On The Hausdorff Distance Used For The Evaluation Of Segmentation Results

by M. Beauchemin, KP.B. Thomson, G. Edwards

RÉSUMÉ

Les techniques de segmentation d'images sont largement utilisées en télédétection, comme le démontre leur présence dans la plupart des systèmes commerciaux de traitement d'images. Un effort considérable a été consacré au développement d'algorithmes de segmentation, résultant en une profusion de techniques différentes. En comparaison, le nombre d'outils développés pour l'évaluation quantitative des résultats de segmentation demeure très limité. En fait, l'inspection visuelle des résultats est encore considérée par plusieurs comme le test ultime. Dans cet article, nous présentons une nouvelle méthode quantitative d'évaluation de résultats de segmentation d'images. Cette dernière s'appuie sur la distance d'Hausdorff. La méthode est basée sur le degré de coincidence des points de frontières entre les résultats d'une segmentation d'image et une carte de référence. L'utilité de la méthode est illustrée avec un exemple de segmentation tiré de la littérature.

SUMMARY

Image segmentation tools are widely used in remote sensing. Their implementation in most commercial image analysis packages demonstrates the interest in them. A considerable effort has been directed toward the development of segmentation algorithms, resulting in a profusion of different techniques. In comparison, the number of tools developed for the quantitative evaluation of segmentation outputs is very limited. In fact, visual inspection is still considered, by many, as the ultimate quality test. In this paper, a new quantitative method for the evaluation of segmentation results is presented. It relies on the partial directed Hausdorff distance. The method focuses on boundary point matching between a segmentation output and a reference partition. The method is illustrated with a case taken from the literature and is shown to provide useful information to assess the quality of a segmentation.

INTRODUCTION

Image segmentation is an important topic in image analysis. It basically consists of partitioning an image into distinct units that are homogeneous (see Hoover *et al.*, 1996 for a discussion on the definition of image segmentation). Much of the efforts in that field have been directed toward the development of segmentation algorithms. Consequently, there exist hundreds of segmentation techniques published in the literature (Pal and Pal, 1993). Most commercial packages used by the remote sensing community include segmentation techniques *(e.g.,* K-means clustering). However, attempts to develop methods for the quantitative evaluation of segmentation results have been much less prolific (Pal and Pal, 1993). Still, there is no widely accepted standard method equivalent to, for example, the error matrix and the kappa coefficient utilized for thematic classification evaluation. In fact, visual inspection is still considered by many as the best evaluation process. The present paper reports on a new low level method for the quantitative evaluation of image segmentation. Here, low level refers to the fact that only the segmentation output information is considered. The original image information is not taken into account under this defInition. In low level methods, the assessment of the quality of a segmentation is achieved by comparing the segmentation output to a reference partition or ground reference data *(e.g.,* Levine and Nazif, 1982; Hoover *et al.,* 1996; Edwards, 1995; Lim and Lee, 1990; Rand, 1971). Beauchemin and Thomson (1997) show that, although they focus on different properties or interpretations of the problem, several of the low level methods developed to date are linked to the exploitation of a common entity: the overlapping area matrix. This matrix gives the number of pixels in each segment of a segmentation output that coincide spatially with each segment of a reference partition. Considering the importance of the subject and the restricted number of tools developed to date, the introduction of other measures is of interest.

- Mario Beauchemin is with the Canada Centre for Remote Sensing, 588 Booth Street, Ottawa, Ontario, Canada, K1A OY7
- Keith P. B. Thomson and Geoffrey Edwards are with the Centre Recherche en Géomatique, Pavillon Casault, Université Laval, Québec, Canada G1K 7P4
- Geoffrey Edwards is also with the Chaire industrielle en geomatique appliquée à la foresterie, Pavillon Casault, Université Laval Québec, Canada G1K 7P4

In this paper, a new low level method for the evaluation of segmentation outputs is presented. It is based on boundary point matching between a segmentation output and a reference partition (ground reference data). The proposed method is not linked to the overlapping area matrix. The approach relies on the partial directed Hausdorff distance. This measure is conceptually simple to understand and is trivial to compute. It is assumed that the segmentation output and the reference partition are perfectly registered. The paper is organized as follows. The mathematical framework is first introduced. Then the method is illustrated with a case taken from the literature. Finally, some characteristics of the method are discussed.

THE HAUSDORFF DISTANCE

In the following, we adopt the terminology used in Huttenlocher *et al.* (1993). If $A= \{al, ..., ap\}$ and $B= \{b1, ..., bq\}$ represent two sets of points, the Hausdorff distance is then defined as:

$$H(A, B) = \max(h(A, B), h(B, A)), \tag{1}$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \| a - b \|. \tag{2}$$

The Hausdorff distance represents a measure of the spatial distance between two sets of points. The function *h(A, B)* is referred to as the directed Hausdorff distance from *A* to *B*. It ranks each point of *A* according to its distance to the nearest point of *B*. The largest of these distances determines the value of *h(A, B)*. For this study, the Euclidean norm has been used for $\| a - h \|$. In algorithmic terms, the evaluation of *h(A, B)* consists of the following steps:

i. select a point al in *A*,

ii. calculate the distance from *al* to all points b_i ($1 \le i \le q$) in *B*,

iii. store in memory the minimum distance found,

iv. repeat steps (i) to (iii) for all other points a_i ($2 \le i \le p$) in *A*.

Following this, *h(A, B)* is defined as the largest of the minimum distances stored in memory. A simple interpretation of *h(A, B)* and *H(A, B)* is as follows. If $h(A, B) = \delta$, this means that all points of *A* are within a distance δ of some point in *B*. If $H(A, B)= d$, this indicates that all points of *A* are within a distance *d* of some point in *B* and vice versa. The notion of spatial proximity between two sets of points is encoded in Equation 1. Huttenlocher *et al.* (1993) exploited the Hausdorff distance and other related measures (see next section) to compare binary images. Le Moigne and Tilton (1995) used the Hausdorff distance to refine image segmentation by the integration of edge and region data. They used the Hausdorff distance to compare an edge-detected map with a boundary map resulting from region growing.

THE EVALUATION OF SEGMENTATION OUTPUTS USING THE PARTIAL DIRECTED HAUSDORFF DISTANCE

An extended formulation of Equation 2, the partial directed Hausdorff distance from *A* to *B,* is defined as (Huttenlocher *et at.,* 1993):

$$h_K(A, B) = \underset{a \in A}{K^{th}} \min_{b \in B} \| a - b \|, \tag{3}$$

where $1 \le K \le p$. This is a modified version of Equation 2 based on ranking, where $K^{th}_{a \in A}$ *represents* the K^{th} ranked value in *A*. In the algorithmic terms described in the previous section, this corresponds to ranking in ascending order all the minimum distances stored in memory during step (iii), and then select the K^{th} minimum distance as the value for $h_K(A, B)$. When $K=p$, then $h_K(A, B)$ is equivalent to Equation 2. A pertinent interpretation of Equation 3 is as follows. For $h_K(A, B)= \delta$, *K* represents the number of points in *A* that are within a distance δ of some point in *B*. Equation 3 is the central measure of the proposed evaluation process. It is used to establish the number of points in a binary image, *e.g.,* boundary points in the reference partition, that are each within a specified distance (say δ) of some points in another binary image, *e.g.,* boundary points in the segmentation output. Under that interpretation, the partial distance δ can be seen as a tolerance or an error bound inside which one boundary map matches another one (δ is defined around the set *A* for $h_K(A, B)$). The degree of matching for a given value of δ is established by the number of points *K* that satisfies the relation

$h_K(A, B)= \delta$.

The inputs for the proposed evaluation process are the maps of the boundary in the reference partition and the segmented scene. Since the method uses discrete points, the boundary maps must be in a discrete form. If the map is in a vector form, it must be transformed (to raster data for example). Most segmentation algorithm outputs and ground reference data comes as thematic maps in raster form. One simple way to obtain boundary points from thematic maps consist in recording the coordinates of the inter-pixel positions each time a change in thematic class occurs. Such a transformation must be carried out in the horizontal and the vertical directions. Let N_X denote the number of boundary points in a given map *X* and let the set of points in the reference partition be denoted as *REF* and the one in the segmentation output be labeled *SO.* To evaluate the level of boundary matching between a segmentation output and a reference partition, we consider the relation between the fraction *f* of points in one partition that are each within a distance h_K of some point in the other partition. This is done for both distances $h_K(REF, SO)$ and $h_K(SO, REF)$. First, the case between the reference partition and the segmentation output is considered. The evaluation is performed with the help of a table or a graph of f_{REF} as a function of $h_K(REF, SO)$, where $f_{REF} = K/N_{REF}$. Then, the same procedure is applied for the converse, *i.e.* the distance from the segmentation ouput to the reference partition. This is achieved by deriving f_{so} as a function of $h_K(SO,GT)$, Where $f_{SO} = K/N_{SO}$. Note that the measure is based on f instead of K because in most cases $N_{REF} \neq N_{SO}$, so that normalization is required. The values of both fractions provide a way to quantify to what extent the segmentation output matches the ground reference data boundary structure as a function of h_K. The higher the fractions for low values of h_K, *the* better the matching. Exact coincidence between two maps will occur for values of f_{REF} and f_{SO} equal to unity at $h_K = 0$ (and obviously for $h_K > 0$). By comparing the reference partition with different segmentation results, the SO map that will match more closely the reference partition will be the one where the fractions are the highest for all values of h_K (or a selected set of h_K depending on the goal of the evaluation process).

The partial directed Hausdorff distance can be integrated into a formulation that has a more familiar interpretation. This helps in understanding the meaning of f. Let us propose the following relations:

$$O'(\delta) = 1 - \frac{N_{h_K}(REF, SO) = \delta}{N_{REF}} \tag{4a}$$

or equivalently,

$$O'(\delta) = 1 - f_{REF_{h_K} = \delta} \tag{4b}$$

and

$$C'(\delta) = \frac{N_{SO} - N_{h_K}(SO, REF) = \delta}{N_{REF}} \tag{5a}$$

or equivalently,

$$C'(\delta) = \left(1 - f_{SO_{h_K} = \delta}\right) \frac{N_{SO}}{N_{REF}} \tag{5b}$$

where ${}^{N}h_K(A,B) = \delta$ represents the number K of points in A that are within a distance δ of some point in B. When $\delta = 0$, O' corresponds to omission errors and C' to commission errors (O'Brien, 1991). The omission error is given by one minus the fraction of boundary points that match together. It is a measure of the number of missing points in SO compared to the *REF* map. The commission error represents the number of boundary points in the segmentation output that do not match with the reference partition boundary points, divided by the number of boundary points in the reference partition. It is a measure of the number of points in SO that are in excess of the *REF* map. The lower the values of O' and C', the better the matching between the two maps involved ($0 \leq O' \leq 1$; $C' \geq 0$). For $\delta > 0$, Equations 4 and 5 extend this concept inside an 'error' or tolerance radius of length δ around the boundary set of points under consideration (for instance, around points in A for $h_K(A,B)$). Since O' is directly related to f_{REF} and C' is directly related to f_{SO}, this illustrates the meaning of f_{REF} and f_{SO}. It can be seen that, from the definitions of Equations 4 and 5, the value of O' will increase with under-segmentation (f_{REF} will decrease) while the value of C' will increase with over-segmentation (f_{SO} will decrease). However, values of O' and C' greater than zero (i. *e* f_{REF} and $f_{SO} < 1$) do not necessary indicate the presence of over- and under- segmentation. This may also reflect systematic or random boundary errors around the reference boundaries (reference partition). Regarding the latter case, the method may have potential to quantify the global spatial distribution of error if the two boundary maps have the same basic structure but differ only slightly in location. However, being a nearest-neighbour algorithm, the difference between systematic and random errors (boundary displacements) will be impossible to discern. An illustration of the method applied to a real case is provided in the next section.

APPLICATION TO A SEGMENTED SCENE

Let us consider a case published by Ait Belaid *et al.* (1992) reporting on segmentation experiments in the presence of small agricultural fields. Among other things, the inclusion of partial cartographic information in the segmentation process of a multi-spectral SPOT image was considered by Ait Belaid *et al.* (1992). **Figure 1** shows the ground reference boundary points (G1) and the results of two different segmentations that will be referred to as the unstructured segmentation (*U*) and the structured segmentation (S) (see Ait Belaid *et al.*, 1992 for details). Although this example involves mainly straight line boundaries, there is no boundary shape restriction imposed by the method. Boundary points were obtained from thematic maps by recording the coordinates of the inter-pixel positions each time a change in thematic class occurred (h_K defined in pixel unit). This transformation was carried out in both the horizontal and the vertical directions ($N_{GT} = 533$, $Ns = 812$ and $Nu = 856$). **Figure 2** shows a plot of f_{GT} versus $h_K(GT,S)$, and f_{GT} versus $h_K(GT,S)$, where $f_{GT} = K/N_{GT}$. This graph represents the fraction of ground reference boundary points that are each within a distance h_K of some point in U in one case, and S in the other case. Even though there are 5% more points in U compared to S, 52% of the boundary points in *GT* coincide exactly ($h_K = 0$) with some point in S while this fraction is 5% lower for U with 47%. The fraction of *GT* points that are each within a distance $h_K = 1$ of some point in S in one case, and U in the other one, are respectively 93% and 87%. In fact, for values of $h_K \leq 3$, there is always a higher number of *GT* points nearby S boundary points than in U. A similar plot is displayed in **Figure 3** for f_S versus $h_K(S, GT)$ and f_U versus $h_K(U, GT)$, where f_S and f_U equal K/N_S and K/N_U respectively. This graph represents the fraction of boundary points in the segmentation outputs that are each within a distance h_K of some boundary point in *GT*. As in **Figure 2**, the fraction of matching boundary points is higher for the structured result than the unstructured one for $h_K \leq 3$. In both figures, the fractions are nearly similar for $h_K \geq 4$. This is caused by the fact that h_K approaches the order of magnitude of most segment dimension; h_K becomes too large to distinguish differences between boundary structure.

Table 1 gives $O'(\delta)$ and $C(\delta)$ derived from Equations 4 and 5 for values of $\delta = 0, 1, 2$ and 3 pixels. It can be seen that for all

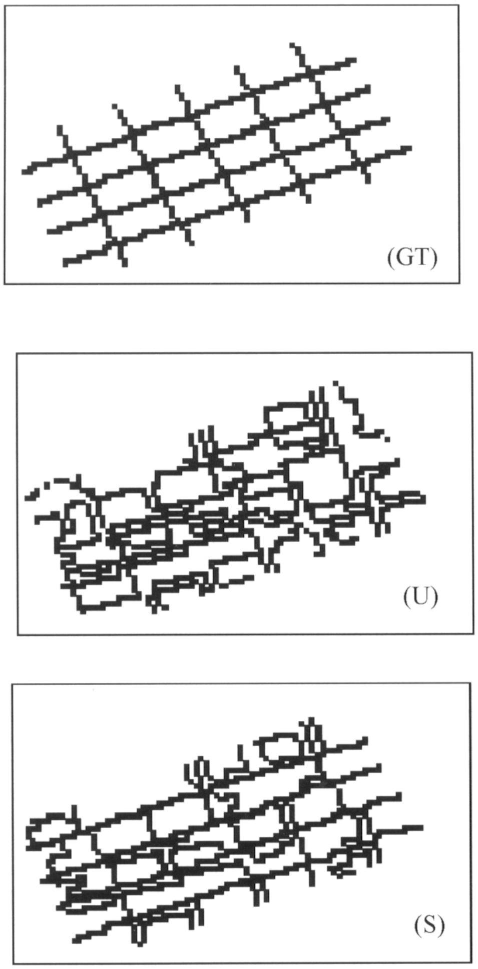

Figure 1.
Boundary maps derived from thematic maps: (*GT*) ground reference, (*U*) unstructured segmentation, and (*S*) structured segmentation.

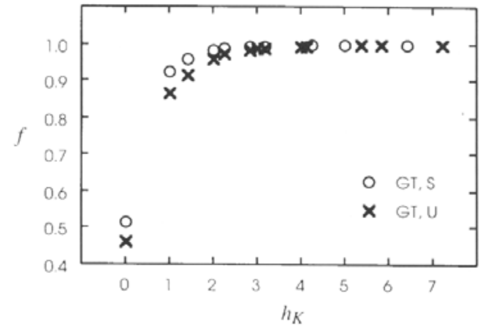

Figure 2.
Fraction (f) of GT points that are each within a distance h_K of some point in U in one case, and S in the other case (h_K in pixels).

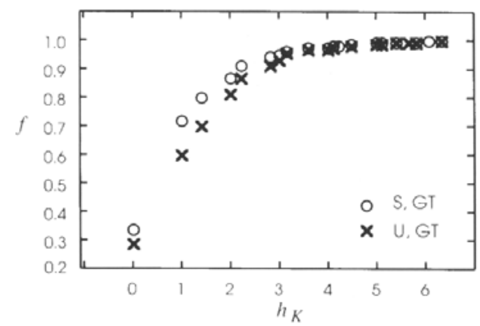

Figure 3.
Fraction (f) of U and S points that are each within a distance h_K of some point in GT (h_K in pixels).

Table 1.
Values of O' and C' as a function of δ (pixels) for the boundary maps shown in Figure 1.

δ	$O'_S(\delta)$	$O'_U(\delta)$	$C'_S(\delta)$	$C'_U(\delta)$
0	0.48	0.54	1.01	1.14
1	0.08	0.13	0.43	0.64
2	0.02	0.04	0.20	0.30
3	0.008	0.011	0.08	0.11

δ, the values of *O*' and *C'* are always lower for the structured result than the unstructured one. This clearly indicates that the structured segmentation provides boundary points that match more closely the ground reference data points than the unstructured one.

It can be concluded from **Figures 2, 3** and **Table 1** that the structured segmentation reproduces the ground reference data boundary structure better than the unstructured segmentation. This is in agreement with the overall conclusion reached in a study carried out by Edwards (1995) using a different method. Although we use a subset a bit smaller than Edwards, we expect that the comparison is valid. In particular, Edwards (1995) found that l)his boundary dispersion index is nearly two times superior for the unstructured segmentation than for the structured one, and 2) the so-called boundary error measure is 40% for the unstructured segmentation and 30% for the structured one.

The partial directed Hausdorff distance seems useful for evaluating segmentation outputs by establishing the degree of resemblance between two boundary point maps. The method should be considered as complementary to existing tools since some proprieties, to be discussed next, restrict its power.

DISCUSSION

The term *boundary points* has been used throughout the text instead of *boundary.* Because the Hausdorff distance is based on sets of points (not continuous lines) and on a nearest-neighbor algorithm, the measure is completely blind to the link that exists between segment area and their surrounding boundary. Consider the hypothetical case illustrated in **Figure 4** which shows a reference partition *(G1)* and a segmentation output *(SO).* In both cases, the straight lines represent the segment boundaries. If segment l' is merged with 3', the fraction of misclassified pixels in *SO,* compared to *GT,* is 22%. The transformation from thematic maps to boundary points is superimposed on the straight lines as square points. It can be seen that $h_K(GT, SO)$ as well as O' will be equal to zero for all but the largest value of K. This should be interpreted as a nearly perfect boundary match. However, the segments labeled l' and 2' are completely disjointed and the misclassified pixel rate is high (if l' and 3' are merged). If the gap shown by the arrow is closed by adding a boundary, the number of misclassified pixels will drop to zero if there is no penalty for over-segmentation. However, the values of O' and C' and the graphs of f versus h_K will resemble the 'not filled' case because the difference resides in only one boundary point. Hence, care must be taken in interpreting the results. The measures based on the Hausdorff distance compare boundary points without consideration to the link they possess with respect to the segments. Depending on the goal of the evaluation process, other measures may be required to complement the Hausdorff-based method. Note that the method presented in this paper can also be considered for performance evaluation implying binary image representations such as edge detection images or road detection mapping (see Wang and Liu, 1994).

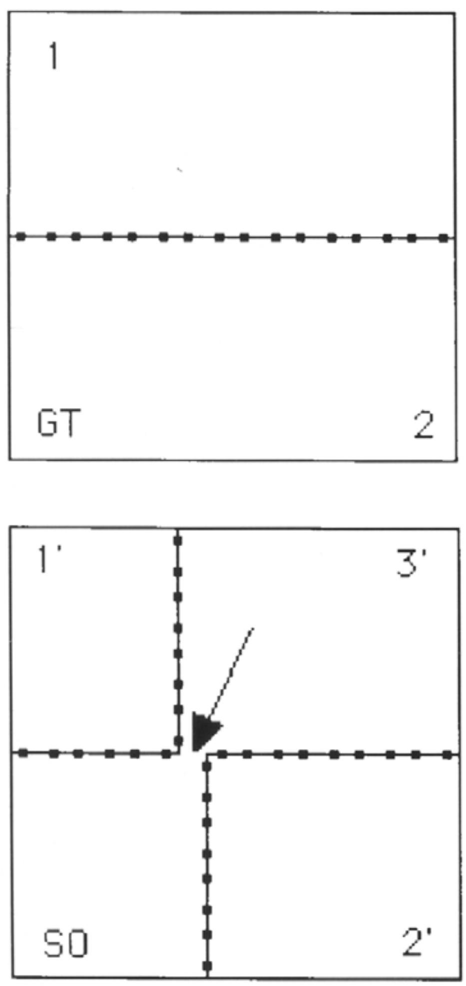

Figure 4.
Hypothetical situation. (*GT*) ground reference map, *(SO)* segmentation output map.

CONCLUSIONS

A method based on the partial directed Hausdorff distance has been presented for evaluating segmentation outputs. The method quantifies how boundaries in a segmentation output resembles the boundaries in the ground reference data map. The method is simple and trivial to compute. It considers the relation between the fraction of points in one partition that are each within a given distance of some point in the other partition. Measures based on an extended definition of commission and omission errors have also been proposed. The evaluation process has been applied to segmentation results previously published in the literature and has been shown to agree with the general conclusions reported in another quantitative evaluation study. Some properties of the method have been discussed. Particularly, it has been stressed that the derived measures compare boundary points without consideration to the link they possess with regard to segments. This can result in misleading interpretations as illustrated by a hypothetical example. Hence, depending on the goal of the evaluation process, other measures may be required to complement the Hausdorff-based approach.

ACKNOWLEDGEMENTS

K. P. B. Thomson acknowledges the financial support of the Natural Science and Engineering Research Council through an individual operating grant. G. Edwards acknowledges the financial support of the "Association des Industries Forestieres du Quebec" and the Natural Sciences and Engineering Research Council through the establishment of an industrial research chair. The authors acknowledge the Geomatics Research Centre at Laval University which provided partial support for M. Beauchemin.

REFERENCES

Ait Belaid, M., J.- M. Beaulieu, G. Edwards, A. Jaton, and K. P. B. Thomson. 1992. "Post-Segmentation Classification of Images Containing Small Agricultural Fields," *Geo-Carto*, Vol. 7, pp. 53-60.

Beauchemin, M., and K. P. B. Thomson. 1997. "The evaluation of segmentation results and the overlapping area matrix," *International Journal of Remote Sensing Letters,* Vol. 18, pp. 3895-3899.

Edwards, G. 1995. "Methods for Assessing Local Map Accuracy in Thematic Classifications Derived from Remotely Sensed Images, " *Proceedings of the 17th International Cartographic Conference,* Barcelona, Espana, pp. 1521-1530.

Hoover, A., G. Jean-Batiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgoh, K. Bowyer, D. W. Eggert, A. Filtzgibbon, and R. B. Fisher. 1996. "An experimental comparison of range image segmentation algorithms," *IEEE on Pattern Analysis and Machine Intelligence,* Vol. 18, pp. 673-689.

Huttenlocher, D. P., G. A. Klanderman, and w. J. Rucklidge. 1993. "Comparing Images Using the HausdorffDistance, " *IEEE on Pattern Analysis and Machine Intelligence,* Vol. 9, pp. 850-863.

Le Moigne, J. and J. C. Tilton. 1995. "Refining Image Segmentation by Integration of Edge and Region Data, " *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 33, pp. 605-615.

Levine, M. D. and A. M. Nazif. 1982. "An Experimental Rule-Based System for Testing Low Level Segmentation Strategies. " In *Multicomputers and Image Processing: Algorithms and Programs* (New York: Academic Press), pp. 149-160.

Lim, Y. W. and S. U. Lee. 1990. "On the Color Image Segmentation Algorithm Based on the Thresholding and Fuzzy c-means Techniques, " *Pattern Recognition*, Vol. 23, pp. 935-952.

O'Brien, D. 1991. "Computer Assisted Feature Extraction (InterEx)," *Proceedings of the 14th Canadian Symposium on Remote Sensing,* Calgary, Alberta, pp. 423-428.

Pal, N. P. and S. K. Pal. 1993. "A Review on Image Segmentation Techniques, " *Pattern Recognition*, Vol. 26, pp. 1277-1294.

Rand, W. M. 1971. "Objective Criteria for the Evaluation of Clustering Methods, " *Journal of the American Statistical Association*, Vol. 66, pp. 846-850.

Wang, *I.*, and W. Liu. 1994. "Road Detection from Multi-Spectral Satellite Imagery, " *Canadian Journal of Remote Sensing,* Vol. 20, pp. 180-191.