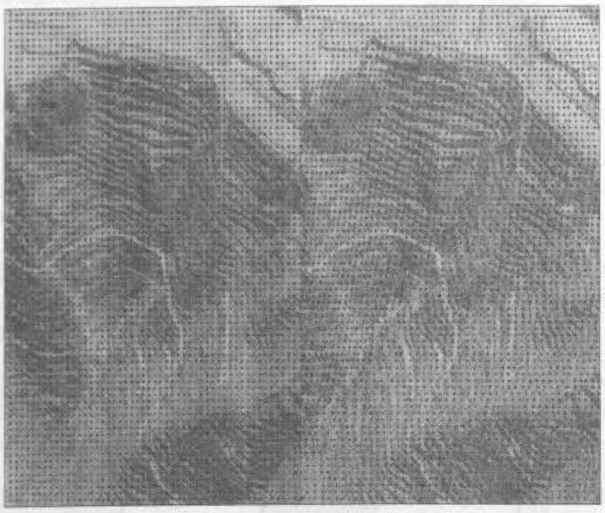
$$\rho(c,r) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(g_{i,j} - \overline{g})(g'_{i+r,j+c} - \overline{g}'_{r,c})}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(g_{i,j} - \overline{g})^2 \sum_{i=1}^{m}\sum_{j=1}^{n}(g'_{i+r,j+c} - \overline{g}'_{r,c})^2}} \tag{5}$$

Where:

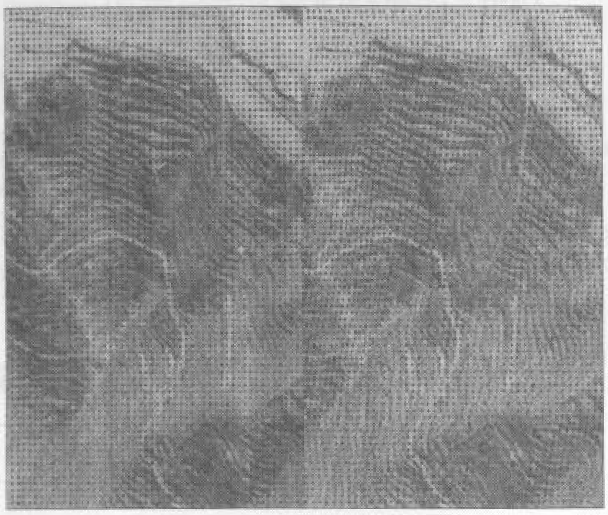
$$\left.\begin{aligned} g &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} g_{i,j} \\ \overline{g}'_{r,c} &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} g'_{i+r,j+c} \end{aligned}\right\} \tag{6}$$

This novel operator of image matching imitates the distribution of cone photoreceptors in the retina of man. In fovea there is high density of cone photoreceptors and high resolution of scenery, whereas in circumference there is low density of cone photoreceptors and low resolution of scenery. This mechanism not only can centralize the resource of computation on the centre part of match-window and use more wide scope image information, but also reduces the influence of geometric distortion on image match, which is the difference between relief displacements of two stereoscopic images in essence.

The aerial images of some area of china (1:20000) are used to test this novel image matching algorithm, the right matching ratio of this algorithm is about 70%-95%, about 15% higher than that of algorithm only using correlation coefficient, one of the example is showed in figure 6.



(a) Image matching algorithm based on no uniform sampling



(b) Image matching algorithm only using correlation coefficient

Figure 6. Image matching results

3. THE JIAJIA SOFTWARE DSM

With recent progress in software DSM and connection networks, software DSM can achieve an acceptable performance. On the one hand, optimization methods such as multiple-writer protocol and lazy implementation of release consistency can significantly improve the performance of software DSM. On the other hand, recent progress on networks reduces the sensitivity of parallel application performance to the communication amounts. Previous study shows that on a large variety of programs, the performance of well-optimized software DSM system is comparable to that of a message passing system. The fast parallel image-matching algorithm introduced in this paper is developed on the home-based software DSM JIAJIA.

JIAJIA is a home-based software DSM which supports scope consistency [6]. It has two distinguishing features compared to other recent software DSM systems. First, it combines physical memories of multiple computers to form a larger shared space. Second, it combines the lock-based cache coherence protocol and many optimization methods.

3.1 Programming Interface

The API of JIAJIA is similar to that of other software DSM systems. It provides following basic routines supporting SPMD (Single program multiple datum) shared memory parallel programming to the applications:

(1) jia_init(argc,argv) and jia_exit(): Initialize and shut down JIAJIA.
(2) jia_alloc(size): Allocate size bytes of shared memory.
(3) jia_lock(lockid) and jia_unlock(lockid): Acquire and release a lock specified by lockid.
(4) jia_barrier(): Performs a global barrier.

Besides, JIAJIA offers some subsidiary calls such as jia_Configure() to set optional optimization methods, jia_setcv(), jia_resetcv(), jia_waitcv() to provide the conditional variable synchronization method.

JIAJIA provides two global variables: jiapid to specify the host identification number and jiahosts to specify the total number of hosts of a parallel program. JIAJIA looks for a configuration file called jiahosts in the directory where the application runs. This file contains a list of hosts to run the applications, one per line. The first line of .jiahosts should be the master on which the program is started.

JIAJIA allows the programmer to control the initial distribution of homes of shared locations. The basic shared memory allocation function in JIAJIA is jia_alloc3(size,blocksize,starthost) which allocates size bytes cyclically across all hosts, each time blocksize bytes. The starthost parameter specifies the host from which the allocation starts. The simple call jia_alloc(size) equals jia_alloc3(size,Pageszie,0).

JIAJIA also provides some MPI-like message passing calls: jia_send(),jia_recv(),jia_bcast (),and jia_reduce().

3.2 Coherence Protocol

Figure 7 shows memory organization of JIAJIA. As shown, each shared pages has a home node and home of shared pages are distributed across all nodes. References to home pages hit locality, references to no-home pages cause these pages to be fetched from their home and cached cache locally. A cached page may be in one of three states: Invalid (INV), Read-Only (RO), and Read-Write (RW). When the number of locally cached pages is larger than the maximum number allowed, some aged cache pages must be replaced to its own home to make room for the new page. This allows JIAJIA to support shared memory that is large than physical memory of one machine.

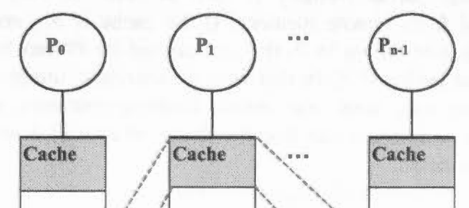


Figure 7. Memory organization of JIAJIA

JIAJIA implements the scope memory consistency model. Multiple writer technique is employed to reduce false sharing. In JIAJIA, the coherence of cached pages is maintained through requiring the lock-releasing processor to send to the lock write notices about modified pages in the associated critical section, the lock-acquiring processor to invalidate cached pages that are notified as obsolete by the associated write-notices in the lock. This protocol maintains coherence through write notices kept on the lock and consequently eliminates the requirement of directory. Some optimization methods of the protocol include single-writer detection [3], incarnation number technique [3], write vector technique [5], home migration [4], lazy home page write detections, and SMP optimization. These optimizations can reduce page faults, message amounts, and diffs dramatically and consequently improve performance significantly.

4. FAST PARALLEL IMAGE MATCHING ALGORITHM ON THE SOFTERWARE DSM

Parallel program segments can be categorized into five patterns; single-process sequential, mutual-exclusive sequential, data-parallel, task-parallel, and common-parallel. The data-parallel pattern is used in parallelizing the sequential matching program with the API of JIAJIA. In our experiment, a cluster of eight Pentium II PCs connected by a 100Mbps switched Ethernet is used. The stereo images are divided into eight parts as showed in figure 8, each PC carries out the matching task of one parts of stereo image. For example, P_0 carries out the matching task of （Limg0,Rimg0,Φ）, the Φ is the matching operator; P_1 carries out the matching task of （Limg1,Rimg1,Φ）; ... P7 carries out the matching task of（Limg7,Rimg7,Φ）。P_0,P_1.......P_7 compute parallelly.

Data locality is crucial for performance in DSM machines, due to the difference in access times between local and remote memories. In image matching algorithm, high data locality can be reached by proper data dividing and proper task mapping to each computer. We put each part of stereo image data in the computer that used them, so that every computer can get the needed image data locally, in this way, massive communication operations are avoided. Even though the optimizing method above mentioned is used, at boundary, matching operator based on window have to fetch datum from the memory of other computers (remote memory). As shown in figure 8, in P_0, when the matching operator deals with the points at boundary, it needs the data in area FL0 and FR0, and the data in area FL0 and FR0 are in computer P_1 (FL1 and FR1), so that P_0 has to fetch the data in area FL0 and FR0 from P_1. As shown in figure 7, for reducing this kind of remote fetching operations, in each computer partial memory is used as cache to keep the data fetched from remote memory. If the cache is big enough the total line of datum in P_1 that are needed by P0 can be fetched into the cache of P_0. In this way, at boundary, image matching operator only need one remote fetching operation, after this remote fetching, it can find the datum of area FL0 and FR0 in its own cache.

The main steps of the program are:

```
(1) Edit .jaconf, .jiahosts files to give the initial parameters.
(2) Call Jia_init(argc,argv) to create multithreads.
(3) Call Jia_alloc() to allocate memory for each thread.
(4) loop:
        Computer $P_0$- $P_n$ reads part of image separately,
        Call jia_barrier() to Synchronize.
(5) loop:
        Computer $P_0$- $P_n$ creates pyramid of part image separately,
        Call jia_barrier() to Synchronize.
(6) loop:
        Computer $P_0$- $P_n$ matches image separately,
        Call jia_barrier() to Synchronize.
(7) Output results.
```

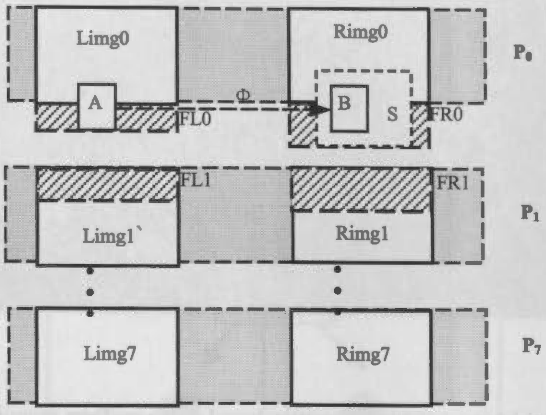


Figure 8. Division of image matching task

Satisfactory speedups are achieved for this application on a cluster of eight Pentium II PCs connected by a 100Mbps switched Ethernet. As showed in table 1,when two PCs are used, the speedup ratio is 1.8; when four PCs are used, the speedup ratio is 3.7; when eight PCs are used, the speedup ratio is 7.5.The speedup ratio is near the ideal linearity speedup ratio. The speed of finding corresponding points reaches 3200 pair/second, when eight PCs are used.

Table 1 depicts the speedup ratio of parallel image matching algorithm.

Number of computers	Two	Four	Seven	Eight
Speedup ratio	1.8	3.7	6.4	7.5

Table 1. Speedup ratio of parallel image matching algorithm

5. CONCLUSIONS

In this paper, we have presented a novel operator of image matching that imitates the distributing of cone photoreceptors in retina of man and its fast parallel algorithm based on the softer ware DSM.

The advantage of suggested image matching algorithm base no-uniform sampling is that this mechanism not only can centralize the resource of computation on the center part of match-window and use more wide scope image information, but also reduces the influence of geometric distortion on image match, which is the difference between relief distortions of two stereoscopic images in essence.

From the experimental results, we have observed that it is important to pay great attention to the development of cluster system. A well-designed cluster system can provide highly available access to applications, system resources and data, together with other features such as scalability and single system image. There is a great demand of large memory size and more high calculation speed in the field of photogrammetry and remote sensing, cluster system can meet this demand at low cost, especially in processing and analysing remote sensing image of great size and establishment of image data bank.

6. REFERENCES

[1] Bershad,B., 1993f. The Midway Distributed Shared Memory System. *In Proc. of the 38th IEEE Int'l CompCon Conf.*, pp. 528-537.

[2]Carter, J., 1991o. Implementation and Performance of Munin. *In Proc. Of the 13th Sym. on Operating Systems Principles,* pp.152-164.

[3] Hu, W., 1999a. Reducing System Overhead in Home-Based Software DSMs. *In Proc. Of the 13th Int'l Parallel Processing Symp.*, pp. 167-173.

[4] Hu, W., 1999j. Home Migration in Home-Based Software DSMs. *In Proc. of the 1st Workshop on Software Distributed Shared Memory,* pp. 21-26.

[5]Hu,W., 1999j. Reducing Message Overheads in Home-Based Software DSMs. *In Proc. of the 1st Workshop on Software Distributed Shared Memory,* pp. 7-11.

[6] Iftode, L., 1998a. *Home-based Shared Virtual Memory.* Ph. D. Thesis, Princeton University.

[7]Poggio,T., 1985s. Computational vision and regularization theory. *Nature,*317(26),p.324-319.

[8] Qiu,Z., 2001o. *The Theory/Algorithms and Realization of Stereo Vision.* Ph.D.Thesis, Zhengzhou Institute of Surveying and Mapping.

INDEX OF AUTHORS

A

Ababspour Rahim 82
Agouris Peggy 412, 781
Alecu Corina 684
Allili M. 386
Almer Alexander 436
Altan M. Orhan 656
Ameri Babak 308,.419
Apers P. M. G. 202
Archinal Brent 713, 743
Armenakis Costas 668, 792
Arndt W. C. 457 786
Asami Yasushi 72
Asano Hiroshi 442
Auclair-Fortier M. F. 386
Azizi Ali 599, 601

B

Bayrante L. F. 488
Boaming Zhang 515
Baral Toya Nath 613
Batistella Mateus 557
Bédard Yvan 142, 196, 344
Bernier Eveline 344
Bertolotto Michela 190
Bian FuLing 39, 231,456
Bien O. C. 488
Bing Tan 515
Blais Rod 482
Bobrich Joachim 502
Bofinger Jan-Martin 157
Bohmann Guido 447
Bouchard André 532
Boutin Denis 166
Brisebois Alexandre 142
Brodeur Jean 142, 196
Buckley Simon 581
Burshtynska Khrystyna 509
Buswell Michael 350

C

Cai Guoray 48, 100
Camit Rex A. 488
Campos dos Santos J. L. 202
Carbonneau Y. 547
Carswell James 190
Chen Jun 10, 106
Chen Liang-Chien 16, 620
Cheng Eric D. 536
Cheng Qiuming 48
Chénier R. 547
Chi Kwang-Hoon 54
Choi Kyung-Mee 88
Christakos George 88
Chung Chang-Jo 54, 72, 124, 154
Chung C. F. 132
Clarke Peter J. 581
Csato Eva 602
Cyr Isabelle 792

D

De By R. A. 202
Di Kaichang 733
Doihara T. 78, 255, 259
Dongyang Ma 515
Duxbury T. C. 743

E

Edwards Dan 167
Edwards Stuart J. 581
Ehlers Manfred 425
Eickhorst Kristin 412
Elias Birgit 249
Espiridion, J. A. 488

F

Fabbri Andrea G. 124, 154
Fabiyi Oluseyi 690
Fadaie Kian 313
Faundeen John L. 350
Finn Michael P. 704
Fischer Christian 460
Forberg Andrea 225
Fryer John G. 644

G

Gao Xiaolu 72
Ghosh Suddhasheel 361
Giese B. 723
Gilbert Michel 185
Goldstein Norman 419
Gong Janya 208
Gong Zhihui 798
Gorte Ben 465
Grabovac Nickolas 175
Grunsky Eric 563
Guarnieri Alberto 406
Guindon Bert 375
Guo Wei 33
Gupta Hari Shanker 31
Gyftakis Sotirios 781

H

Habbane Mohammed 313
Hahn Michael 411, 495, 599, 613
Haj-Yehia Basheer 775
Han Gang 106
Han-Sze-Chuen Dominique 269
Harrie Lars 237
Hatger Carsten 575
Hay Geoffrey. J. 532
Heipke Christian 470
Hellwich Olaf 786
Hilbring Désiree 476
Hilker Thomas 411
Højholt Peter 776
Holland David 341
Holme Dorthe 776
Hong Jung-Hong 60, 395, 755
Hoyer M. 738
Howington-Kraus Elpitha 713, 723
Hu Shunfu 496
Huang Duo 319
Huang Gordon 83
Hunter Andrew 593

I

Illert Andreas 356

J

Jansa Josef 569
Jaumann R. 738
Jeong Jong Hyeok 626
Jiang Jie 106
Jiang Jingjue 456
Jordan Thomas 481
Jung Stefan 425

K

Kada Martin 243
Kagawa Yoshiaki 662
Kandawasvika Admire 495
Kanengieter Ronald L. 350
Keighan Edric 331
Kéna-Cohen Serge 331
Kikuchi Yuki 442
Kim Jung Rack 637
Kirk Randolph 713, 723, 743
Knöpfle Walter 675
Koch A. 470
Kojima Hirohito 27
Kraak Menno-Jan 279
Kristóf D. 602

L

Larrivée, Suzie 344
Lee Hae-Yeoun 631
Lee Scott D, ...297, 728
Lee Y. C. 216
Lehner Manfred 675
Lehto Lassi 237, 356
Li Bonan 100
Li Cheng-ming 111, 113, 163, 371, 778
Li Jonathan 83
Li Luqun 111
Li Qingquan 148, 212
Li Ron 733
Li Zhi-lin 10
Liao Hsiung-Peng 755
Lin Zong-jian 111, 163, 371
Lohmann Peter 470
Loubier Eric 608
Lu An-min 371
Lu Dengsheng 557
Lu Wei 78, 255, 259
Lu Yi 536

M

Madden Marguerite 481
Magalhães C. 202
Manandhar Dinesh 650
Mansourian Ali 305
Marceau Danielle J. 532
Marchand Pierre 142
Martin Marie-Eve 608
Matthies Larry H. 733
Matz K-D. 738
Mayer Helmut 225
McGeown Liam 190
Michaelsen Eckart 401
Mills Jon P. 581
Min Deng 113
Miquel Maryvonne 142
Mitchell Harvey L. 581, 644
Molenaar Martien 263, 279
Morales Javier 324
Moran Emilio 557
Moshkovitz Avner 419
Moulin Bernard 269, 274
Mountrakis Georgios 781
Muller Jan-Peter 637
Mustière Sébastien 269, 274

N

Nakagawa Masafumi 662
Nebiker Stephan 391
Neidhart Frank 613
Neumann G. A. 743
Niedermaier G. 738
Nissen Flemming 356

O

Obayashi Shigeyuki 27

Oberst J. 723
Ohsawa Yutaka 33
Olson Clark F. 733
Orth Robert 308
Otto Stephen 502

P

Panem C. C. 488
Papanikolaou E. 792
Pâquet Robert 644
Park No-Wook 54
Park Wonkyu 631
Peled Ammatzia 774, 775
Penna de Moraes Claudio 434
Peteri Renaud 768
Pethick John 581
Pilouk Morakot 66
Poulin P. 386
Pouliot Jacynthe 142
Prasad Nupoor 361
Prisniakov V. 506
Prisniakova L. 506
Proulx Marie-Josée 344

Q

Qian Zeng Bo 798
Qiu Zhenge 798

R

Radwan Mostafa 324
Rahman Alias 66
Raizman Yuri 774
Rajabi Mohammad 482
Ramos Fabien 43
Ranchin Thierry 768
Rase Wolf-Dieter 430
Rau Jiann -Yeou 16
Reinartz Peter 675
Reinhart Wolfgang 495
Reiss Peter 526
Ritter David 602
Roatsch Th. 738
Rochon Guy 331
Roth Achim 675
Rottensteiner Franz 569
Roy Parth Sarathi 361
Ruisong Xu 699
Ruotsalainen Reino 356
Rystedt Bengt 356

S

Sabourin Marcel 166
Sakurai Masayuki 33
Salo-Merta Leena 355
Samadzadegan Farhad 82, 599, 601
Saran Sameer 361
Sarjakoski Tapani 356
Sarjakoski L. Tiina 237, 356
Sarpoulaki Mohammad 82, 601
Savin Elena 684
Savopol Florin 668
Scheidt Douglas J. 704
Schetselaar Ernst 132
Schiewe Jochen 380, 447
Scholten F. 738
Schröder Dietrich 411
Sester Monika 291, 356
Sha Zong-Yao 231
Shahriari Nadia 118
Shan Jie 297, 521, 728
Shankaie Ali 308
Sharma S. K. 697
Shi L. 208
Shi Wen-Zhong (John) 148
Shibasaki Ryosuke 650, 662
Shoji Tetsuya 94
Shu-long Zhu 552
Simpson Justin 167
Skaloud Jan 614
Soderblom Laurence 713, 723
Sokolova Valentina B. 680
Stambouloglou Efstratios 521
Stancalie Gheorghe 684
Stefanidis Anthony 781
Stelzl Harald 436
Stilla Uwe 401
Strunz Günter 675
Stroemer Katrin 425
Su Zeal 60, 395

T

Takagi Masataka 442, 626
Talebzadeh Ahmad 601
Talukdar Gautam 361
Tan Xiaojun 39
Tao C. Vincent 118, 593
Tempfli Klaus 263
Teo Tee-Ann 620
Thiemann Frank 286
Thijs Peter 190
Tolley Cecilia 313
Torun Abdulvahit 179
Toutin Thierry 547
Trinder John 536

U

Ulubay A. 656
Usery E. Lynn 704

V

Valadan Zoej M. J. 305
Vallet Julien 614
Van Gasselt S. 738

Vettore Antonio 406
Vieira Dias Luiz Alberto 434
Volz Steffen 157

W

Wählisch M. 738
Walter Volker 749
Wang Chong 212
Wang Guizhi 138
Wang Hongmei 48
Wang Jiguo 778
Wang Ping 536
Wang Pingtao 255, 259
Wang Y.Q. (Yeqiao) 22, 208
Wang Zhijun 536
Wehn H. 419
Welch Roy 481
Wessel B. 786
Wewel F. 738
Wiedemann Christian 786
Wilson Mark L. 88
Willrich Felicitas 761
Wu Simon 83

X

Xiaoyong Chen 113
Xiong Hanjiang 208
Xu F. 733
Xu Qing 515, 798
Xu Z. 216

Y

Yang Bisheng 148
Yao Xuefeng 319
Yaolin Liu 279
Yin Jie 163, 778
Ying Ming 456
Ying Xie Zhi 212
Yong Liu 212
Yoon Jong-Suk 728

Z

Zayac Alexandr 509
Zhan Qingming 263
Zhang Yetin 319
Zhang Ying 375
Zhang Yun 453, 587
Zhao Renliang 10
Zhu Guomin 778
Zhu Qing 319
Ziou Djemel 386
Zlatanova Siyka 66
Zwick Harold 419