Natural Resources Canada Ressources naturelles Canada

GEOMATICS CANADA
OPEN FILE 50

Literature review on the potential of high-resolution optical satellite imagery for the extraction of mapping information

M. Turgeon-Pelchat

2019

Canada

GEOMATICS CANADA
OPEN FILE 50

Literature review on the potential of high-resolution optical satellite imagery for the extraction of mapping information

M. Turgeon-Pelchat[1]

[1]Natural Resources Canada, 50 Place de la Cité, Suite 212, P.O. Box 162, Sherbrooke, Québec J1H 4G9

2019

Permanent link: https://doi.org/10.4095/314763

This publication is available for free download through GEOSCAN (https://geoscan.nrcan.gc.ca/).

Recommended citation
Turgeon-Pelchat, M., 2019. Literature review on the potential of high-resolution optical satellite imagery for the extraction of mapping information; Geomatics Canada, Open File 50, 58 p. https://doi.org/10.4095/314763

Publications in this series have not been edited; they are released as submitted by the author.

Table of contents

List of tables

List of equations

List of acronyms

ACO	Ant Colony Optimization
ANN	Artificial Neural Network
AWEI	Automated Water Extraction Index
BDTQ	Base de données topographiques du Québec
BSP	Binary Space Partitioning
CCMEO	Canada Centre for Mapping and Earth Observation
ConvNet	Convolutional Network
CRIM	Computer Research Institute of Montréal
DEM	Digital Elevation Model
DMP	Differential Morphological Profile
DSM	Digital Surface Model
DTM	Digital Terrain Model
FLST	Fast Level Set Transform
GAN	General Adaptive Neighbourhood
GDB	Geospatial Database
GLI	Green Leaf Index
HR	High Resolution
HRWI	High Resolution Water Index
HSV	Hue, Saturation, Value
IDW	Inverse Distance Weighted
LiDAR	Light Detection and Ranging
MLC	Maximum Likelihood Classification
MNDWI	Modified Normalized Difference Water Index
MSAVI	Modified Soil Adjusted Vegetation Index
MSI	Morphological Shadow Index
nDSM	normalized Digital Surface Model
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NGRDI	Normalized Green Red Difference Index
NIR	Near Infrared
NRCan	Natural Resources Canada
NSVDI	Normalized Saturation-Value Index
OSM	Open Street Map
OTB	Orfeo ToolBox
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
RADAR	Radio Detection and Ranging
RGB	Red, Green, Blue
SAR	Synthetic Aperture Radar
SOM	Self-Organizing Map
SVM	Support Vector Machine
TGI	Triangular Greenness Index
TIN	Triangulated Irregular Network

1 Introduction

The extraction of mapping information from optical satellite imagery is a remote sensing application that has been documented for years. Methods and technology are constantly evolving, as well as the quantity and quality of available images. The advent of high resolution (HR) optical satellite imagery for all federal departments will provide national coverage and annual updates for certain Canadian regions. The availability of these images will provide a wide range of possibilities for information analyses, processing and extraction. On the basis of scientific literature, this report documents the operational methods and applications concerning the extraction of mapping information with HR optical imagery. The purpose of this report is to establish the imagery's potential for extraction in order to prioritize development efforts, identify limitations and, most importantly, provide an overview of current knowledge on the topic. It should be noted that this report is not intended to be comprehensive. Furthermore, the term "high resolution" refers to images with a spatial resolution between 0.5 m and 2.0 m (Navulur, 2007).

Sections 2 to 5 present the methods based on different themes. These themes have been selected to ensure the continuity of the extraction projects underway and due to their alignment with the priorities of the Canada Centre for Mapping and Earth Observation (CCMEO). The four extraction themes selected are buildings, hydrography, the road network and urban vegetation.

2 Building extraction

This section identifies and summarizes building extraction methods and is divided into three categories: the review of research using only optical imagery for extraction; research performed by combining optical imagery with another type of sensor, i.e. Light Detection and Ranging (LiDAR) or Radio Detection and Ranging (RADAR); and operational applications related to building extraction.

2.1 Optical imagery-based methods

There are different methods available to extract buildings from HR imagery. This review is primarily based on the method proposed by Dubois (2014). Most of the articles cited by Dubois (2014) are also found in the review written by Ghanea *et al*. (2016). A small selection of articles was added to provide a more comprehensive and updated overview of building extraction methods. Table 2.1 at the end of the section presents an overview of the studies addressed in this section, as well as their results.

Building edge detection method

Jin and Davis (2005) use the Differential Morphological Profile (DMP) to detect buildings. The DMP detects shapes or structures in an image of variable size (Jin and Davis, 2005). Edge detection is based on the building structure itself or the detection of the building's shadow (contextual). The process is iterative and each iteration refines the edge detection to eventually shape the building's outline. The algorithm has been improved by adding extraction based on the spectral signature of buildings. The final product of this method combines three extraction techniques, i.e. structural, contextual and spectral. The study by Jin and Davis focuses on urban areas because of the high density and regularity of the buildings in these areas.

Hough transform

The Hough transform extracts linear patterns or curves from an image (Jung and Schramm, 2004). The transformation identifies all the linear equations, which can explain each pixel. This method delivers excellent results when extracting rectangular or round buildings from imagery (San and Turker, 2010). Conversely, the algorithm does not detect buildings with complex shapes.

Shapes that may be used to qualify a building

The Karantzalos and Paragios (2009) method uses a bank of shapes to establish its research. Based on these shapes, the algorithm can identify similar patterns in the image. For each iteration, the algorithm changes the shape of the research polygon so that it will correspond to the reality of the patterns found. The downside of this method is that complex shapes, which are not included in the bank, will not be detected. This method stands out due to its rate of accurate representations of detected footprints (90% and over). Furthermore, certain study areas are located in regions comprising abundant vegetation.

Mean shift segmentation method

Aytekin *et al*. (2011) detect buildings by deduction. The authors used a Normalized Difference Vegetation Index (NDVI) to remove the forest, morphological operations to mask roads and a Principal Component Analysis (PCA) to remove remaining artifacts. They applied a mean shift segmentation to the remaining areas to extract the buildings (Aytekin *et al.*, 2011). Mean shift segmentation is explained in Appendix I. The method proposed by Aytekin *et al.* (2011) is effective regardless of the shape of the buildings to be extracted (Jiang *et al.*, 2008). Aytekin *et al*. (2011) achieved interesting, albeit very variable, results among the different areas tested. The results were unsatisfactory (62-70% accurate detections) in larger neighbourhoods (approximately 30 buildings), while they were more positive (80-96% accurate detections) in smaller areas (approximately 10 buildings). The study areas were located solely in urban settings, were quite restricted (neighbourhoods) and comprised very little vegetation. This algorithm may be used in rural areas, as different types of roofs and the presence of parking lots result in poor detection rates.

Stereogrammetry

This method has been widely used to manually digitize buildings. Depth perception provides the interpreter with additional information in order to differentiate the elements comprised in the image. In turn, this method is very demanding; it requires several images of a given scene taken at different angles in order to recreate the model. Furthermore, this method is essentially manual, thereby limiting the work scope. The detection error is similar (approximately 85%) to that observed with the other methods presented here (Fraser *et al.*, 2002). This approach extracts the height of the buildings while modelling them in 3D (Baillard and Maître, 1999; Chen *et al.*, 1996; Fradkin *et al.*, 2001; Fraser *et al*., 2002).

Shape and texture features

Izadi and Saeedi (2010) achieved the best results. First, their method used a hierarchical segmentation technique based on the mean shift image segmentation algorithm. Second, their method used four mathematical features of building roofs. These features reflected symmetry, the difference between a roof and its surroundings (grass, street, shadows, etc.), and the low variance in the surface of a roof. Thanks to these features, a roof probability (rooftopness) was attributed to each region for each level of the hierarchical mean shift. Lastly, the authors applied an iterative research process targeting the areas most likely to be categorized as a “roof” during the final extraction.

Sirmacek and Unsalan (2011) proposed another interesting method. They used a Gabor filter to identify the potential edges of buildings. The Gabor filter is a linear filter that is applied to the image in order to extract linear features according to the specified direction. To detect the center of buildings, they then applied an algorithm based on statistical probabilities. They applied this method to 32 IKONOS images containing 911 buildings.

Active contour algorithm

Ahmadi *et al*. (2010) developed a building extraction method based on active contours. The authors modified the active contours model developed by Chan and Vese (2001) and optimized it for automatic building extraction. Contours were initialized by manually selecting two building points and two “non-building” points. The initial contours were then generated automatically and arranged regularly on the image. The last step of their method involved generalizing the contours in order to smooth out the edges of the buildings. They succeeded in extracting 281 of the 347 buildings (80%) and the edges of 96% of the extracted buildings were accurately mapped out.

Liasis and Stavrou (2016) also optimized the algorithm by Chan and Vese (2001) to extract building edges. The initial contours were identified using the centroids of segmented areas through a k-means algorithm. The k-means is outlined in Appendix I. The algorithm created by Liasis and Stavrou (2016) was configured by studying the image’s histograms in terms of hue, saturation and value (HSV). Vegetation areas were identified and eliminated based on the intensity in the green band, while shadow areas were eliminated due to their low intensity in all bands. The centroids extracted from the remaining clusters were considered to be points on the buildings. The next step was the application of the optimized active contour algorithm. A morphological filter was then applied to eliminate small polygons and refine building contours. The authors mentioned that their algorithm sometimes considers parking lots as buildings because their shapes are similar.

Machine (deep) learning algorithm

Yuan's (2016) method used the Convolutional Network (ConvNet) to extract building footprints from HR images. ConvNets are explained in Appendix I. The author indicated that he used 2,000 tiles to train this algorithm, as well as a mask on the pixels corresponding to anything other than buildings for these tiles. Post-training, the extraction can be carried out quickly and easily on very large images. The tests were performed outside the training sites. The author also extracted buildings for five other cities in the United States without any image preparation or adjustments to its algorithm. The images used contained a multitude of spectral and spatial features, atmospheric conditions, sensors, lighting conditions and building architectures. Although only visual, the results showed that this method has real potential.

Conclusion

The review of these methods shows that building extraction from HR images has been thoroughly documented and many such methods are available. Traditional extraction methods using the spectral signature lead to less accurate results than more recent methods using shape and texture features. Technological limits are less concerning when processing large amounts of data. This gives way to new methods such as machine learning, which have achieved promising results in remote sensing applications. The main challenge in extracting buildings from HR satellite images is the strong similarity between buildings and non-buildings. More specifically, buildings are difficult to distinguish from shadow areas and parking lots, whose shape and contrast are very similar to that of buildings (Ghanea *et al.*, 2016). The literature also shows that building extraction is often processed within land cover classification. This is mentioned in three articles in particular, i.e. Myint *et al*. (2011), Li *et al.* (2016) and *Scott et al.* (2017). These three articles are presented in the section on the extraction of urban vegetation. The method proposed by Liasis and Stavrou (2016) stands out from other methods due to its simplicity, excellent results and the possibility of using only Red, Green and Blue (RGB) images. Otherwise, if you are considering the implementation of a machine learning algorithm, the method developed by Yuan (2016) which uses a ConvNet to detect and extract building footprints has achieved excellent results. The use of these two methods seems appropriate in both rural and urban areas. Table 2.1 provides an overview of the building extraction methods which use optical imagery only.

Table 2.1 - Overview of building extraction methods using optical imagery

Method	Benefits and limits	Automation	Maturity	Number of buildings	PA*	Quality**
Jin and Davis (2005)	- Multispectral - Disappointing results	- Strong - Few parameters to be configured (thresholds)	- Low - Complex process to implement - Long processing chain	121	73%	59%
San and Turker (2010)	- Rectangular and round buildings only - Multispectral - nDSM necessary	- Strong - Few parameters to be configured	- High - Fairly simple process - Processing chain tools are installed in most software	394	78-95%	67-80%
Karantzalos and Paragios (2009)	- Requires information on shapes to search for - Variable results	- Poor - Requires knowledge of the shapes of the buildings to search for	- Very low - All algorithms are to be developed	Unknown	65-98%	65-95%
Aytekin *et al*. (2011)	- Effective regardless of the shape of the building to be extracted - Multispectral - Very variable results from one scene to another	- Strong - Few parameters, simple to set	- Excellent - Processing chain tools are installed in most software - Linear process	Unknown	62-96%	34-66%
Fraser *et al.* (2002)	- Enables 3D extraction - Image pair necessary - Manual operation	- Very poor - Manual operation	- High - Processing chain tools are installed in most software - Linear process	Unknown	±85%	Unknown
Izadi and Saeedi (2010)	- Excellent results - Sufficient RGB - Requires significant development	- Excellent - No parameters	- Very low - Search tree to be developed - Roof features to be developed (100%)	140	97%	90%

Sirmacek and Unsalan (2011)	- Excellent detection rate - Sufficient RGB - Only able to detect the centroid of the buildings	- Excellent - No parameters	- High - Statistical probability algorithm to be developed only	911	94%	80%
Ahmadi *et al*. (2010)	- Sufficient RGB - Average results	- Strong - Few parameters	- High - Simple processing chain - Implementation of the active contour algorithm	347	80%	Unknown
Liasis and Stavrou (2016)	- Sufficient RGB - Excellent results	- Strong - Semi-automated method - Requires the analysis of HSV histograms	- High - Simple processing chain - Implementation of the active contour algorithm	5,063	96%	Unknown
Yuan (2016)	- Excellent success rate - Easily transposable (from one location to another) - Sufficient RGB - Requires a significant amount of training data	- Strong - Varies according to the availability of training data	- Excellent - Open access tools	1,437	89%	85%

* Producer's accuracy (PA) is defined as follows: $100 * \frac{TP}{TP+FN}$

** Quality is defined as follows: $100 * \frac{TP}{TP+FP+FN}$

Where: FN = False negatives; FP = False positives; TP = True positives

2.2 Methods combining optical imagery with another sensor

This section includes the literature review on building extraction from HR images combined with another data type. In 2016, Natural Resources Canada (NRCan) mandated the firm AECOM to conduct a study on the extraction of cartographic objects other than elevation using LiDAR data. The report documents the methods and tools used to extract buildings, hydrographic features, the road network and power transmission lines. Several excerpts from AECOM's review (2016) are included in this section, as the literature review on this topic was addressed in its report. However, only the most relevant articles in the review were retained and further details were provided. Extraction methods using imagery and another sensor only use LiDAR. For this theme, the research did not show any results using RADAR as a second sensor. The results of the methods presented in this section are outlined in Table 2.2 at the end of the section.

Methods combined with the LiDAR point cloud

Sohn and Dowman (2007) developed a method using the LiDAR point cloud and a multispectral IKONOS image. The first step of their process was the use of an algorithm to differentiate ground points and non-ground points. Vegetation points were then filtered out using the NDVI produced with the IKONOS image. By following these steps, it was possible to isolate the points belonging to buildings. To reconstruct building contours, an envelope was applied to building points. Their method also included the extraction of linear elements around convex envelopes in order to refine building delineation. The outcome of their classification was of approximately 90%. Li *et al.* (2013) developed a method that extracts building footprints using RGB images and LiDAR data. Their method consisted of four steps: the integration of the LiDAR data into the image, the removal of ground points and tree points, the coarse detection of the building's footprint, and the refinement of the footprint. The authors used an edge detection algorithm to identify the vertexes and linear segments in both types of data, so as to determine the optimal geometric transformation combining both types of data. The authors used an adaptive Triangulated Irregular Network (TIN) to filter out ground points in the LiDAR data, while they used a region growing method to remove tree points. The coarse detection of footprints was carried out using LiDAR data and an edge detection algorithm. Lastly, the refined footprints were extracted by combining the edges extracted from the optical image and the LiDAR data. This method showed excellent potential for differentiating adjacent buildings.

Methods combined with Digital Elevation Models (DEMs)

Rottensteiner *et al.* (2007) used LiDAR to create a Digital Terrain Model (DTM) and a Digital Surface Model (DSM). To extract buildings, they used two DSMs corresponding to the first and last LiDAR returns, a DTM, and the result of an NDVI extracted from a multispectral image. In order to determine the statistical probability of belonging to the building class, they used the Dempster-Shafer theory on data fusion based on these four pieces of information. The Dempster-Shafer theory classifies elements, taking into account all classes based on probabilities. The following excerpt is drawn from AECOM (2016) and completes the summary of the method: *"The authors mentioned that 95% of the buildings over 70 m^2 were detected, while those below 30 m^2 were not detected. They also mentioned that adding multispectral imagery to the algorithm improved the detection of small residential buildings by 20%."* [Translation]

Hermosilla *et al.* (2011) compared two methods combining LiDAR and HR imagery. The following summary appears in AECOM (2016): *"A threshold-based classification and an object-oriented classification were compared by Hermosilla et al. (2011). The threshold approach was developed by selecting the minimum elevation considered to be part of a building and the presence of vegetation identified in the imagery. The classification approach used a segmentation combination with a decision tree. A metric describing the shadow in relation to neighbouring objects was used to obtain the contextual relationships. Algorithms were configured differently to apply to urban, periurban and industrial areas. The threshold method was the most successful. It only uses two parameters: the minimum height of a building and the maximum reflectance in the NDVI of the imagery. The method was effective in urban and industrial areas, but the results are less compelling in suburban areas."* [Translation] The summary prepared by AECOM (2016) is fairly comprehensive. However, it is important to add that morphological filters were applied in the threshold method to smoothen the extracted buildings.

The AECOM report (2016) refers to a method developed by Awrangjeb *et al.* (2012). *"In 2012, Awrangjeb et al. presented a method using the normalized DSM [nDMS] and imagery to separate buildings from trees. It used elevation thresholds to eliminate shrubs and trees, as well as the entropy and reflectance of the imagery. A* *procedure was then developed to extract the edges. A histogram illustrating building edge orientation was then produced in order to eliminate false positives."* [Translation] However, further clarification is required. The NDVI, as well as a width and area threshold, was used to eliminate trees. This nicety implies that the use of a multispectral image is necessary. The authors also provided their method's open-access MATLAB code.

Conclusion

LiDAR is suitable for building detection through its three-dimensional component. Through the alignment and fusion of LiDAR and imagery, we can reap the benefits of both types of data and thus improve detection and representation results. The combined use of both types of data adds spectral, shape and textural information to the precise elevation of LiDAR data, thereby providing context for LiDAR points (Sohn and Dowman, 2007). One method, presented by Zhou and Qiu (2015), enables building and vegetation extraction from imagery and LiDAR data. This method is presented in the section on the extraction of urban vegetation. Table 2.2 provides an overview of the methods presented in this section.

Table 2.2 - Overview of building extraction methods using optical imagery combined with another sensor

Method	Benefits and limits	Automation	Maturity	Number of buildings	PA*	Quality**
Sohn and Dowman (2007)	- Requires a multispectral image	- Strong - Few parameters to configure	- High - Processing chain tools are installed in most software	±100	88%	80%
Li *et al.* (2013)	- Sufficient RGB imagery - Excellent in differentiating the edges of adjacent buildings - Little quantitative information on results	- Strong - Significant processing - Few parameters	- High - Algorithm chain already in place	Unknown	95%	Unknown
Rottensteiner *et al.* (2007)	- Requires a multispectral image - Buildings <30 m^2 impossible to detect	- Poor - Variable parameters from one scene to another	- High - Processing chain tools are installed in most software	4,159	87-91% for buildings >30 m^2	Unknown
Hermosilla *et al.* (2011)	- Adapted to urban, periurban and industrial scenes - Requires a multispectral image	- Excellent - Threshold method requires very little parameters	- High - Very simple processing chain - Processing chain tools are installed in most software	Unknown	86-98%	66-92%
Awrangjeb *et al.* (2012)	- Requires a multispectral image	- Strong - Few parameters	- Excellent - Code available for open access	950	95%	91%

* Producer's accuracy (PA) is defined as follows: $100 * \frac{TP}{TP+FN}$

** Quality is defined as follows: $100 * \frac{TP}{TP+FP+FN}$

Where: FN = False negatives; FP = False positives; TP = True positives

2.3 Operational applications

Building extraction from HR optical imagery can be used to assess damages during natural disasters. Dubois (2014) used automatic building extraction for damages assessment of the 2010 earthquake in Haiti. He performed building extraction on images before and after the disaster, using a method similar to the one proposed by Izadi and Saeedi (2010) before detecting changes in order to assess the earthquake's impact. The author used several types of images (Quickbird, GeoEye and IKONOS), in addition to working on reasonably large areas (containing up to 533 buildings). The work zones included dense urban areas. The Orfeo ToolBox (OTB) software was used to carry out the building extraction processes and roof features were modelled using a Support Vector Machine (SVM). An explanation of SVMs is provided in Appendix I.

Shamaoma *et al.* (2006) also extracted buildings in the case of natural disasters. Their study used the combination of LiDAR data and HR imagery to extract buildings in the case of flooding. The eCognition software was used to extract buildings using Quickbird images. The results of this extraction were merged with the classification based on LiDAR data to improve building delineation and incorporate elevation into buildings.

Two operational applications were developed using the methods cited in sections 3.1 and 3.2. The first application was developed by Yuan (2016). It is free to use, but not intended for commercial purposes. It can be downloaded using the following link: http://jiangyeyuan.com/bldgExt.html. The second application was developed by Awrangjeb *et al.* (2012) and requires both LiDAR data and multispectral images. Their method was implemented in the free open-source Barista software[1].

[1] Link to the software site: http://www.baristasoftware.com.au/

3 Hydrographic extraction

3.1 Optical imagery-based methods

Hydrographic extraction from HR optical satellite imagery is mainly based on the use of indexes using the infrared bands of multispectral images. The results of the methods described in this section are compared in summary table 3.2 at the end of the section.

Methods using only water indexes

The most popular approach due to its simplicity and processing speed is the use of water indexes. The Normalized Difference Water Index (NDWI) uses the very low reflectance of water in near infrared to differentiate water pixels (Yang *et al*., 2017). The NDWI equation is written as follows (McFeeters, 1996):

$$NDWI = \frac{Green - NIR}{Green + NIR}$$

Equation 3.1 - NDWI (McFeeters, 1996)

Xu (2006) derived a new index from the NDWI: the modified NDWI (MNDWI). Xu's variant uses mid-infrared instead of near infrared in the equation. He claims that the use of near infrared (NIR) in the extraction of water bodies leads to an overestimation of the proportion of water in the image. The author compared his algorithm with that of McFeeters in three areas of a Landsat image. The accuracy of the classification results was above 99% in the three areas, compared to 85-95% for the NDWI. The method was developed for Landsat images.

Feyisa *et al.* (2014) have used the spectral wealth of the Landsat 5 satellite to develop a water index, which is calculated using five bands of the image. Their algorithm, the Automated Water Extraction Index (AWEI), enhances the contrast between the water and the rest of the image, thereby minimizing the shadow effect. They also compared the results of their algorithm with the MNDWI on several portions of images in four regions of the world. The results showed that the AWEI performed better in each test scene, with the results varying between 93% and 98%, compared to 89-95% for the MNDWI.

In 2015, Zhai *et al.* published an article comparing the NDWI, MNDWI and AWEI with the use of Landsat-5 and Landsat-8 images. Although their study did not propose a new method, the article provided an objective assessment of the algorithms. The authors compiled the results by distinguishing urban areas from rural areas.

The results showed that the accuracy of water classification was similar for the MNDWI and the AWEI (96-97%), and that the NDWI was less successful in rural areas.

Method combining the NDWI and NDVI

Sun *et al.* (2012) developed two methods to extract information from HR images. Their first method combined the MNDWI to extract water and the NDVI to categorize the quality of the water extracted into one of four classes: turbid, clear, green and other. Their second method began with an object-oriented segmentation of the image to cluster the homogenous elements. They then applied the pixel-by-pixel approach from the first method to assess the percentage of water in each segmented region. Regions with more than 10% water were then classified as "water." They obtained the best results with the last method.

Methods using shadow removal algorithms

The NDWI, MNDWI and AEWI algorithms are quick and easy to apply. However, these indexes cannot distinguish between water bodies and the shadows created by buildings in urban areas in HR images (Yang *et al.*, 2017; Yao *et al.*, 2015; Xie *et al.*, 2016; Massalabi *et al.*, 2004). Several methods were developed to improve the detection and elimination of these shadows. These methods can be grouped into three categories: object-oriented classification (Li *et al.*, 2014; Yang *et al.*, 2017); SVM classification (Yao *et al.,* 2015); and the Morphological Shadow Index (MSI) (Xie *et al.*, 2016).

Li *et al.* (2014) proposed a two-step method for extracting water bodies from HR imagery. The first was to segment the image using spectral features by combining the NDWI with the Normalized Saturation–Value Difference Index (NSVDI). The NSVDI equation is written as follows:

$$NSVDI = \frac{(Saturation - Value)}{(Saturation + Value)}$$

Equation 3.2 - NSVDI (Li *et al*., 2014)

In the previous equation, saturation and value are HSV colour space components. They then used the shape of the elements to classify and remove shadow areas. The criteria for determining the shape of a shadow were density, angle of geometry, and symmetry. The results of their method were qualitatively compared with five other water classification methods, including the NDWI and two other SVM classification methods. Their tests were conducted using five QuickBird images. Unfortunately, the quantitative results were not disclosed in the article.

Yang *et al.* (2017) developed a method using two equations derived from the NDWI to extract potential water areas. In their first equation, they replaced the NDWI's green band with the blue band. Their second equation replaced the NDWI's green band with the values resulting from the PCA analysis. They then combined the results of their equations. Shadow detection based on an object-oriented approach was applied to the extracted areas of their equation. To distinguish shadow areas from water areas, the authors analyzed the spectral relationships between the image bands in known shadow areas and developed three spectral relationship models. If the objects identified with the equations from the first step responded to one of the three models, they were classified as such and the other objects were considered water bodies.

Yao *et al.* (2015) developed a method to eliminate shadow areas using the SVM. Contrary to the object-oriented approach, the SVM algorithm requires training, i.e. an additional implementation effort. First, they developed a new water extraction index: the High Resolution Water Index (HRWI). The idea is that the NDWI is not able to distinguish shadow and water areas for HR images. The parameters of this index were calculated using SVM classification and the resulting equation is as follows:

$$HRWI = 6 * Green - Red - 6.5 * NIR + 0.2$$

Equation 3.3 – HRWI (Yao *et al.*, 2015)

The results achieved with this equation show that it is robust and useable on several work sites. They once again used an SVM to generate a prediction model to detect shadow areas. This model is based on the two following principles: (1) the difference in value of the shadow according to the NDWI is almost equivalent to that of surrounding buildings, while the NDWI of the water is very different from that of surrounding elements (soil, vegetation); and (2) the reflectance in the green band of the water is closer to the surrounding pixels. This prediction model is used to distinguish between the shadows of open bodies of water. The extracted shadow areas serve as a mask on the HRWI results to obtain the final water bodies. This method's extraction results show a clear improvement over the NDWI. For the five images tested, the kappa statistic varies between 0.90 and 0.98. The assessment index of the kappa results is described in Appendix I.

Method adapted to the WorldView-2 sensor

Xie *et al.* (2016) wanted to leverage the potential offered by the eight spectral bands of the WorldView-2 sensor. Based on this information, they developed a new algorithm to extract surface water bodies. First, they calculate the normalized ratio of a high reflectance band (coastal, blue, green and yellow) and a low reflectance band (red-edge, PIR1 and PIR2), as well as their respective averages. They calculated each ratio possibility (5 high reflectances x 4 low reflectances = 20 possible ratios) and used the optimal ratio to determine the presence or absence of water for each pixel of the image. To remove shadow areas, the authors used an MSI developed by Huang and Zhang (2012). The MSI uses the low reflectance of shadows and their contrast with adjacent buildings to detect shadow pixels. Subtracting the MSI result from the potential water bodies identified in stage 1 eliminates the false positives caused by shadows. The results were compared with a Maximum Likelihood Classification (MLC) method and an SVM method. It was found that the SVM achieved better results (kappa between 0.70 and 0.97) and was even more successful when combined with the MSI to eliminate shadow areas. Table 3.1 summarizes the results for the three areas tested.

Table 3.1 - Aggregated results of the tests by Xie *et al.* (2016)

Method	Area 1 (kappa)	Area 2 (kappa)	Area 3 (kappa)
NDWI	0.669	0.966	0.975
NDWI-MSI	0.863	0.977	0.996
MLC	0.503	0.919	0.940
SVM	0.702	0.972	0.951
MLC-MSI	0.827	0.978	0.977
SVM-MSI	0.881	0.982	0.995

Active contour method (snake)

The extraction of rivers with a width of several pixels on the image was addressed by Dillabaugh *et al.* (2002). First, their process uses medium resolution multispectral images (±20 m) to extract rivers that appear as lines. The rivers extracted on a smaller scale serve as the starting point in the application of a snake algorithm to search for the river's contour in a higher resolution panchromatic image.

Self-organizing map (SOM) method

Zaremba (2011) presented a method using an SOM to extract hydrographic elements and generate the hydrographical network in order to obtain an adequate cartographic representation. First, he trained and used an SOM to classify the image pixel by pixel according to its spectral features. The NIR was mainly used to distinguish water. This classification resulted in a binary image of water and non-water areas. Second, a new SOM was used to create the hydrographic network based on the principal-curve procedure. The author did not mention the results achieved with his method, but he indicated that it was applied to Spot, QuickBird and Landsat-7 images. A description of the SOMs is provided in Appendix I.

Methods using the variance and intensity of a panchromatic image

The method developed by Khurshid and Khan (2012) automatically extracts rivers from panchromatic images. The method can be divided into two iterative steps. The first step uses variance to calculate an approximate mask on the contour of rivers. The second step refines the results of the previous step using the panchromatic image's intensity values. The results showed that the method extracts the contour of the river in an "efficient and detailed" manner without, however, presenting the results table. The authors used a SPOT-5 image with a resolution of 2.5 m to test their method.

Wang *et al.* (2008) also used the variance of panchromatic images to extract internal waters over 100 m^2. Their process includes two steps: extracting the water based on variance and removing false positives based on the correlation between the water classes and the rest. The tests were carried out on an image with a resolution of 1 m and measuring 6,126 X 4,800 pixels.

Conclusion

The NDWI water indexes and their derivatives were used, documented and tested on a number of occasions. Regarding images with medium spatial resolution (5-30 m), these methods can achieve excellent results when used as is. However, they cannot be applied by themselves to HR optical images, particularly in urban scenes. For this reason, several authors mentioned the use of shadow removal algorithms subsequent to the application of indexes using spectral information. These methods are simple to implement and very effective; the results presented in the various articles all exceed 90% accuracy. The method proposed by *Yao et al.* (2015) represents the most promising opportunity for development among the methods which eliminate shadow areas. The methods developed by Khurshid and Khan (2012) and by Wang *et al.* (2008) do not use multispectral information and seem very interesting and simple to implement. In both cases, the authors mentioned having achieved interesting results (over 90%). However, the details of these results were not included in the articles and the authors made no mention of the issue of shadow areas in HR images. Based on the research conducted, no article referred to the use of an Artificial Neural Network (ANN) or ConvNet to extract hydrographic features. ANNs are explained in Appendix I. Table 3.2 provides an overview of the methods presented in this section.

Table 3.2 - Overview of hydrographic extraction methods using optical imagery

Method	Benefits and limits	Automation	Maturity	UA*	PA**	Quality*** Kappa****
Xu (2006)	- Very simple pixel by pixel approach - Multispectral (mid-infrared) - Method developed for Landsat images	- Excellent - Only one threshold to be determined	- Excellent - Index is simple to implement	Unknown	Unknown	99% 0.99
Feyisa *et al.* (2014)	- Very simple pixel by pixel approach - Multispectral (mid-infrared and near infrared) - Method developed for Landsat images	- Excellent - Only one threshold to be determined	- Excellent - Index is simple to implement	95-100%	91-99%	90-99% 0.95-0.98
Sun *et al.* (2012)	- Multispectral (mid-infrared and near infrared) - Does not take into account shadows in HR imagery - Method developed for Landsat images	- Strong - A few thresholds to be determined	- Excellent - Process is simple to implement	95%	72%	99% 0.81
Li *et al.* (2014)	- Eliminates shadows - Multispectral - Lacking information on results	- Strong - Index thresholds to be determined	- Excellent - Indexes and operators on shapes which are easy to implement	Unknown	Unknown	Unknown
Yang *et al.* (2017)	- Excellent results - Eliminates shadows - Multispectral	- Poor - Spectral relationship models depend on the extraction area	- High - Fairly complex process, but consisting of algorithms which have already been developed	90-99%	88-99%	97-99% 0.87-0.96
Yao *et al.* (2015)	- Eliminates shadows - Independent water and shadow detection models - Excellent results - Multispectral	- Excellent - Parameters have already been calculated by the SVM	- High - Fairly simple process - To be completely developed	93-98%	88-98%	Unknown 0.90-0.98

Xie *et al.* (2016)	- Eliminates shadows - Independent water and shadow detection models - Excellent results - WV-2 images (8 bands)	- Poor - Requires WV-2 images and their eight bands	- High - Significant processing, although fairly simple	88-99%	86-100%	Unknown 0.86-0.99
Dillabaugh *et al.* (2002).	- Panchromatic (HR) and multispectral (BR) - Results lacking in detail	- Excellent - No parameters	- High - Implementation of the active contour only	Unknown	Unknown	Unknown
Zaremba (2011)	- Multispectral - Lacking information on results	- Strong - Seemingly few parameters	- Low - SOM to reconstruct the hydrographic network to be developed	Unknown	Unknown	Unknown
Khurshid and Khan (2012)	- Sufficient RGB imagery - Lacking information on results	- Strong - Few parameters	- High - Fairly simple processing chain	93%	94%	Unknown
Wang *et al.* (2008)	- Sufficient panchromatic image - Few details on results	- Strong - A single parameter	- Excellent - Very simple process	>90%	>90%	Unknown

* User's accuracy (UA) is defined as follows: $100 * \frac{TP}{TP+FP}$

** Producer's accuracy (PA) is defined as follows: $100 * \frac{TP}{TP+FN}$

*** Quality is defined as follows: $100 * \frac{TP}{TP+FP+FN}$

**** The definition of kappa is found in the compendium of methods.

Where: FN = False negatives; FP = False positives; TP = True positives

3.2 Methods combining optical imagery methods with another sensor

This section identifies the methods which combine HR imagery with RADAR or LiDAR. The combined use of RADAR and HR imagery is a subject of interest, but is not addressed by a significant amount of research on hydrographic extraction. There are slightly more methods using LiDAR. Table 3.3 summarizes the methods presented in this section.

Method combined with RADAR

Joshi *et al.* (2016) conducted an interesting review that identifies and compares 50 scientific articles on the fusion of optical images and RADAR in order to map land use. Only four of these articles use optical imagery with a spatial resolution over four metres. Among the four articles using HR imagery, two authors did not indicate if the fusion of both types of images had improved the quality of the results. The two remaining articles did not address the themes presented here. In addition to this review, the research did not identify methods other than the one proposed by Hong *et al.* (2015), which combines RADAR and imagery.

Hong *et al.* (2015) used a threshold method based on amplitude to determine water areas on a RADAR image. A Landsat image was classified using a maximum likelihood algorithm in seven land use classes (urban, agriculture, forest, pasture, arid land, wetland and water). This classification was used to determine the limit of the threshold method. This has considerably improved the results of RADAR image classification.

Methods combined with LiDAR

Wu *et al.* (2013) used LiDAR and aerial imagery to extract water polygons. First, their method consisted of creating a TIN using LiDAR data. Triangles of water were then distinguished from other triangles. To do this, the authors used a threshold based on the density of the points and the width of the stream to be extracted. The triangles identified in step 1 were superimposed on the aerial imagery and a mean-shift classification was applied to the covered areas on these triangles in order to refine the detection of the riverbanks. The authors indicated in their study that the method proposed may be limited by the presence of turbidity or pollution in the water. Under such conditions, the LiDAR is not completely absorbed by the water and the proposed method would not be able to distinguish the triangles of water.

Zhan *et al.* (2002) also used the capacity of the water to absorb the LiDAR signal to extract information on surface hydrography. Their method was used to extract buildings, hydrographic features and green spaces in order to map land use in an IKONOS image. LiDAR was used to extract buildings and hydrography while imagery was used to classify green spaces. The authors did not present any results regarding their classification.

Conclusion

The combination of HR optical imagery and RADAR is interesting. It would be relevant to test the method developed by Hong *et al.* (2015) on HR satellite images. Regarding the use of imagery combined with LiDAR data, the method proposed by *Wu et al.* (2013) is limited to water which is free of turbidity and pollution. Due to this limitation, this method cannot be applied at the national level. The method developed by Zhan *et al.* (2002) does not have this limitation. Table 3.3 presents an overview of the hydrographic extraction methods which combine imagery with another sensor.

Table 3.3 - Overview of hydrographic extraction methods combining optical imagery with another sensor

Method	Benefits and limits	Automation	Maturity	UA*	PA**	Quality*** Kappa****
Hong *et al.* (2015)	- Application to HR images to be tested	- Poor - Different thresholds	- Low - Complex chain	49-60%	82-83%	84-89% Unknown
Wu *et al.* (2013)	- Sufficient RGB imagery - Water turbidity or quality limits the application of this method	- Poor - Difficult to apply from one scene to another	- Low - Requires calm waters, with no turbidity	99%	92-94%	96-97% Unknown
Zhan *et al.* (2002).	- Can classify buildings, vegetation and water - Multispectral	- Strong - Few parameters	- High - Simple processing chain	Unknown	Unknown	Unknown

* User's accuracy (UA) is defined as follows: $100 * \frac{TP}{TP+FP}$

** Producer's accuracy (PA) is defined as follows: $100 * \frac{TP}{TP+FN}$

*** Quality is defined as follows: $100 * \frac{TP}{TP+FP}$

**** The definition of kappa is found in appendix I.

Where: FN = False negatives; FP = False positives; TP = True positives

3.3 Operational applications

The Computer Research Institute of Montréal (CRIM) has performed a number of studies and a handful of applications on hydrographic extraction from satellite imagery. They developed an algorithm that extracts surface hydrography from Spot and Landsat 8 images. Water bodies, originally derived from the NRCan website, are updated based on extraction. Two applications were derived from this algorithm. The first application is AutoCarto[2]. Developed for NRCan, this application extracts water bodies from SPOT images, compares the results of this extraction with the bodies of waters included in the geospatial database (GDB) and guides the operator in updating the information by indicating the areas where there has been a major change or the addition of a body of water (CRIM, 2017). The application uses an active contour algorithm based on GDB data to establish the differences and guide the update. The second application developed by the CRIM, CanGeo[3], is a Web service for the update of hydrographic data initially published on the GeoBase website (CRIM, n.d.). The updates were done using Landsat 8 images.

The research carried out did not identify any articles using a ConvNet to extract hydrographic features. Nevertheless, a student from the University of Edinburgh developed a ConvNet which extracts water bodies from satellite images and has made it available via a free licence (Reichelt, 2017). The WaterNet[4] network was developed to be used with *Python* and enables the use of Open Street Map (OSM) vector data to train the network.

---

[2] Link to an online demo of the AutoCarto application: http://geo-ti.crim.ca/
[3] CanGeo project website: http://cangeo.crim.ca/hydro
[4] WaterNet download website: https://github.com/treigerm/WaterNet

4 Road extraction

4.1 Optical imagery-based methods

The extraction of roads from HR optical satellite imagery is widely studied and a great many articles have been written on the subject. This section lists those most recently published and with the greatest potential for application. Note that the state of the art described in the article by Das *et al.* (2011) is highly relevant with respect to the methods developed prior to 2010. Although there are several others, this article forms an excellent complement to this review. Table 4.2 presents an overview of the methods described in this section.

Method based on shape features

Maboudi *et al.* (2016) propose an interesting method for extracting roads from HR imagery. The segmentation of the objects in the image is achieved using a multiresolution approach. In order to distinguish and classify roads as opposed to other objects, four characteristics were identified: roads are long structures of a defined width, road segments are quite long, the road surface is generally uniform on a local level, and their curves do not vary drastically. These features have been modelled and help to clearly distinguish road objects from other elements. Lastly, in order to fill the gaps in road detection and complete the network, tensor voting was applied prior to a thinning algorithm to extract the centre line in the roads. The article included a comparison of the results achieved with certain other methods presented in this section. Table 4.1 presents an excerpt of the results observed using each of these methods. The principle of tensor voting is described in detail in Appendix I.

Table 4.1 - Comparison of road extraction results (Maboudi *et al.*, 2016)

Test area	Method	UA (%)	PA (%)	Quality (%)
1	Wang *et al.*, 2015	70.0	75.0	74.0
	Maboudi *et al.*, 2016	93.4	89.9	84.8
2	Ameri *et al.*, 2015	94.0	89.0	84.0
	Maboudi *et al.*, 2016	95.9	93.4	89.8

Among the methods used in the comparison, those proposed by Wang *et al.* (2015) and Ameri *et al.* (2015) are described in this section.

Method using an artificial neural network (ANN)

Mokhtarzade *et al.* (2007) use an ANN to extract roads from HR images. The authors used portions of IKONOS and QuickBird images in their research. An efficient artificial neural network requires an accurate configuration of upstream parameters. The required parameters include network size and the initial parameters of the algorithm. In this case, the initial parameters are spectral information and information about the surroundings and road texture. In their research, Mokhtarzade *et al.* (2007) studied the impact on the results regarding the neural network size as well as certain other parameters such as information about roadside and road texture. They established a maximum network size as well as the initial parameter with the greatest impact on the outcome.

Method using a ConvNet and a finite-state automaton

Wang *et al.* (2015) used convolutional networks to perform road extraction. Their approach considered that roads form a connected network, i.e. a single object in the image. The ConvNet was used to detect road shapes and alignments (straight, left-turning, right-turning, intersecting and blocked). The search for roads was thus performed using a finite-state automaton and according to alignment. The algorithm was trained by combining the HR imagery (QuickBird) and road network vectors applied to an area in China. A sample of 500,000 pairs of imagery vectors was used to train the algorithm. Tests were performed on a number of images of numerous areas in China.

Method using two successive ConvNets

Convolutional networks were also used in a study by Cheng *et al.* (2017). With their method, roads were extracted as polygons and according to their centreline. The authors used two successive convolutional networks to extract their information. The first ConvNet detected the pixels of the roads, while the second was used to extract the centrelines of the roads. For this method to be successful, both networks must be trained. They used 180 images of a minimum size of 300 x 300 pixels and a resolution of 1.2 m for training. In the end, they compared the results of the extraction performed on 30 test images with five other methods (Huang and Zhang, 2009; Shi *et al.*, 2014; Cheng *et al.*, 2016; Long *et al.*, 2015; Navab *et al.*, 2015). The methods proposed by Huang and Zhang (2009) and Shi *et al.* (2014) are also presented in this section. The comparison demonstrates that the algorithm developed by Cheng *et al.* (2017) surpassed the other methods with respect to the level of detection of the road pixels and the extraction of the centrelines. The average quality of the centreline extraction results was notably better.

Methods using distributed intelligence

Maboudi *et al.* (2017) proposed a method using region-based segmentation: multiresolution segmentation. The point of interest in this study was the use of distributed intelligence, more precisely of the Ant Colony Optimization (ACO) method, to classify segmentation results. As its name indicates, ACO is based on the behaviour of ants. The principle is that the algorithm places ants on an image and each ant explores its environment in search of interesting connections. The algorithm assesses the quality of each connection in relation to the spectral, structural, and topological features previously determined by the user. The information about the ant's trajectory to the connection as well as the quality of this connection is subsequently communicated to other ants and assumes a value in the calculation of future trajectories. A strong connection and a short trajectory will influence the path taken by the next ants. Subsequent iterations will help refine the detection of the routes. Multiresolution segmentation and distributed intelligence are explained in Appendix I.

Ameri *et al.* (2015) proposed an approach similar to that of Maboudi *et al.* (2017). They used a fuzzy classification to segment and classify road areas, and used a Particle Swarm Optimization (PSO) distributed intelligence algorithm to vectorize the road network.

Methods using a support vector machine

Huang and Zhang (2009) extracted road centrelines using a method which combines multiscale segmentation using the structural and spectral information of objects along with an SVM to perform the classification. The SVM is applied to the segmentation at each scale and the detection of the centreline is performed using a thinning algorithm applied to the merged multiscale classifications. This method was further tested in the Cheng *et al.* (2017) study.

The method proposed by Shi *et al.* (2014) employed a pixel-by-pixel classification based on the image's spectral and spatial information. In order to arrive at this, they applied a General Adaptive Neighbourhood (GAN) followed by an SVM classification. As its name indicates, the GAN processing approach allows the algorithm to adapt the concept of surroundings. The relative position of surrounding elements was modified in relation to the spectral, textural or spatial features that were sought after. In this case, the classification aimed to differentiate the road pixels from the other elements. A second SVM classification was applied to the initial image in order to examine road properties and refine the differences between false positives (parking areas, bare ground, buildings) and actual roads. The results of both classification methods were then merged. A filter for element shapes eliminated elements classified as roads, but which lacked the linear characteristics of roads. To determine the position of the centrelines of roads, a thinning algorithm was applied. The study used images captured by a ZiYuan-3 satellite with a resolution of six metres. With that resolution, only major roads could be extracted. This method was also tested in the Cheng *et al.* (2017) study.

Road quality qualification method

Generally speaking, it is very difficult to access information about the attributes of extracted objects using only optical imagery. Shahi *et al.* (2016) studied the possibility of qualifying the condition of road asphalt. Their method was tested on WorldView-2 images. They used a classification algorithm based on the chi-square test, enabling them to determine the condition (good or bad) of a road. This approach is valuable in that it provides additional information regarding extracted roads. The results of their study were 83% accurate with a kappa coefficient of 0.68. The chi-square test is presented in Appendix I.

Conclusion

The articles addressing road network extraction using HR optical imagery are very recent. Methods using machine learning algorithms are widely used for this type of extraction. The major stages of road network extraction using HR images are fairly similar across the various methods. The image is classified to determine where the roads are, then the polygons are connected in order to create a road element so that a thinning algorithm can be applied to obtain the centreline of the paved surface of the roads. Furthermore, most authors succeeded in achieving extraction results which were over 90% accurate, demonstrating significant potential. The method proposed by Cheng *et al.* (2017) using two successive ConvNets appears to be highly efficient, robust, and well adapted to many types of sites. In addition, it has been compared with a number of other methods. The implementation of this method requires more resources (training data), but the results are worthwhile for nationwide extraction. This assessment method and the summary of the methods in this section can be found in Table 4.2.

Table 4.2 - Overview of road extraction methods using optical imagery

Method	Benefits and limits	Automation	Maturity	UA*	PA**	Quality*** Kappa****
Maboudi *et al.* (2016)	- Sufficient RGB imagery - Very positive outcomes - Introducing tensor voting is complex	- Strong - Easily transposable from one scene to another	- Low - Implementing tensor voting is complex	93-99%	90-97%	85-96% Unknown
Mokhtarzade *et al.* (2007)	- Sufficient RGB imagery - Outcomes based on very small areas - Small dimension ANN	- Poor - Requires a large amount of training data	- Low - ANN implementation	Unknown	Unknown	92-95% 0.61-0.72
Wang *et al.* (2015)	- Sufficient RGB imagery - Good outcomes - ConvNet training	- Poor - Requires a large amount of training data - Interaction necessary (initialization of road points)	- Low - Complex process - Implementation of ConvNet	70%	75%	74% Unknown
Cheng *et al.* (2017)	- Sufficient RGB imagery - Best results - Training data available - Complex training (2 ConvNets)	- Excellent - No interaction necessary	- Low - Complex process - Implementation of ConvNet	Approx. 95%	Approx. 96%	Approx. 92% Unknown
Maboudi *et al.* (2017)	- Multispectral - Lack of information on the implementation of the ACO	- Poor - Interactions necessary	- Low - Complex process - ACO implementation	89%	87%	78% Unknown
Ameri *et al.* (2015)	- Lack of details in road detection - Multispectral	- Strong - Semi-automated grouping of roads in clusters - Automated vectorization of centrelines	- Low - Complex process - Lack of details concerning the method	90%	91%	83% Unknown
Huang and Zhang (2009)	- Process ensuring network continuity - Multispectral	- Poor - Numerous interactions	- Excellent - Simple process - Tools in place in existing software	90-93%	78-81%	Unknown

Shi *et al.* (2014)	- Multispectral - Results widely variable depending on the sensor and the study area	- Poor - Numerous interactions	- High - Tools in place in existing software	62-98%	34-94%	29-92% Unknown

* User accuracy (UA) is defined as follows: $100 * \frac{TP}{TP+FP}$

** Producer accuracy (PA) is defined as follows: $100 * \frac{TP}{TP+FN}$

*** Quality is defined as follows: $100 * \frac{TP}{TP+FP}$

**** The definition of kappa is found in the compendium of methods.

Where: FN = False negatives; FP = False positives; TP = True positives

4.2 Methods combining optical imagery methods with another sensor

Few studies were found concerning the use of optical satellite imagery in combination with RADAR or LiDAR. The outcomes are presented in Table 4.3 at the end of the section.

Method combined with RADAR

Fard *et al.* (2014) developed a classification method for urban areas, combining synthetic aperture RADAR (SAR) with HR optical imagery. The RADAR imagery used was derived from the ALOS PALSAR sensor, while the optical imagery was taken from IKONOS. Each image was individually classified and the results of these classifications were then merged. A textural classification into three categories (city, fallow land, and road) was applied to the optical image. The authors used the product operator to merge the classifications and indicated that the merging process improved results by 4 to 8%.

Method in combination with LiDAR

Liu and Lim (2016) introduced a road network extraction method combining LiDAR data and HR imagery. Their process can be divided into five steps. First, the image and LiDAR data were merged. To do so, they assigned the RGB value of the image's corresponding pixel to each point in the LiDAR data. Second, they divided the point cloud into several files to reduce the number of points being processed at the same time. Third, the point cloud was then classified based on a set of rules which helped distinguish potential roads from other elements. These rules used the elevation, intensity and RGB value of the points and surrounding elements. Fourth, the authors used a classification system based on nearest neighbours (KNN) to distinguish roads from parking areas, using the long narrow shape of roads as a distinguishing feature. Fifth, an inverse distance weighting (IDW) interpolation was applied to eliminate blank areas resulting from trees hiding the road and complete the extraction. The centreline was identified using a curve fitting algorithm.

Conclusion

Road extraction methods using imagery combined with another sensor are limited in number. The method proposed by Fard *et al.* (2014) using RADAR is relatively easy to implement and they demonstrated that the combination of the two types of data improves the quality of the classification. On the other hand, their method only presents classification, without applying an algorithm to create the road network. Liu and Lim (2016), however, presented a method that can also extract the road network. The results and a summary of both methods are presented in Table 4.3.

Table 4.3 - Overview of road extraction methods using optical imagery in combination with another sensor

Method	Benefits and limits	Automation	Maturity	UA*	PA**	Quality*** Kappa ****
Fard *et al.* (2014)	- Excellent results - Multispectral	- Strong - Little interaction with the user	- High - Relatively simple method	Unknown	Unknown	98% 0.96
Liu and Lim (2016)	- Sufficient RGB imagery - Results to be improved	- Excellent - Able to be 100% automated	- High - Simple process	88-91%	80-89%	79-88% Unknown

* User accuracy (UA) is defined as follows: $100 * \frac{TP}{TP+FP}$

** Producer accuracy (PA) is defined as follows: $100 * \frac{TP}{TP+FN}$

*** Quality is defined as follows: $100 * \frac{TP}{TP+FP}$

**** The definition of kappa is found in the compendium of methods.

Where: FN = False negatives; FP = False positives; TP = True positives

4.3 Operational applications

A few years ago, the Université de Montréal's remote sensing laboratory was mandated to develop an application aiming to automatically detect and to update the road network on topographical maps at a scale of 1:20,000 (Bélanger, 2011). The approach adopted is set out in Bélanger's Master's thesis (2011). In general, the approach aims to modify the Sigma[0] software package for road extraction, comparing the extraction with segments from the Quebec Topographic Database (BDTQ), assessing the changes and qualifying the segments as intact, modified, or missing. The resulting application, SIGMA-ROUTES, was able to detect 78% of the existing roads in the study area. However, the application is unable to extract the roads that have been added. The software uses the panchromatic band of SPOT-5 images and the test area in Bélanger (2011) is a section of the city of Sherbrooke. The software developed cannot be used, as it was created for the provincial government.

Johnson (2016) developed an application called Deep OSM[5], which can download road data from the OSM and prepare evaluation and training data for use in a ConvNet. The ConvNet's outcome is an image with predictions of discrepancies between the OSM data and the input image. The application is available for free and unrestricted use.

[5] Link to application: https://github.com/trailbehind/DeepOSM

5 Urban vegetation extraction

5.1 Optical imagery-based methods

This section identifies the methods of extracting vegetation in an urban environment based on HR optical satellite imagery. The results and a summary of the methods can be found in Table 5.1.

Vegetation index using near-infrared

The extraction of vegetation based on multispectral images has long been studied. Indeed, Rouse *et al.* introduced the NDVI in 1974. This index is the normalized difference between the strong reflectance of vegetation in the near-infrared and its weak reflectance in the red. Even today, the NDVI is widely used to easily distinguish between vegetation and other elements (Greenhill *et al.*, 2003; Tunay *et al.*, 2007; Van Delm and Gulinck, 2011; Zhang and Feng, 2010). The use of this index is limited by the fact that it requires the near-infrared spectral band. Formula 5.1 describes the NDVI calculation:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

Equation 5.1 - NDVI (Rouse *et al.*, 1974)

Vegetation indexes using visible bands

Other indexes have been developed that use only visible bands. The first of these, the Normalized Green Red Difference Index (NGRDI), was developed by Tucker (1979) and constitutes the normalized difference between green and red. Several authors have shown that the NDVI is more effective than the NGRDI to distinguish vegetation (Campbell and Wynne, 2011; Tucker, 1979). On the other hand, Motohka *et al.* (2010) demonstrated that the index achieves results as positive as those of the NDVI.

The second index that uses visible bands is the Green Leaf Index (GLI). Gobron *et al.* (2000) developed this index, which uses the following formula:

$$GLI = \frac{2 * Green - Red - Blue}{2 * Green + Red + Blue}$$

Equation 5.2 - Green Leaf Index (Gobron *et al.*, 2000)

It has been shown that the GLI produces results that are very similar to the NDVI's results (Gobron *et al.*, 2000; Hunt *et al.*, 2011). Lastly, the Triangular Greenness Index (TGI) was developed by Hunt *et al.* (2011). The authors noted that this index is strongly influenced by chlorophyll content and that other studies using the index must be conducted before it can be used as widely as the NDVI.

Object-oriented method of ground cover classification

Vegetation extraction is often an element of ground cover classification. In this regard, Myint *et al.* (2011) developed a method of ground cover classification using an object-oriented approach. They compared their method to a pixel-by-pixel approach. Their method used a multiresolution segmentation algorithm. A nearest neighbour algorithm was used for classification. The authors stated that they used four or five training objects to classify trees and grass. The results showed an overall accuracy of 95% with a kappa coefficient of 0.94 in the classification of seven categories (building, bare ground, grass, impermeable material, swimming pool, tree and lake).

Land use classification method employing a Bayesian network

Li *et al.* (2016) developed a land use classification method based on a Bayesian network. The authors classified ground cover in 11 categories using an object-oriented approach. The categories identified are grass, tree, shade, water, bare ground, dark roof, grey roof, brick-coloured roof, blue roof, light-coloured roof, and other. The multiresolution segmentation here was performed using an algorithm implemented in the eCognition software, the parameters of which were established by trial and error. The authors indicated that they deliberately oversegmented the image in order to improve classification results. Afterwards, an SVM was trained and used to classify the segmented objects. The overall accuracy of the classification on a Pléiades satellite image was of 90% with a kappa coefficient of 0.89. Trees were accurately classified at 100% and grass at 98%. The process proposed by Li *et al.* (2016) allows for the subsequent qualification of the density and complexity of the buildings and, ultimately, for the use of the Bayesian network to extract land use.

Methods using deep learning

Scott *et al.* (2017) tested three convolutional neural networks in order to produce a classification of ground cover. The three networks tested were: CaffeNet (Jia *et al.*, 2014), GoogLeNet (Szegedy *et al.*, 2015) and ResNet (He *et al.*, 2016). The authors used previously trained algorithms for one task and applied them to extraction using satellite imagery. The method, called transfer learning, reduced training time. In order to increase the volume of training data, they applied several rotation and transposition transformations to the training data at their disposal. This technique is called data augmentation. This procedure enabled the authors to increase the amount of training data 120 times. The classification results obtained using the three algorithms were similar, i.e. around 98% accurate. The authors filed their code on GitHub[6].

Convolutional neural networks were used by Dahmane *et al.* (2016) with the aim of extracting individual cars and trees from images captured by the Pléiades satellites. As in Scott *et al.* (2017), the authors used a pre-trained convolutional algorithm, CaffeNet. They used 6,307 areas of 1,024 m^2 from a region in Vancouver to train the CaffeNet network. They then tested their algorithm on an area in the city of Québec. Their methodology was not highly detailed, but they mentioned that they obtained better results when they combined R, V and PIR bands. The probability that individual trees will be accurately detected is greater than 0.99.

Vegetation types classification method using a decision tree

Zhang *et al.* (2010) developed an object-oriented method of classifying vegetation types among five categories: aquatic, deciduous, coniferous, artificial turf and grass. Their method consisted of four steps. During the first step, shaded areas were eliminated using a hierarchical method based on spectral and textural information. The details of this step were not outlined in the article. Next, two segmentation methods were applied one after the other. They first applied the NDVI to view the pixelated vegetation. The second segmentation used the region growing method to obtain objects representing vegetation. Then they calculated a set of spectral, textural and shape-related features for each of the segmented areas of vegetation, before proceeding with a principal component analysis in order to reduce the size of the objects while retaining the relevant information. Finally, the authors used a decision tree (CART) to classify the vegetation in five categories. The overall accuracy of their classification was of 89% with a kappa coefficient of 0.85.

Conclusion

The NDVI is the index most used to detect vegetation in a multispectral image. It is a reliable, easy-to-use and efficient index. The use of indexes employing visible bands is less common. In each case, the thresholds to be

---

[6] https://github.com/scottgs/ML_DataAugmentation

used vary from one scene to the next. Ground cover classification methods can be used to accurately classify vegetation and extract other types of information. However, these methods are more difficult to automate. Methods using deep learning achieve nearly perfect results in vegetation classification and rarely require any multispectral images. In this regard, the approach proposed by Dahmane *et al.* (2016) merits further study. Finally, the vegetation type classification method proposed by Zhang *et al.* (2010) represents an interesting solution for attributing features to the vegetation extracted. Table 5.1 presents an overview of these methods and their results.

Table 5.1 - Overview of urban vegetation extraction methods using optical imagery

Method	Benefits and limits	Automation	Maturity	Quality* Kappa**
Rouse *et al.* (1974)	- Widely used - Multispectral	- Excellent - Ratio of spectral bands	- Excellent - Very easy to implement	Results vary depending on the threshold used
Tucker (1979)	- Sufficient RGB imagery	- Excellent - Ratio of spectral bands	- Excellent - Very easy to implement	Results vary depending on the threshold used
Gobron *et al.* (2000)	- Sufficient RGB imagery	- Excellent - Ratio of spectral bands	- Excellent - Very easy to implement	Results vary depending on the threshold used
Hunt *et al.* (2011)	- Sufficient RGB imagery - Index performance yet to be proven	- Excellent - Ratio of spectral bands	- Excellent - Very easy to implement	Results vary depending on the threshold used
Myint *et al.* (2011)	- Can classify several types of information - Excellent results - Multispectral	- Poor - Supervised classification	- High - Easy to implement	95% 0.94
Li *et al.* (2016)	- Can classify several types of information - Excellent results - Multispectral - Lack of detail concerning the method	- Poor - Difficult to apply from one scene to another	- High - Easy to implement	83% 0.80
Scott *et al.* (2017)	- Uses a previously trained network - Excellent results - Can classify any element, depending on training data - Requires a large amount of data for training	- Strong - Varies depending on training data	- Excellent - Some development required to adapt the network	98-99% Unknown
Dahmane *et al.* (2016)	- Uses a previously trained network - Excellent results - Can detect individual trees	- Strong - Varies depending on training data	- Excellent - Some development required to adapt the network	Approx. 99% Unknown

	- Requires a large amount of data for training			
Zhang *et al.* (2010)	- Distinguishes the different types of vegetation - Multispectral	- Strong - Few parameters - Applicable to different scenes	- High - Simple process - Lacking details about the elimination of shadows	89% 0.85

* Quality is defined as follows: $100 * \frac{TP}{TP+FP+FN}$

** The definition of kappa can be found in the compendium of methods.

Where: FN = False negatives; FP = False positives; TP = True positives

5.2 Methods combining optical imagery with another sensor

This review was unable to identify methods of extracting cartographic information using satellite imagery in combination with RADAR. The following methods use a combination of satellite imagery and LiDAR. The results along with a summary of these methods can be found in Table 5.2.

Pseudo-waveform classification method using LiDAR data

In 2015, Zhou and Qiu developed a method combining the use of WorldView-2 multispectral images and LiDAR data to classify buildings, grass, pavement, shadows and trees. Image segmentation is performed using the ENVI Zoom software. The authors then applied an initial object-oriented classification to the segmented image using spectral histograms for the five categories to be extracted. They simultaneously performed another classification with the segmented image and LiDAR data. The principle was that, for each segmented area of the image, the LiDAR points located in those areas were extracted and a histogram of the points was then calculated based on height, in order to differentiate the heights of the various elements, thus refining the segmentation. This classification method is called pseudo-waveform. The results of both classifications were then merged. The results of the final classification using this method were 97% accurate with a kappa coefficient of 0.97.

Classification of types of vegetation using active learning

Rougier *et al.* (2016) compared the performance of three active learning algorithms for classifying urban vegetation according to two categories: trees and herbaceous plants. Their method used Pléiades satellite images and a normalized DSM. The segmentation of the images was performed by combining a multiresolution approach and an approach based on the spectral difference. Next, for each area, the authors calculated a host of spectral, textural and geometric information (band average, standard deviation of pixel values, minimum and maximum, co-occurrence matrix, area, asymmetry, etc.). They then performed an initial training of the active learning algorithm using stratified random sampling. Finally, they compared three active learning strategies to assess their respective performance. The results achieved with the three strategies were similar. Furthermore, trees were much more accurately classified (approximately 80%) than herbaceous plants (50%). Active learning is explained in Appendix I.

Conclusion

In order to extract buildings and vegetation, the method proposed by Zhou and Qiu (2015) appears promising. The potential for automation is very good and the process is simple to implement. The results show that the algorithm leverages the spectral advantage of the optical sensor and the three-dimensional advantage of the LiDAR data in order to accurately classify the information. The approach proposed by Rougier *et al.* (2016) can be used to classify types of vegetation, but will require further development. Table 5.2 presents an overview of both methods discussed in this section.

Table 5.2 - Overview of urban vegetation extraction methods combining optical imagery with another sensor

Method	Benefits and limits	Automation	Maturity	Quality* Kappa**
Zhou and Qiu (2015)	- Can classify several types of information - Multispectral	- Excellent	- High - Easy to implement	97% 0.97
Rougier *et al.* (2016)	- Distinguishes different types of vegetation - Multispectral - Average results (50% for herbaceous plants)	- Poor - Numerous parameters	- High - Process comprises many steps, but is fairly simple	Unknown

* Quality is defined as follows: $100 * \frac{TP}{TP+FP+FN}$

** The definition of kappa can be found in the compendium of methods.

Where: FN = False negatives; FP = False positives; TP = True positives

5.3 Operational applications

Pham *et al.* (2011) proposed an application for extracting vegetation in an urban environment. Their goal was to map out Montréal's vegetation cover in order to study its relationship with the demographics of the various neighbourhoods. The vegetation cover was extracted from two QuickBird images acquired in 2007. The authors indicated that they executed the segmentation in the first stage, but did not specify which algorithm they used. The types of vegetation were classified by applying the modified soil adjusted vegetation index (MSAVI) at two different scales to objects which had been previously segmented. The overall accuracy of the classification of the two categories (trees and grass) was of 73% with a kappa coefficient of 0.50. The authors next studied the relationship between the vegetation cover and the demographics of Montréal's various neighbourhoods. The authors indicated that they used the eCognition 8.1 software to perform the classification.

6 Conclusions regarding extraction methods

The extraction of cartographic information from HR optical satellite imagery is widely studied and the numerous extraction methods are continuously evolving. The current state of extraction methods shows that pixel-by-pixel and object-oriented classification strategies are still widely used today. Among these strategies, the method that exhibits the greatest potential for extracting buildings is that developed by Liasis and Stavrou (2016), which uses RGB images and a snake algorithm. Concerning hydrography, development efforts should be based on two approaches: the method proposed by Wang *et al.* (2008), which extracts moisture regions from panchromatic images; and the method proposed by Yao *et al.* (2015), which uses spectral information along with an algorithm that eliminates shadows. The optimal object-oriented approach for extracting roads and the road network is the method proposed by Maboudi *et al.* (2016), which uses tensor voting. Regarding vegetation, the most promising object-oriented approach is that developed by Zhang *et al.* (2010), thanks to which various types of vegetation can be classified.

The current trend in cartographic information extraction from high resolution imagery is the use of machine learning and deep learning algorithms. As computer capabilities evolve, so do developments and advances in these fields. The machine learning method developed by Yuan (2016) shows a promising potential for extracting the shapes of buildings. Though no articles have been published to date on the extraction of hydrographic elements using machine or deep learning, the application developed by Reichelt (2017) should be tested further in order to assess its results. Many of the methods using machine learning were created with roads and road network extraction in mind. Development efforts should be undertaken with respect to this subject in order to reproduce the method proposed by Cheng *et al.* (2017), which uses two successive ConvNets. Finally, development efforts should focus on the method proposed by Dahmane *et al.* (2016), which uses deep learning in order to extract individual trees.

Among the methods combining optical imagery with other sensors, the method proposed by Zhou and Qiu (2015) employs LiDAR to improve imagery-based segmentation and offers an interesting technique which leverages the benefits of each sensor. In addition, this method can be used to extract water, buildings and vegetation.

Studying these methods uncovered the limitations of the extraction process. Among the problems examined by the scientific literature is the difficulty of distinguishing certain objects that share similar features (spectral, spatial, and geometric), but the nature of which is completely different (Shan and Hussain, 2011). For instance, it is difficult to distinguish between a roof and a parking area, or between a body of water and the shadow cast by high-rise buildings. Another limit of the extraction of cartographic information from optical imagery is the difficulty in qualifying the extracted objects. The type of building, road and road cover are much harder to identify than the mere shape of the building or road. The features attributed to the extracted elements are quite limited. Collaborative cartography could provide an acceptable solution to this problem. The main limitation on the use of machine learning algorithms is the need for large amounts of a priori data.

The methods examined in the present study were applied to areas of relatively limited size. On a nationwide scale, the elements to be extracted vary from one end of the country to the other. These variations present an additional challenge to the nation-wide implementation of an extraction process. In this respect, the process used must be robust and adapted to work area. It is recommended that several deep learning algorithms be trained on sub-areas of the country.

7 Bibliographie

AECOM. 2016. *Étude sur le potentiel des données LiDAR pour l'extraction des objets cartographiques autres que l'élévation. Méthodes, outils d'extraction des bâtiments, du réseau hydrographique, des infrastructures de transport et des lignes de transport d'énergie. Rapport présenté à Ressources naturelles Canada.*

Ahmadi, Salman, MJ Valadan Zoej, Hamid Ebadi, Hamid Abrishami Moghaddam, and Ali Mohammadzadeh. 2010. Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *International Journal of Applied Earth Observation and Geoinformation* 12 (3): 150-7.

Ameri, Fateme, and Mohammad J. Valadan Zoej. 2015. Road vectorisation from high-resolution imagery based on dynamic clustering using particle swarm optimisation. *The Photogrammetric Record* 30 (152): 363-86.

Awrangjeb, Mohammad, Chunsun Zhang, and Clive S. Fraser. 2012. Building detection in complex scenes thorough effective separation of buildings from trees. *Photogrammetric Engineering & Remote Sensing* 78 (7): 729-45.

Aytekin, Ö., A. Erener, I. Ulusoy, and H. S. B. Düzgün. 2011. Automatic and unsupervised building extraction in complex urban environments from multi spectral satellite imagery.

Baillard, C., and H. Maître. 1999. 3D reconstruction of urban scenes from aerial stereo imagery: A focusing strategy. *Computer Vision and Image Understanding* 76 (3): 244-58.

Bélanger, Jean. 2011. Mise à jour de la base de données topographiques du Québec à l'aide d'images à très haute résolution spatiale et du progiciel Sigma0 : Le cas des voies de communication. M. sc. en Géographie., Université de Montréal.

Campbell, James B., and Randolph H. Wynne. 2011. *Introduction to remote sensing*. 5th ed. New York: Guildford Press.

Caselles, Vicent, and Pascal Monasse. 2010. Geometric description of images as topographic maps.

Chan, Tony F., and Luminita A. Vese. 2001. Active contours without edges. *IEEE Transactions on Image Processing* 10 (2): 266-77.

Chen, Sei-Wang, George C. Stockman, and Kuo-En Chang. 1996. SO dynamic deformation for building of 3D models. *IEEE Transactions on Neural Networks* 7 (2): 374-87.

Cheng, Guangliang, Ying Wang, Shibiao Xu, Hongzhen Wang, Shiming Xiang, and Chunhong Pan. 2017. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*.

Cheng, Guangliang, Feiyun Zhu, Shiming Xiang, Ying Wang, and Chunhong Pan. 2016. Accurate urban road centerline extraction from VHR imagery via multiscale segmentation and tensor voting. *Neurocomputing* 205 : 407-20.

Contassot-Vivier, Sylvain, Lo Bosco, Giosué, and Nguyen, Chanh D. 1996. *Multiresolution approach for image processing.* Pays-Bas.

CRIM. CanGeo project. n.d. [cité 06/12 2017]. Available from http://cangeo.crim.ca/hydro.

CRIM. Autocarto, outil pour la détection automatique des contours hydrographiques (ressources naturelles canada). 2017 [cité 06/12 2017]. Available from http://www.crim.ca/en/achievements/autocarto-a-tool-for-the-automatic-detection-of-hydrographic-contours-form-satellite-images.

Dahmane, Mohamed, Samuel Foucher, M. Beaulieu, F. Riendeau, Yacine Bouroubi, and Mathieu Benoit. 2016. Object detection in Pleiades images using deep features. Paper presented at Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International.

Das, Sukhendu, TT Mirnalinee, and Koshy Varghese. 2011. Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing* 49 (10): 3906-31.

Dillabaugh, Craig R., K. Olaf Niemann, and Dianne E. Richardson. 2002. Semi-automated extraction of rivers from digital imagery. *Geoinformatica* 6 (3) : 263-84.

Dubois, David. 2014. *Étude du profil d'échelle des formes et de mesures d'énergie de texture pour l'évaluation semi-automatique des dégâts sur les bâtiments dans les images satellitaires de très haute résolution*. Doctorat., École de technologie supérieure.

Fard, T. Alipour, M. Hasanlou, and H. Arefi. 2014. Classifier fusion of high-resolution optical and synthetic aperture radar (SAR) satellite imagery for classification in urban area. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (2): 25.

Feyisa, Gudina L., Henrik Meilby, Rasmus Fensholt, and Simon R. Proud. 2014. Automated water extraction index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment* 140 : 23-35.

Flouzat, Guy, Olivier Amram, Florence Laporterie, and Selma Cherchali. 2001. Multiresolution analysis and reconstruction by a morphological pyramid in the remote sensing of terrestrial surfaces. *Signal Processing* 81 (10): 2171-85.

Fradkin, M., H. Maître, and M. Roux. 2001. Building detection from multiple aerial images in dense urban areas. *Computer Vision and Image Understanding* 82 (3): 181-207.

Fraser, C. S., E. Baltsavias, and A. Gruen. 2002. Processing of IKONOS imagery for submetre 3D positioning and building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 56 (3) (4): 177-94.

Ghanea, Mohsen, Payman Moallem, and Mehdi Momeni. 2016. Building extraction from high-resolution satellite images in urban areas: Recent methods and strategies against significant challenges. *International Journal of Remote Sensing* 37 (21): 5234-48.

Gobron, Nadine, Bernard Pinty, Michel M. Verstraete, and J-L Widlowski. 2000. Advanced vegetation indices optimized for upcoming sensors: Design, performance, and applications. *IEEE Transactions on Geoscience and Remote Sensing* 38 (6): 2489-505.

Gonzalez, Rafael C., and Richard E. Woods. 2008. *Digital image processing*. 3rd ed. Upper Saddle River, N.J: Pearson Prentice Hall.

Greenhill, Darrel R., Lennart T. Ripke, Adrian P. Hitchman, Graeme A. Jones, and Graeme G. Wilkinson. 2003. Characterization of suburban areas for land use planning using landscape ecological indicators derived from ikonos-2 multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 41 (9): 2015-21.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition.

Hermosilla, Txomin, Luis A. Ruiz, Jorge A. Recio, and Javier Estornell. 2011. Evaluation of automatic building detection approaches combining high resolution images and LiDAR data. *Remote Sensing* 3 (6): 1188-210.

Hong, Seunghwan, Hyoseon Jang, Namhoon Kim, and Hong-Gyoo Sohn. 2015. Water area extraction using RADARSAT SAR imagery combined with landsat imagery and terrain information. *Sensors* 15 (3): 6652-67.

Huang, Xin, and Liangpei Zhang. 2012. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (1): 161-72.

Huang, Xin, and Liangpei Zhang. 2009. Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines. *International Journal of Remote Sensing* 30 (8): 1977-87.

Hunt, E. Raymond, CST Daughtry, Jan UH Eitel, and Dan S. Long. 2011. Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal* 103 (4): 1090-9.

Izadi, Mohammad, and Parvaneh Saeedi. 2010. Automatic building detection in aerial images using a hierarchical feature based image segmentation.

Jia, Yangqing, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. Paper presented at Proceedings of the 22nd ACM international conference on Multimedia.

Jiang, N., J. X. Zhang, H. T. Li, and X. G. Lin. 2008. Semi-automatic building extraction from high resolution imagery based on segmentation.

Jin, Xiaoying, and Curt H. Davis. 2005. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *Eurasip Journal on Applied Signal Processing* 2005 (14): 2196-206.

Johnson, Andrew L. Deeposm. 2016 [cited 09/25 2017]. Available from https://github.com/trailbehind/DeepOSM.

Joshi, Neha, Matthias Baumann, Andrea Ehammer, Rasmus Fensholt, Kenneth Grogan, Patrick Hostert, Martin Rudbeck Jepsen, *et al.* 2016. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing* 8 (1).

Jung, Cláudio Rosito, and Rodrigo Schramm. 2004. Rectangle detection based on a windowed hough transform.

Karantzalos, Konstantinos, and Nikos Paragios. 2009. Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. *IEEE Transactions on Geoscience and Remote Sensing* 47 (1): 133-44.

Khurshid, M. Hasnat, and Muhammad Faisal Khan. 2012. River extraction from high resolution satellite images. Paper presented at Image and Signal Processing (CISP), 2012 5th International Congress on.

Li, Bangyu, Hui Zhang, and Fanjiang Xu. 2014. Water extraction in high resolution remote sensing image based on hierarchical spectrum and shape features. Paper presented at IOP Conference Series: Earth and Environmental Science.

Li, Hui, Cheng Zhong, Xiaoguang Hu, Long Xiao, and Xianfeng Huang. 2013. New methodologies for precise building boundary extraction from LiDAR data and high resolution image. *Sensor Review* 33 (2): 157-65.

Li, Mengmeng, Alfred Stein, Wietske Bijker, and Qingming Zhan. 2016. Urban land use extraction from very high resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing* 122 : 192-205.

Liasis, Gregoris, and Stavros Stavrou. 2016. Building extraction in satellite images using active contours and colour features. *International Journal of Remote Sensing* 37 (5): 1127-53.

Liu, Li, and Samsung Lim. 2016. A framework of road extraction from airborne lidar data and aerial imagery. *Journal of Spatial Science* 61 (2): 263-81.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation.

Maboudi, Mehdi, Jalal Amini, Michael Hahn, and Mehdi Saati. 2017. Object-based road extraction from satellite images using ant colony optimization. *International Journal of Remote Sensing* 38 (1): 179-98.

Maboudi, Mehdi, Jalal Amini, Michael Hahn, and Mehdi Saati. 2016. Road network extraction from VHR satellite images using context aware object feature integration and tensor voting. *Remote Sensing* 8 (8).

Massalabi, Amani, Dong-Chen He, G. B. Bénié, and Eric Beaudry. 2004. Detecting information under and from shadow in panchromatic IKONOS images of the city of sherbrooke.

McFEETERS, S. K. 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* 17 (7) (05/01): 1425-32.

Mokhtarzade, M., and M. J. Valadan Zoej. 2007. Road detection from high-resolution satellite images using artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation* 9 (1): 32-40.

Motohka, Takeshi, Kenlo Nishida Nasahara, Hiroyuki Oguma, and Satoshi Tsuchida. 2010. Applicability of green-red vegetation index for remote sensing of vegetation phenology. *Remote Sensing* 2 (10): 2369-87.

Myint, Soe W., Patricia Gober, Anthony Brazel, Susanne Grossman-Clarke, and Qihao Weng. 2011. *Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery*. Vol. 115.

Navab, Nassir, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi. 2015. Medical image computing and computer-assisted intervention – MICCAI 2015: 18th international conference munich, germany, october 5-9, 2015 proceedings, part II.

Navulur, Kumar. 2007. *Multispectral image analysis using the object-oriented paradigm*. Boca Raton: CRC Press/Taylor & Francis.

Pham, Thi-Thanh-Hiên, Philippe Apparicio, Anne-Marie Séguin, and Martin Gagnon. Mapping the greenscape and environmental equity in Montreal: An application of remote sensing and GIS.

Reichelt, Tim. Waternet. 2017 [cited 08/22 2017]. Available from https://github.com/treigerm/WaterNet.

Rottensteiner, Franz, John Trinder, Simon Clode, and Kurt Kubik. 2007. Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (2): 135-49.

Rougier, Simon, Anne Puissant, André Stumpf, and Nicolas Lachiche. 2016. Comparison of sampling strategies for object-based classification of urban vegetation from very high resolution satellite images. *International Journal of Applied Earth Observation and Geoinformation* 51 : 60-73.

Rouse Jr, J_W, RH Haas, JA Schell, and DW Deering. 1974. Monitoring vegetation systems in the great plains with ERTS.

San, D. Koc, and M. Turker. 2010. Building extraction from high resolution satellite images using hough transform.

Scott, Grant J., Matthew R. England, William A. Starms, Richard A. Marcum, and Curt H. Davis. 2017. Training deep convolutional neural networks for Land–Cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters* 14 (4): 549-53.

Shahi, Kaveh, Helmi Zulhaidi Mohd Shafri, and Alireza Hamedianfar. 2016. Road condition assessment by OBIA and feature selection techniques using very high-resolution WorldView-2 imagery. *Geocarto International*: 1-18.

Shamaoma, H., N. Kerle, and D. Alkema. 2006. Extraction of flood-modelling related base-data from multi-source remote sensing imagery. Paper presented at ISPRS mid-term symposium.

Shi, Wenzhong, Zelang Miao, and Johan Debayle. 2014. An integrated method for urban main-road centerline extraction from optical remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing* 52 (6): 3359-72.

Sirmacek, Beril, and Cem Unsalan. 2011. A probabilistic framework to detect buildings in aerial and satellite images. *IEEE Transactions on Geoscience and Remote Sensing* 49 (1): 211-21.

Sohn, Gunho, and Ian Dowman. 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (1): 43-63.

Sun, Fangdi, Wanxiao Sun, Jin Chen, and Peng Gong. 2012. Comparison and improvement of methods for identifying waterbodies in remotely sensed imagery. *International Journal of Remote Sensing* 33 (21): 6854-75.

Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition.

Tucker, Compton J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8 (2): 127-50.

Tunay, M., MA Marangoz, S. Karakis, and A. Atesoglu. 2007. Detecting urban vegetation from different images using an object-based approach in Bartin, Turkey. Paper presented at Recent Advances in Space Technologies, 2007. RAST'07.

Van Delm, An, and Hubert Gulinck. 2011. Classification and quantification of green in the expanding urban and semi-urban complex: Application of detailed field data and IKONOS-imagery. *Ecological Indicators* 11 (1): 52-60.

Wang, Hua, Li Pan, and Hong Zheng. 2008. Multi-texture-model for water extraction based on remote sensing image.

Wang, Jun, Jingwei Song, Mingquan Chen, and Zhi Yang. 2015. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *International Journal of Remote Sensing* 36 (12): 3144-69.

Wu, Hangbin, Chun Liu, Yunling Zhang, Weiwei Sun, and Weiyue Li. 2013. Building a water feature extraction model by integrating aerial image and lidar point clouds. *International Journal of Remote Sensing* 34 (21): 7691-705.

Xie, Cong, Xin Huang, Wenxian Zeng, and Xing Fang. 2016. A novel water index for urban high-resolution 8-band WorldView-2 imagery. *International Journal of Digital Earth* 9 (10): 925-41.

Xu, Hanqiu. 2006. Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* 27 (14): 3025-33.

Yang, Fan, Jianhua Guo, Hai Tan, and Jingxue Wang. 2017. Automated extraction of urban water bodies from ZY-3 Multi-Spectral imagery. *Water* 9 (2): 144.

Yao, Fangfang, Chao Wang, Di Dong, Jiancheng Luo, Zhanfeng Shen, and Kehan Yang. 2015. High-resolution mapping of urban surface water using ZY-3 multi-spectral imagery. *Remote Sensing* 7 (9): 12336-55.

Yuan, Jiangye. 2016. Automatic building extraction in aerial scenes using convolutional networks. *ArXiv Preprint arXiv:1602.06564*.

Zaremba, Marek B. 2011. Automated mapping of hydrographic systems from satellite imagery using self-organizing maps and principal curves. In *Self-organizing maps-applications and novel algorithm design.*InTech.

Zhai, Ke, Xiaoqing Wu, Yuanwei Qin, and Peipei Du. 2015. Comparison of surface water extraction performances of different classic water indices using OLI and TM imagery in different situations. *Geo-Spatial Information Science* 18 (1): 32-42.

Zhan, Qingming, Martien Molenaar, and Klaus Tempfli. 2002. Hierarchical image object-based structural analysis toward urban land use classification using high-resolution imagery and airborne LIDAR data. Paper presented at Proceedings of the 3rd international conference on remote sensing of urban areas, Istanbul, Turkey.

Zhang, Xiuying, Xuezhi Feng, and Hong Jiang. 2010. Object-oriented method for urban vegetation mapping using IKONOS imagery. *International Journal of Remote Sensing* 31 (1): 177-96.

Zhou, Yuhong, and Fang Qiu. 2015. Fusion of high spatial resolution WorldView-2 imagery and LiDAR pseudo-waveform for object-based image analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 101 : 221-32.

Appendix I – Compendium of methods

This section explains concepts used in image processing. It contains information about recurring concepts in the literature as well as some machine learning algorithms.

General concepts

- **Multiresolution segmentation**

Segmentation is the first stage in an object-oriented classification. It is used to create homogenous areas comprising similar elements in the image (Shahi *et al.*, 2016). Multiresolution segmentation is a widely used method of segmentation and is implemented in most image-processing tools and software. The parameters of this algorithm are the shape, compaction and scale. The parameter of shape determines whether the focus is on the shape of the object as compared to spectral homogeneity. Compaction controls the choice between compact and smooth respecting the contours of the objects. Finally, the parameter of scale is used to determine the size of the segmented objects. Multiple scales will generate large homogenous objects, while a small scale value will produce small groupings (Myint *et al.*, 2011).

- **Mean Shift Segmentation**

The mean shift uses the pixels' surroundings to group together elements comprising similar values (Dubois, 2014). The shape and number of groupings are not established by the user when the algorithm is executed. The general operating principle comprises three stages. First, a window is placed around one pixel of the image. The size of the window is specified by the user. Next, the algorithm calculates the average of the pixels in the window. Finally, the window is moved towards the average and the algorithm starts over at step one until the average and the points of the group converge (Thirumuruganathan, 2010).

- **Tensor Voting**

During the extraction, it often occurs that the information extracted is incomplete and noise is present in the results. Tensor voting is used to complete the extraction. The goals of the algorithm are as follows: identify the anomalous values (noise) in order to eliminate them, estimate the size of the geometrical structures described by the data, estimate the orientation of the geometrical structures, and enable an accurate reconstruction of these structures. The mathematical basis of this algorithm is relatively complex, but its application is quite simple as only the parameter of scale is controlled by the user (Milkers, 2015; Medioni *et al.*, 2000).

- **Chi-square**

Chi-square is a statistical test which helps determine whether the difference between two distributions is statistically significant or simply due to chance. The chi-square formula is as follows:

$$\chi_c^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

Equation 1 - Chi-square (adapted from Hopkins, 2017)

- **Kappa**

The kappa value is used in evaluating classification results. This calculation adjusts the overall accuracy of the classification in order to eliminate the "chance" factor in classification (Campbell and Wynne, 2011). The kappa equation is expressed as follows:

$$\kappa = \frac{Observed - Expected}{1 - Expected}$$

Equation 2 - kappa (adapted from Campbell and Wynne, 2011)

The value of kappa varies from -1 to 1, where -1 indicates that the classification produces results worse than a random classification, a value of 0 indicates that the classification is not better than a random classification, and 1 indicates that the classification is significantly better than a random classification (Campbell and Wynne, 2011).

- **K-Means classification**

The k-means classification algorithm includes, iteratively, the datasets that share similarities. AECOM (2016) proposes the following definition: "*K-means partitioning is a data partitioning method and a combinatorial optimization issue. Considering the points and whole K, the issue lies in dividing the points into K partitions, often referred to as clusters, in order to minimize a specific function. The distance separating a point from the* *average of the points in its cluster is considered; the function to be minimized is the sum of the squares of these distances."* [Translation]

Machine learning

Machine learning methods have existed since the 1980s, but their application requires a significant calculating capacity that has only become available recently. A number of authors addressed the application of these methods to extract information from HR satellite imagery (LeCun *et al.*, 2015).

- **Artificial neural network (ANN)**

An artificial neural network comprises a training (or learning) phase and an execution phase. During the training phase, the algorithm scans the images, knowing precisely what it is seeking (Mokhtarzade *et al.*, 2007). These search parameters are called neurons. With each iteration, the algorithm adjusts and reinforces the neurons, as well as the connections between them, in order to reduce detection errors. For instance, in order to extract roads from the imagery, the algorithm must be provided with images in which all elements other than roads are masked. The algorithm will therefore scan the image in order to confirm, validate, and adjust the parameters of the search. These algorithms tend to mimic human learning through reinforcement and repeated associations (Campbell and Wynne, 2011). The size of the network (number of neurons) is determined by the user. A lack of neurons can limit the algorithm in its ability to recognize important patterns, while an excessive number of neurons can fragment important patterns, causing them to be indistinguishable (Zaremba, 2011). In order to achieve accurate results, an algorithm must include many iterations with a significant amount of training data (LeCun *et al.*, 2015). Finally, the execution phase consists of applying the trained algorithm to the study area.

- **Self-Organizing Map (SOM)**

SOMs are a type of artificial neural network which are trained in an automated and unsupervised manner. This means that the connections are determined by the algorithm itself. As in a neural network, the information provided is fully connected to the layers of neurons, so that with each iteration the connections between neurons are adjusted and strengthened by the algorithm (Gao, 2009). The training aspect is achieved through a competitive process between neurons. When encountering certain patterns, certain neurons react more strongly than others. With each iteration, the neuron showing the strongest reaction becomes the “winning” neuron, and its connections are reinforced within the network (Gao, 2009).

- **Convolutional neural network**

Traditionally, the convolutional neural network (ConvNet) is used in face recognition and with artificial intelligence. It comprises a series of operations: convolutions, non-linear operations (ReLU) and pooling (Cheng *et al.,* 2017; Karn, 2016). Convolution operations are simple linear operations such as filters (Gaussian, contour detection, normalized, etc.). Non-linear (polynomial) operations are applied pixel-by-pixel and increase the model’s complexity by introducing non-linearity (Long *et al.,* 2015). Spatial pooling helps reduce or increase the size of the image while preserving as much information as possible. Finally, an artificial neural network using results from previous operations enables the ConvNet to classify the elements appearing in the image (Karn, 2016). Just like an artificial neural network, a ConvNet comprises a training phase and an execution phase. Learning takes place as the convolutional filters are adjusted so as to reduce the difference between the predicted and real values of the training data (Muruganandham, 2016). The ConvNet includes fewer parameters than the artificial neural network, due to the local connectivity between the layers (Cheng *et al*., 2017).

- **Swarm Intelligence**

The concept of distributed intelligence algorithms is derived from certain animal species. They attempt to represent collective interactions and behaviours found in nature, for instance among bees and ants. In general terms, the algorithm distributes the members of the colony on the image, and their movements and interactions simulate the behaviour of the individual members of the swarm (Maboudi *et al.*, 2017).

- **Support vector machine (SVM)**

This method of supervised learning is frequently used in classification. SVMs are used to address discrimination and regression issues (AECOM, 2016). In order to separate the categories, the SVM seeks the optimal linear expression to distinguish among the categories. Insofar as the data cannot be separated linearly, the algorithm will transform the information into a (possibly infinite) dimensional space X in which the data can be separated linearly (Raschka, 2016).

- **Active Learning**

Active learning is a form of semi-supervised machine learning. Algorithms in this category refine their learning by actively questioning the user (Rougier *et al.*, 2016). Various strategies may be used to question the user. Among these, the following two should be noted: the model is unsure of the classification, or the results of the classification performed by several models differ.

Bibliographie du compendium

Campbell, James B., and Randolph H. Wynne. 2011. *Introduction to remote sensing*. 5th ed. New York: Guildford Press.

Chan, Tony F., and Luminita A. Vese. 2001. Active contours without edges. *IEEE Transactions on Image Processing* 10 (2): 266-77.

Cheng, Guangliang, Ying Wang, Shibiao Xu, Hongzhen Wang, Shiming Xiang, and Chunhong Pan. 2017. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*.

Cheng, Guangliang, Ying Wang, Shibiao Xu, Hongzhen Wang, Shiming Xiang, and Chunhong Pan. 2017. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*.

Dubois, David. 2014. *Étude du profil d'échelle des formes et de mesures d'énergie de texture pour l'évaluation semi-automatique des dégâts sur les bâtiments dans les images satellitaires de très haute résolution*. Doctorat., École de technologie supérieure.

Gao, Jay. 2009. *Digital analysis of remotely sensed imagery*. New York: McGraw Hill.

Hopkins, Johns. Statistics how to. 2017 [cited 09/25 2017]. Available from http://www.statisticshowto.com/probability-and-statistics/chi-square/.

Karn, Ujjwal. An intuitive explanation of convolutional neural networks . 2016 [cité 05/23 2017]. Available from https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521 (7553): 436-44.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation.

Maboudi, Mehdi, Jalal Amini, Michael Hahn, and Mehdi Saati. 2017. Object-based road extraction from satellite images using ant colony optimization. *International Journal of Remote Sensing* 38 (1): 179-98.

Medioni, Gérard, Chi-Keung Tang, and Mi-Suen Lee. 2000. Tensor voting: Theory and applications. Paper presented at Proceedings of RFIA.

Milkers, S. Lecture 05 gerard medioni - tensor voting: Fundamentals and recent progress. 2015 [cited 07/10 2017]. Available from https://www.slideshare.net/milkers/lecture-05-gerard-medioni-tensor-voting-fundamentals-and-recent-progress.

Mokhtarzade, M., and M. J. Valadan Zoej. 2007. Road detection from high-resolution satellite images using artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation* 9 (1): 32-40.

Muruganandham, Shivaprakash. 2016. Semantic segmentation of satellite images using deep learning.

Raschka, Sebastian. 2016. Maximum margin classification with support vector machines . In *Python machine learning.* 71-80Packt Publishing Ltd.

Shahi, Kaveh, Helmi Zulhaidi Mohd Shafri, and Alireza Hamedianfar. 2016. Road condition assessment by OBIA and feature selection techniques using very high-resolution WorldView-2 imagery. *Geocarto International*: 1-18.

Thirumuruganathan, Saravanan. Introduction to mean shift algorithm. 2010 [cited 07/07 2017]. Available from https://saravananthirumuruganathan.wordpress.com/2010/04/01/introduction-to-mean-shift-algorithm/.

Zaremba, Marek B. 2011. Automated mapping of hydrographic systems from satellite imagery using self-organizing maps and principal curves. In *Self-organizing maps-applications and novel algorithm design.*InTech.