Assessment of Very High Spatial Resolution Satellite Image Segmentations

A.P. Carleer, O. Debeir, and E. Wolff

Abstract
Since 1999, very high spatial resolution satellite data represent the surface of the Earth with more detail. However, information extraction by per pixel multispectral classification techniques proves to be very complex owing to the internal variability increase in land-cover units and to the weakness of spectral resolution. Image segmentation before classification was proposed as an alternative approach, but a large variety of segmentation algorithms were developed during the last 20 years, and a comparison of their implementation on very high spatial resolution images is necessary. In this study, four algorithms from the two main groups of segmentation algorithms (boundary-based and region-based) were evaluated and compared. In order to compare the algorithms, an evaluation of each algorithm was carried out with empirical discrepancy evaluation methods. This evaluation is carried out with a visual segmentation of Ikonos panchromatic images. The results show that the choice of parameters is very important and has a great influence on the segmentation results. The selected boundary-based algorithms are sensitive to the noise or texture. Better results are obtained with region-based algorithms, but a problem with the transition zones between the contrasted objects can be present.

Introduction
The first commercial very high-resolution satellite (Ikonos) became accessible in autumn 1999, the QuickBird satellite in October 2001, and OrbView-3 was launched on 26 June 2003. These sources of very high-resolution images are increasing the amount of information available for land-cover from local to national scales (Aplin *et al.*, 1999). These data provide amazing detail of the Earth's surface, but information extraction using computer-assisted classification techniques appears to be much more complex.

Internal Variability
These sources of very high spatial resolution images do not provide necessarily better classification; this observation had already been made by Irons *et al.* (1985) at the time of the marketing of the Landsat-4 Thematic Mapper (TM) images. The refinement of spatial resolution from 80 m to 30 m did not often improve classification accuracy, even though the advantages of a higher resolution system appeared obvious when visual comparisons were made between TM and MSS imagery. This incongruous result of earlier studies had been attributed to one consequence of change of spatial resolution. With the spatial resolution refinement, the internal variability within homogenous land cover units increases (Cushnie, 1987; Woodcock and Strahler, 1987; Aplin *et al.*, 1997; Zhang, 2001; Thomas *et al.*, 2003). The increased variability decreases the statistical separability of land-cover classes in the spectral data space. This decreased separability tends to reduce per pixel classification accuracies, such as maximum-likelihood classification algorithms. The increased variability was attributed to the imaging of diverse class components by higher resolution sensors, whereas at coarser resolutions, sensors integrated the reflected spectral radiance of the various components, and classes appeared more homogeneous (Irons *et al.*, 1985). For example, the sunlit and shady sides of the same tree have vastly different spectral responses, even though they belong to the same class (Thomas *et al.*, 2003).

Spectral Resolution
Another disadvantage with the very high spatial resolution satellite images is the relatively poor spectral resolution (Herold *et al.*, 2003b). While the spatial resolution is fine, spectral capabilities are limited compared to sensors like Landsat TM. Generally, there is a trade-off between the spatial resolution and the spectral resolution (Aplin *et al.*, 1997; Key *et al.*, 2001). The spectral sensibility of the receptor cell requires a sufficient instantaneous field of view (IFOV). The spectral resolution depends on the ratio of signal to noise, and this ratio is linked to the IFOV, the height of flight, and the opto-electronic characteristics of receptor cell (Lillesand and Kiefer, 1994).

Region-based Procedure
To overcome these problems, a region-based procedure can be used. The first step of this procedure is the segmentation. The segmentation process is successful at removing much of the structural clutter, and performs well in comparison with traditional majority filtering (Barr and Barnsley, 2000). This alternative approach had been proposed by Hill (1999) to reduce local spectral variation. Image regions are more homogeneous within themselves than with nearby regions

A.P. Carleer is with the Institut de Gestion de l'Environnement et d'Aménagement du Territoire, Université Libre de Bruxelles, CP 130/02, 50 av. F. Roosevelt, 1050 Brussels, Belgium, (acarleer@ulb.ac.be).

O. Debeir is with the Information and Decision Systems, Université Libre de Bruxelles, CP 165/57, 50 av. F. Roosevelt, 1050 Brussels, Belgium.

E. Wolff is with the Institut de Gestion de l'Environnement et d'Aménagement du Territoire, Université Libre de Bruxelles, CP 130/02, 50 av. F. Roosevelt, 1050 Brussels, Belgium.

Photogrammetric Engineering & Remote Sensing
Vol. 71, No. 11, November 2005, pp. 1285–1294.

0099-1112/05/7111–1285/$3.00/0

and represent discrete objects or areas in the image. Each image region then becomes a unit analysis for which a number of attributes, on top of spectral attributes, can be measured and used during the classification (Herold *et al.*, 2002a; 2003b; Thomas *et al.*, 2003).

The attributes which can be measured are, for example, the spatial information. This information can be used in order to increase the classification accuracy for spectrally heterogeneous classes (Lillesand and Kiefer, 1994) and to overcome the current spectral limitations of very high spatial resolution satellite images (Guindon, 2000; Herold *et al.*, 2002b). The attributes can be the area, the perimeter, the compactness (area/perimeter^2), and the degree and kind of texture (Johnsson, 1994). In the per pixel methods, the spatial attributes are calculated on an arbitrary neighborhood using mobile windows. The image segmentation enables to obtain the specific spatial attributes of the regions without taking into account nearby regions. Moreover, the spatial attributes, like texture, calculated with mobile windows, smooth the boundaries between discrete land cover region (Herold *et al.*, 2003a; Carleer and Wolff, 2004) and create between-class texture, which is often more distinctive than within-class texture. This between-class texture causes an edge effect in the classification (Ferro and Warner, 2002). This edge effect problem could disappear with segmentation.

The classification of these segmented images is often more accurate than per pixel classification, overcoming the misclassification problem due to the internal variability of the regions (Aplin *et al.*, 1999).

Objective

A large variety of segmentation algorithms were developed these last 20 years (Zhang, 1997; Haralick and Shapiro, 1984), and a comparison between them on very high spatial resolution satellite images is necessary. This paper presents an evaluation and comparison study of different types of segmentation algorithms on very high spatial resolution satellite images. The evaluation and comparison are made on different landscapes as a whole and not on particular objects like in other studies (Karantzalos and Argialas, 2003; Neubert and Meinel, 2003).

Segmentation Algorithms

Segmentation algorithms can conveniently be classified as boundary-based and region-based (Zhang, 1997; Nicolin and Gabler, 1987; Janssen and Molenaar, 1995; Guindon, 1997). Boundary-based algorithms detect object contours explicitly by using the discontinuity property, and region-based algorithms locate object areas explicitly according to the similarity property (Zhang, 1997). The boundary approach gathers the edge detection techniques. These methods do not lead directly to a segmentation of the image because contours obtained are seldom closed; therefore, it is necessary to carry out closing edges algorithm if one wishes a complete segmentation of the image. Indeed, after contours closing, the duality contours/regions appears clearly. A region is defined as the inside of a closed line. On the other hand, the methods of the region approach lead directly to a segmentation of the image, each pixel being assigned to a single region (Cocquerez and Philipp, 1995). For this study, algorithms from each group have been selected.

Boundary-based Algorithms

Two algorithms of this group have been selected: "optimal edge detector" (Canny, 1986; Cocquerez and Philipp, 1995) and "watershed segmentation" (Vincent and Soille, 1991; Debeir, 2001).

In "optimal edge detector" the procedure first filters the image with the Canny-Deriche filter. This filter provides the derivative of the images and takes into account the Canny criterion (Canny, 1986): namely *good* detection (low probability of missing real edges and detecting noise), *good* position estimation and single response to each edge, which also accommodates edges of a finite width. Next, a hysteresis thresholding (Hou and Koh, 2003; Ding and Goshtasby, 2001) is achieved on the image to preserve the coherent boundaries. Finally, contours are closed by the way of best count. The count is calculated by the sum of the pixels norm constituting the way in the gradient image. This procedure is proposed in the Signal and Images software Packages (SIMPA) by the GDR Information, Signal, Images et Vision (ISIS) of the Centre National de la Recherche Scientifiques, France (CNRS).

In the "watershed segmentation" (Vincent and Soille, 1991; Debeir, 2001), the procedure first transforms the original data into a gradient image. The resulting grey tone image is considered as a topographic surface. This surface is flooded from its minima and the merging of the waters coming from different sources is prevented, thus the image is divided into a set of watershed lines. The catchment's basin should correspond to the homogeneous regions of the image. Before transforming the original data into a gradient image, a median filter can be applied on the image to reduce the noise. The presence of noise in an image causes an over-detection of edges by the morphological gradient. The median filter locally homogenizes the image and avoids extreme gradients, and thus disturbing contours. It is also a good means not to take into account object texture during contours detection. The image gradient can also be thresholded to limit the contour sensitivity; e.g., if the threshold is 10, we keep pixels with a gradient higher than 10, and the others are put at 0 as if there are no edges. This watershed segmentation procedure was implemented by the department of Logical and Digital Systems (SLN) of the Université Libre de Bruxelles-Belgium (Debeir, 2001).

Region-based Algorithms

In this group, two algorithms have been selected: "multilevel thresholding technique" (Deravi and Pal, 1983) and a "region-growing" technique.

The "multilevel thresholding" algorithm is a global thresholding of a grey tone image which uses second-order grey level statistics. The segmentation is carried out with a non-symmetric, co-occurrence matrix. For all grey levels, the conditional estimated probabilities of intensity transition between two regions separated by the grey level n are calculated. The lower the value of this estimated probability, the lower the probability that the next transition will be in a different intensity class. Therefore, it is assumed that meaningful sets of thresholds would correspond to the minima of this measure. Then, the minima are searched on a range of a number of neighboring grey levels on both sides of the grey level n. These minima are used as image segmentation thresholds (Deravi and Pal, 1983; Biswas and Pal, 2000). This procedure is proposed in the SIMPA software GDR ISIS at CNRS.

In the selected "region-growing" technique, the procedure starts at each point in the image with one-pixel objects, and in numerous subsequent steps, smaller image objects are merged into bigger ones, throughout a pair-wise clustering process. The merger is based on three criteria: color, smoothness, and compactness. These criteria are combined in numerous ways to obtain varying output results. The combination of these criteria is defined as a *within-segment* heterogeneity. These criteria optimize the region's spectral homogeneity and the spatial complexity. The balance at which these criteria are applied depends on the desired output (Thomas *et al.*, 2003). If the smallest growth exceeds a heterogeneity tolerance defined by the user, the process stops (Definiens Imaging, 2003; Burnett and Blaschke, 2003; van der Sande *et al.*, 2003). The heterogeneity tolerance affects the relative size of output polygons. This procedure was developed by Definiens Imaging, Inc. in its eCognition® software.

Test Images

The test images are VHR satellite images of various types of landscape. The test images are extracted from an Ikonos panchromatic scene of 08 June 2000 of the Brussels area in Belgium (Figure 1). Only the panchromatic band was used because the both boundary-based algorithms and the "multi-level thresholding" algorithm do not segment several bands at the same time, and the panchromatic band offers the finest spatial resolution. The extent of the extracts is 256 m by 256 m, and the spatial resolution is 1 meter. The size of 256 pixels by 256 pixels is very common in the image libraries which serve this kind of assessment. The various landscape types in Figure 1 are:

- rural (RURAL): a rural area with fields, hedges and some isolated houses;
- residential (RESI): a residential area with isolated houses, large gardens, and many trees;
- urban administrative zone (ADM): a business area with many office-blocks, perpendicular roads, and shadows;
- urban dwelling zone (DW): a residential area with adjoining houses, little gardens, some trees, roads, and shadows; and
- forest (FOREST): a forest area with different age and species trees.

Each extract is visually segmented. The visual segmentations are used as reference for the assessment (Lee *et al.*, 1990; Pal and Bhandari, 1993; Neubert and Meinel, 2003; Carleer *et al.*, 2004). These visual segmentations were digitized manually in a GIS, and a visual segmentation method was defined and consists in three main points:

- an appropriate minimum size of four pixels is selected for the smallest region to be represented (Welch, 1982);
- the boundaries of adjacent regions are placed at the center of the transition zone between them, but if the transition zone occurs consistently, it is considered to be a separated object (Campbell, 1996); and
- a contrast of twenty is selected, on a grey level of 256, as the tonality difference between two adjacent regions.

We could have used, in this study, synthetic images which can be obtained from image generation procedure, but this kind of images do not posess all the characteristics of the real images, like textures, contrasts, and forms.

Evaluation Methods

Segmentation algorithms can be evaluated analytically or empirically (Zhang, 1996). The analytical methods directly examine and assess the segmentation algorithms themselves by analyzing their principles and properties. The empirical methods indirectly judge the segmentation algorithms by applying them to test images, and measuring the quality of segmentation results compared with a reference segmentation (empirical discrepancy methods), or by measuring some desirable properties of segmented images (empirical

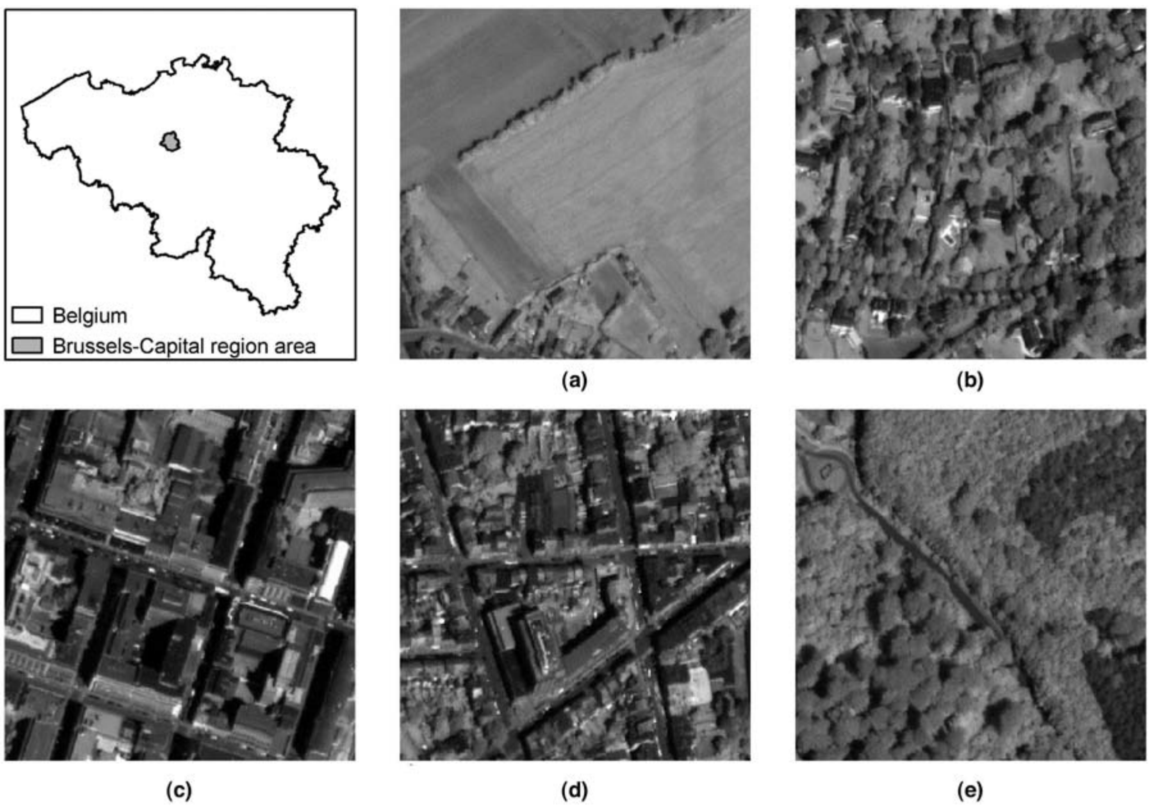

Figure 1. Extracts of the Ikonos image: (a) Rural zone (RURAL), (b) Residential zone (RESI), (c) Administrative zone (ADM), (d) Dwelling zone (DW) and (e) Forest zone (FOREST).

goodness methods). For this study, we have chosen empirical discrepancy evaluation methods. Indeed, analytical methods are easier to apply, but they often provide only qualitative properties of algorithms. Empirical methods are normally quantitative, as the values of quality measures can be numerically calculated. Among them, desirable properties for the goodness methods are chosen subjectively, and moreover, it should be assured that the properties are not used by segmentation algorithms to avoid a bias assessment. Discrepancy methods can be both objective and quantitative (Zhang, 1996).

The first selected method is the discrepancy, based on the number of mis-segmented pixels in the segmented images compared with the visually segmented reference images (Neubert and Meinel, 2003; Carleer *et al.*, 2004).

Considering the segmentation as a pixel classification process, the percentage of mis-classified pixel is a measurement of discrepancy. Suppose an image made up of N_{ref} pixel classes (the number of reference regions), a confusion matrix C of dimension N_{ref} can be constructed, where each entry C_{ij} represents the pixel number of class j classified as class i by the segmentation algorithms. The evaluation measure is defined as:

$$E = \frac{\left[\left(\sum_{i=1}^{N_{ref}}\sum_{j=1}^{N_{ref}} C_{ij}\right) - \sum_{k=1}^{N_{ref}} C_{kk}\right]}{\left(\sum_{i=1}^{N_{ref}}\sum_{j=1}^{N_{ref}} C_{ij}\right)} \times 100 \tag{1}$$

where the numerator represents the number of pixels mis-classified, and the denominator is the total number of pixels in the test image. The mis-segmented pixels are the pixels of the regions, for the most part, classified in a reference region but being in the parts spanning adjacent reference regions (Figure 2).

In this evaluation method, the same importance is not given to small and great regions. A segmentation with large regions could give a small percentage of mis-classified pixels, whereas the majority of the small regions are mis-segmented. We should give the same importance to small and large regions, therefore, we modify the first discrepancy measure. This new measure is based on the same principle as the first method, but the percentage of mis-classified pixels is calculated on each reference region, and then, the mean percentage is calculated. The new evaluation measure is defined as:

$$E_{N_{ref}} = \frac{\sum_{i=1}^{N_{ref}}\left(\frac{\sum_{j=1}^{N_{ref}}(C_{ij} - C_{ii})}{n_{iref}} \times 100\right)}{N_{ref}} \tag{2}$$

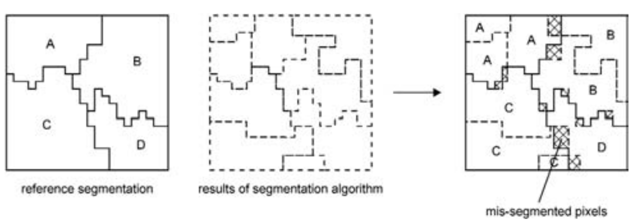

Figure 2. Intersection between the reference segmentation and the result of segmentation algorithm to identify the mis-segmented pixels.

Table 1. Mean Local Variance of Images

Test Images	Local Variance
FOREST	10,46
DW	14,92
RESI	13,21
RURAL	6,21
ADM	12,82

where n_{iref} is the number of pixels in the region i (class i) of the reference segmentation image.

The second evaluation method selected is a simple ratio between the number of regions in the segmented image and the number of regions in the reference segmentation. This ratio is the *generalization*, and is defined as:

$$Gen = \frac{N_s}{N_{ref}} \tag{3}$$

where N_s is the number of regions in the segmented image and N_{ref} is the number of regions in the reference segmentation. This measure allows for evaluation of the over-segmentation ($Gen > 1$) or the under-segmentation ($Gen < 1$) of tested algorithms (Debeir, 2001). Over-segmentation is not a defect in itself because it could be recovered during the classification of the regions (Janssen and Molenaar, 1995); but if the over-segmentation is significant, the advantages of carrying out image segmentation before classification are lost. Usually, with the increase in number of regions, errors are expected to decrease (Biswas and Pal, 2000). However, under-segmentation cannot be recovered during classification because the objects are mis-identified (Debeir, 2001).

The most important evaluation measure is the average error by region (E_{NRef}) which makes it possible to give the same weight to all the regions of the visual segmentation. It is all the same interesting to analyze the total error (E). A high total error value, being the consequences of either small or great regions, is not desirable. The *Gen* evaluation measure is not the most important measure in the segmentation methods evaluation except if an under-segmentation is present. It enables to know if the advantage of carrying out a segmentation is not lost. In order to complete and to help the interpretation of the evaluation, the mean local variance (Table 1), and the gradient image of each extract (test images) was also calculated. The mean local variance is calculated within a 3 $\times$ 3 pixels variance filter. The mean local variance provides a measure of internal variability of the extracts (Cushnie, 1987).

Results

First, the results of algorithm evaluations are presented separately after which they will be compared at the end of the next section.

Optimal Edge Detector (Canny-Deriche Optimal Filter)

Three sets of high- (ht) and low-thresholds (lt) are selected for the hysteresis thresholding (ht = 4 and lt = 2, ht = 8 and lt = 5, ht = 15 and lt = 1). The results are presented in Table 2. The over-segmentation values (*Gen*) are reasonable, and even go as far as under-segmentation with the DW image test, which is an important defect of the segmentation. The under-segmentation of DW and RURAL (Deriche-C) test images and the very low over-segmentation of ADM, RESI, and RURAL (Deriche-B) test images are due to a low identification of object contours. Part of the object contours were removed after the optimal filtering in these test images. This poor contour detection is confirmed by high total errors (E) and high average errors by region (E_{NRef}).

TABLE 2. OPTIMAL EDGE DETECTOR EVALUATION RESULTS

Algorithms	Test Images	*E* (%)	*ENRef* (%)	*Gen*	Nb of Regions
Deriche_A (ht = 4, lt = 2)	FOREST	11,03	46,87	5,44	1083
	DW	34,83	58,04	0,98	839
	RESI	29,54	50,52	1,31	868
	RURAL	9,57	56,81	3,12	742
	ADM	24,17	49,86	1,57	721
Deriche_B (ht = 8, lt = 5)	FOREST	11,13	45,82	5,08	1011
	DW	34,90	58,07	0,96	823
	RESI	29,80	50,11	1,27	840
	RURAL	9,85	57,92	1,45	345
	ADM	24,93	50,60	1,45	667
Deriche_C (ht = 15, lt = 1)	FOREST	11,95	50,13	3,71	738
	DW	36,55	59,18	0,88	752
	RESI	30,31	51,02	1,14	755
	RURAL	33,46	82,06	0,88	209
	ADM	26,64	52,23	1,22	560

The change of hysteresis thresholding thresholds does not change anything about the errors and over-segmentation for the ADM, RESI, and DW test images. Part of the objects in these three images are adequately contrasted and homogeneous, and thus, the object borders are characterized by strong gradients which are always detected with a high high-threshold. A low high-threshold does not detect more borders because the objects are more homogeneous (less local maxima within those). The fact that the RURAL and FOREST test images are more textured than the others, explains that over-segmentation decreases as the threshold increases. With a low high-threshold, the result of the segmentation is disturbed by the local maxima resulting from the object texture. With a high high-threshold, the borders due to the texture disappear, but the object borders which have less contrasted can also disappear, as it is the case with the RURAL test image. With the RURAL image the total error (*E*) strongly increases with the high-threshold going from 8 up to 15.

The ascending order of the absolute number of regions resulting from the segmentation gives us a curious order (Table 2): FOREST > RESI > DW > RURAL > ADM. This order does not follow the ascending order of the mean local variance (Table 1) as one could expect. This could be explained by the distribution of the maxima within the gradient images (Figure 3). The fact that the FOREST image test is more segmented while not having the greatest mean local variance, can be explained by the objects within this image being comparably textured and little contrasted; this leads to the detection of many local maxima in the gradient image, and thus, detection of many contours of many regions. On the other hand, for the ADM test image for which the mean local variance is high, the number of regions resulting from the segmentation is lower. The objects in this image are contrasted and more homogeneous (less textured); this explains the greater mean local variance and a fewer number of regions. Indeed, the local maxima of gradient image correspond to the object borders and not to their texture. Contrary to what is mentioned in other studies (Cushnie, 1987; Woodcock and Strahler, 1987), a high mean local variance does not necessarily indicate the presence of strongly textured objects. An image of homogenous contrasted objects can provide higher mean local variance than an image of textured non-contrasted objects, as is the case with the ADM and FOREST images (Figure 3).

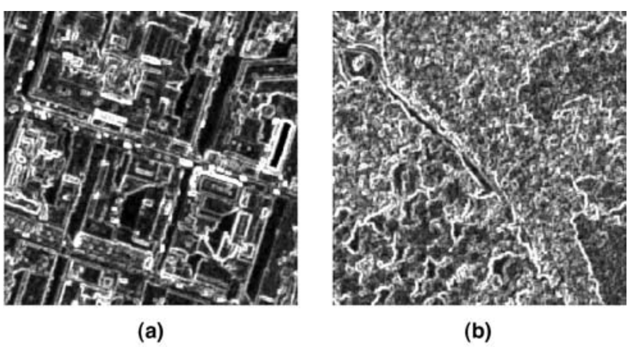

Figure 3. The local maxima in the gradient image of (a) the ADM test image correspond to the object borders and not to their texture as in (b) the gradient image of the FOREST test image.

The segmentations of the RURAL and FOREST test images give a total error (*E*) below 15 percent, but have an average error by region (E_{Nref}) over 45 percent which is not acceptable in a segmentation. One notices that the average errors by region are significant for all test images. This is explained by the fact that an image does not only consist of homogeneous and contrasted objects or textured and little contrasted objects, but of a proportion of both; for this reason, the choice of suitable thresholds is difficult. If the high-threshold is high, many contours will be missing. It is not profitable to take it too low because in this case, the result can be parasitized: by having a lower high-threshold, some pixels not corresponding to desired contours are preserved and are then unfortunately supplemented by connection in complete contours in the final result.

This contour detection method is adapted for homogeneous and contrasted objects detection, such as buildings, but not for whole image segmentation.

Watershed Segmentation

As for the contour detection method by Canny-Deriche operator (optimal filtering), the test images are filtered. In this method the filter is a median filter of 3 × 3 pixels instead of the Canny-Deriche optimal filter. Median filtering makes it possible to remove the noise while preserving contours. Moreover, the gradient images are thresholded by a simple threshold, and the contour detection is carried out by "watershed." Two simple thresholds are selected for this segmentation (10 and 5). The results are presented in Table 3.

It is noticed that the change of threshold does not change the total errors (*E*) and average errors by region (E_{Nref}) for the ADM, DW, RESI, and FOREST test images, there is

Table 3. Watershed Segmentation Evaluation Results

Algorithms	Test Images	E (%)	ENRef (%)	Gen
Watershed_1 (t = 10)	FOREST	6,03	22,59	23,68
	DW	17,39	27,95	5,61
	RESI	14,14	23,10	7,48
	RURAL	34,96	32,56	5,23
	ADM	15,64	27,20	7,82
Watershed_2 (t = 5)	FOREST	6,05	23,02	38,79
	DW	14,74	26,04	7,63
	RESI	13,34	22,70	10,19
	RURAL	8,52	27,12	12,33
	ADM	11,58	25,41	11,79

a light decrease in these two errors. On the other hand, the change of threshold has a great influence on the total error of the RURAL test image for which the total error (*E*) falls from 35 percent to 8 percent. Which is explained by the detection of new contours delimiting objects of the visual segmentation which were not detected previously. One also notes a big increase of the over-segmentation for FOREST test image with a reduction of the threshold. The reduction of the threshold allows the identification of maxima not corresponding to contours in the gradient image: the segmentation is disturbed. One can say regarding the FOREST test image that identifiable contours by this method had already been detected with a threshold of 10 since there is no improvement of the total error and average error by region. Broadly, the use of a threshold of 5 works well on all test images; the best results were obtained with this threshold. Similar results were already obtained with a threshold of 10 except for RURAL test image where contours were missing.

Multilevel Thresholding Technique (Deravi and Pal Segmentation)

Two numbers of neighboring grey levels (10 and 15) are selected for the search of the minima, the evaluation results are in Table 4. In the multi-level thresholding segmentation, there is a significant salt and pepper effect which contributes to the over-segmentation of images. Because of the high local variance, the adjacent pixels are not in the same range of gray values between two thresholds, which explain the salt and pepper effect (Figure 4b). Between 49 and 60 percent of the regions are composed of one pixel. A modal filter could have been applied after the thresholding, but that would not improve the errors because the one pixel regions are contained in a reference region and are consequently well segmented.

The total errors (*E*) are above 20 percent (Deravi_10) except for FOREST, but then the over-segmentation is very significant. The thresholding methods are less effective on images presenting a unimodal histogram. Such histograms are typical in images with many small different objects, such as aerial photographs or very high spatial resolution satellite images. The histogram of the image RURAL (Figure 4c) test is bimodal, but the two peaks do not represent contrasted objects and background. These images do not exhibit a clear background-foreground distinction (Deravi and Pal, 1983) (Figure 4a).

This method could be used to isolate particular objects in an image, e.g., objects having a particular spectral characteristic (peak in the histogram), but not to segment a whole image, whatever the landscape. The method is appropriate for images of objects on homogeneous background or for images of homogeneous, adjacent, and contrasted objects, as it is the case for the ADM test image which presents the smallest average error by region (Deravi_10). However, a problem remains with the images of contrasted objects, like urban images: if the number of thresholds is sufficient, the transition zone between contrasted objects will be isolated as an object. The value of the transition zone falls in an other range of value as the two adjacent objects. This effect disappears if the number of thresholds decreases (from Deravi_10 to Deravi_15), but the errors are too significant.

Finally, it should be mentioned that the method is applicable when the scene is of object-background nature (Pal and Bhandari, 1993), and the texture level is not very high.

Region-growing Technique

In this region-growing technique, the main merger criterion is the color (spectral homogeneity); the other merger criterion is a region form criterion combining smoothness and compactness (spatial complexity). A different weight can be given to each one of these criteria in the heterogeneity value calculation which cannot exceed an *a priori* fixed heterogeneity tolerance. Four heterogeneity tolerances (het = 5, 10, 15, 20) as well as two weights were selected for both criterions and led to eight different segmentations. The **rg1** algorithm does not take into account the region form criterion, whereas the **rg3** does. The results are presented in Table 5.

One notices that the best results are obtained by taking the form into account (**rg3**), all heterogeneity tolerances taken together and for all test images. The results are obviously the best with the smallest heterogeneity tolerance (5). With a small heterogeneity tolerance, mergers of pixels and regions are limited, and there is a considerable over-segmentation, which results into decreasing errors.

While the heterogeneity tolerance goes from 5 up to 10, the total errors (*E*) for **rg3** and for all landscapes remain lower than 15 percent and increase in a limited way for RURAL and FOREST test images. The average errors by region (E_{NRef}) for "**rg3** het = 10" are included between 20 and 30 percent, and there is a very significant decrease of the over-segmentation. In this method, it can be noticed that all test images respond in the same way to the change of

Table 4. Multilevel Thresholding Technique Evaluation Results

Algorithms	Test Images	E (%)	ENRef (%)	Gen	Number of Threshold	One Pixel Regions (%)
Deravi_10	FOREST	7,48	22,61	41,78	6	52,66
	DW	25,11	24,92	11,10	8	52,23
	RESI	26,57	33,76	9,70	6	50,89
	RURAL	18,84	42,99	10,00	3	60,15
	ADM	20,39	21,29	15,21	6	52,59
Deravi_15	FOREST	41,40	77,69	8,91	2	55,30
	DW	54,48	58,17	3,65	5	51,78
	RESI	50,02	53,73	5,44	3	50,82
	RURAL	20,89	43,65	10,00	3	60,15
	ADM	79,87	69,13	3,46	3	49,12

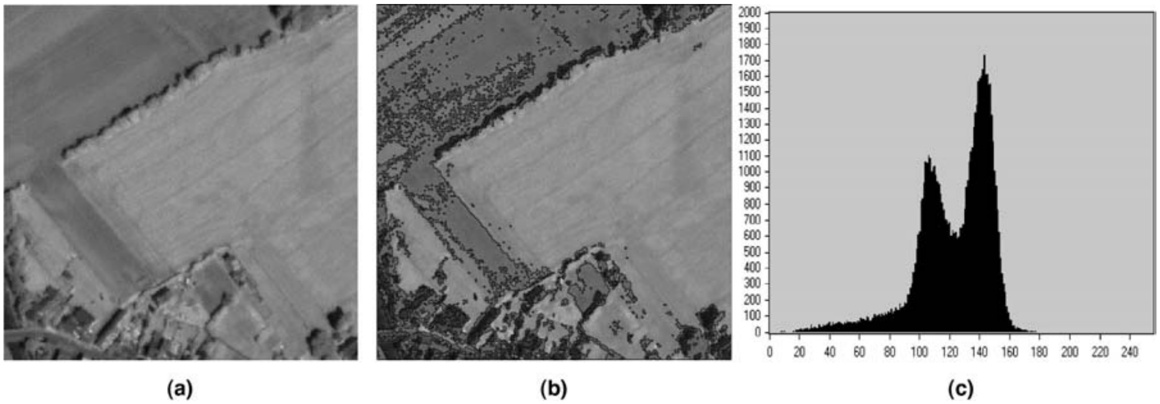

Figure 4. (a) RURAL test image, (b) the deravi_10 RURAL test image segmentation is disturbed by a salt and pepper effect, and (c) the bimodal histogram of the RURAL test image do not represent contrasted objects and background.

TABLE 5. REGION-GROWING TECHNIQUE EVALUATION RESULTS

Algorithms	Test Images	*E* (%)	*ENRef* (%)	*Gen*
rg1 (het = 5)	FOREST	4,18	11,04	30,87
	DW	8,91	12,65	9,79
	RESI	8,09	12,53	11,45
	RURAL	3,07	15,30	12,76
	ADM	6,89	9,83	15,65
rg1 (het = 10)	FOREST	6,88	26,84	7,53
	DW	19,44	29,23	2,56
	RESI	17,55	27,64	2,99
	RURAL	6,51	33,35	3,13
	ADM	13,19	22,84	3,98
rg1 (het = 15)	FOREST	9,15	37,09	3,67
	DW	28,23	42,91	1,28
	RESI	25,43	40,30	1,54
	RURAL	8,88	47,81	1,45
	ADM	20,05	34,49	1,98
rg1 (het = 20)	FOREST	11,81	50,99	2,15
	DW	37,06	56,53	0,79
	RESI	34,84	52,45	0,96
	RURAL	11,66	57,02	0,97
	ADM	24,69	44,24	1,25
rg3 (het = 5)	FOREST	3,47	8,91	29,28
	DW	8,06	11,61	8,63
	RESI	7,04	11,12	10,36
	RURAL	2,52	11,36	13,66
	ADM	6,30	9,56	14,37
rg3 (het = 10)	FOREST	5,78	22,83	7,88
	DW	14,73	24,33	2,41
	RESI	13,02	22,91	2,90
	RURAL	5,20	29,79	3,54
	ADM	10,80	19,21	3,78
rg3 (het = 15)	FOREST	8,19	39,61	3,68
	DW	21,39	39,09	1,24
	RESI	19,29	35,67	1,48
	RURAL	7,34	47,78	1,66
	ADM	14,89	31,49	1,88
rg3 (het = 20)	FOREST	10,66	53,27	2,14
	DW	29,69	53,78	0,75
	RESI	26,30	50,02	0,90
	RURAL	9,25	56,89	1,05
	ADM	19,24	43,13	1,15

parameters. We also observe that with small heterogeneity tolerance and with contrasted objects, the transition zone (more or less, one pixel wide) between objects is isolated as an object as with the *Multilevel Thresholding Technique*

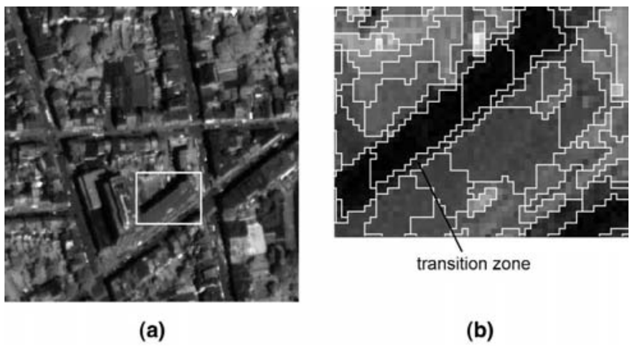

Figure 5. (a) ADM test image, and (b) the **rg3** (het = 10) ADM segmentation isolates the transition zones between contrasted objects.

(Figure 5). The heterogeneity tolerance is not high enough for the transition zone to be included in one of the two objects. When the heterogeneity tolerance increases from 10 to 15, this transition zone disappears, but the errors are too big.

When the heterogeneity tolerance increases up to 20, the over-segmentation (*Gen*) decreases as far as under-segmentation for some images. The total error remains lower than 10 percent for FOREST and RURAL test images but the average error by region (E_{NRef}) reaches 57 percent which is unacceptable.

Segmentation Comparison

For similar high-thresholds, the contour detection by "Watershed" is more effective than by the Canny-Deriche Operator from the point of view of the total error and average error by region. On the other hand, watershed segmentation produces more segments than the Canny-Deriche Operator, which could partly explain these best results. The median filter retains more contours than the Canny-Deriche optimal filter. Figure 6 shows the differences between the results of these two algorithms, the over-segmentation of the Watershed segmentation (Figure 6d) and the missing contours with the "Optimal Edge Detector" with regard to the reference segmentation (white circles in Figure 6c).

The "Watershed Segmentation" with a threshold of 5 is equivalent to the "Region-growing" segmentation with a

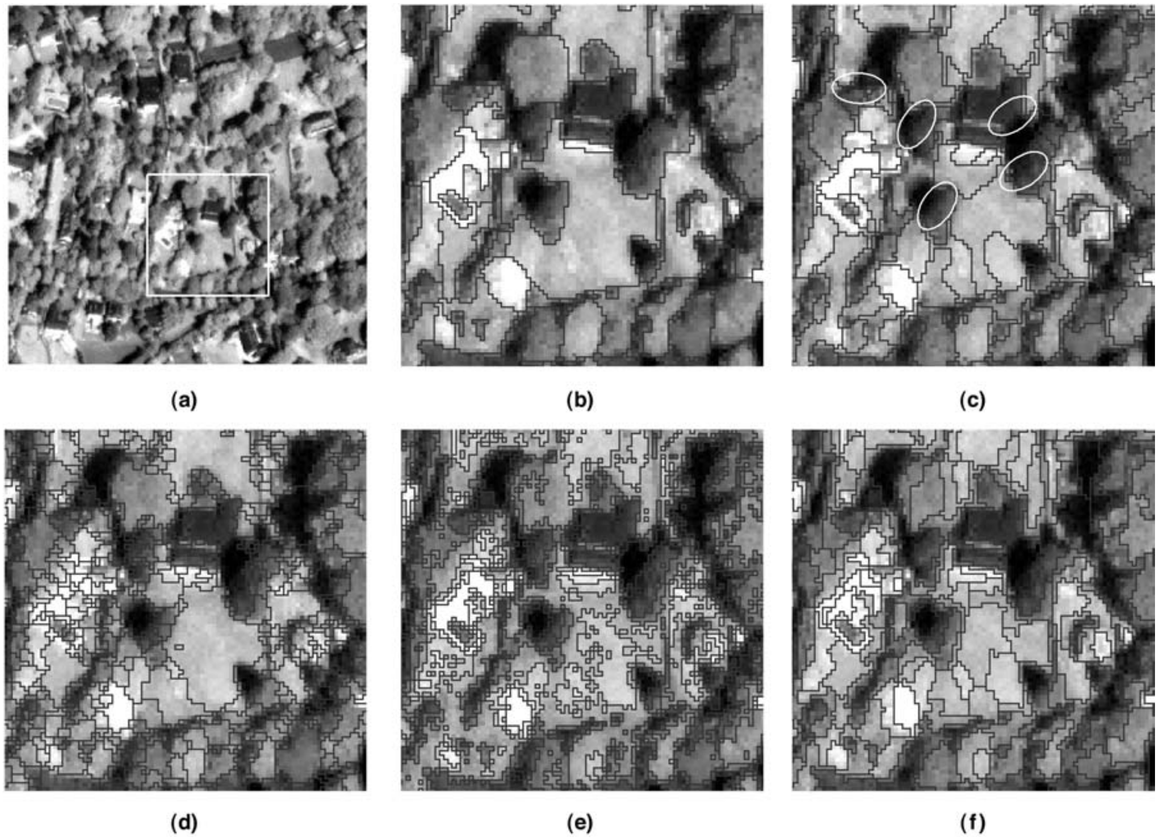

Figure 6. The best results or the best compromises between the E_{NRef} and the *Gen* evaluation measures for the four segmentation algorithms: (a) RESI test image, (b) reference segmentation, (c) Optimal edge detector segmentation (Deriche_b), the white circles show the missing contours, (d) Watershed segmentation (Watershed_1), (e) Multi thresholding segmentation (Deravi_10), and (f) Region Growing segmentation (**rg3**, het = 10).

heterogeneity tolerance of 10, but over-segmentations are higher for the "Watershed" segmentation. The results would be better with a smaller threshold, but over-segmentation would be high as is the case with a decrease of the heterogeneity tolerance in the "Region-Growing" technique.

The "Multilevel Thresholding" technique is effective for images having a histogram including peaks and valleys, which is seldom the case for very high spatial resolution satellite images, and, if the segmentation is good, it is to the disadvantage of a reasonable over-segmentation value. Figure 6e shows the high over-segmentation and the salt and pepper effect with the "Multi-thresholding" technique. The "Region-growing" method works well with textured images and images with not high contrast objects like in RURAL and FOREST test images. The results obtained with the other types of landscapes are not bad (Figure 6f). It is perhaps the method which gives a good image segmentation of any type of landscape without having high over-segmentation values, and this method does not require pre- and post-processing, such as filters or contours closing. Moreover, there is not any salt and pepper effect as with "Multilevel Thresholding" segmentation. These two segmentation methods are in the same group of methods (region-based algorithms), but the segmentation procedures are very different. The "Multilevel Thresholding" segmentation uses the global information of the image to segment it (the detection of thresholds is influenced by all pixels in the image) while the "Region-growing" segmentation uses the local information (region variance and form) to segment the image. The global information methods are more sensitive to texture (internal variability), i.e., less immune to texture than the local information based methods (Pal and Bhandari, 1993).

Conclusions

The miraculous segmentation method which segments in a correct way for all types of landscape does not exist. In each of the four used methods, the choice of the parameters (thresholds) is important and has a great influence on the segmentation results.

The contour detection methods are sensitive to noise or texture in the images, and pre-processing is essential in the majority of the cases (median filter and optimal filter). These methods prove to be effective for the detection of homogeneous and contrasted objects within the images (Janssen and Molenaar, 1995) as in the images of urban zones where these types of objects are very common (for example, buildings).

The two region-based methods are very different, and a common conclusion is not obvious. However, a problem with the transition zones between the contrasted objects can be present in both segmentation techniques. The "Region-growing" segmentation is less sensitive to texture, which is a significant advantage in the segmentation of very high spatial resolution satellite images.

All the objects in an image cannot be extracted with one segmentation without over-segmentation. If each region of the segmentation is supposed to represent one object in the image, multi-level segmentation must be applied. In each level, different objects are identified according to their characteristics (such as, texture and form). The lower levels are made up of small objects, larger homogenous objects, and pieces of the larger textured objects. The upper levels are made up of the merger of the regions of the lower levels and allow the identification of larger textured objects. The implementation of multi-level segmentation is easier with the "Region-growing" technique, as long as the heterogeneity tolerance is increased.

Acknowledgments

The Ikonos images were acquired by the OSTC (Federal Office for Scientific, Technical, and Cultural Affairs) within the framework of a feasibility study carried out by the I.G.E.A.T. (Institut de Gestion de l'Environnment et d'Aménagement du Territoire, U.L.B.) within the framework of the TELSAT4 program. The authors give special thanks to the GDR ISIS of the CNRS (Groupe de Recherche en "Information-Signal-Image-Vision" du Centre National de Recherche Scientifique français) for the access to the SIMPA software library. Alexandre Carleer would like to thank the Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture (FRIA) for its support.

References

Aplin, P., P.M. Atkinson, and P.J. Curran, 1997. Fine spatial resolution satellite sensors for the next decade, *International Journal of Remote Sensing*, 18(18):3873–3881.

Aplin, P., P.M. Atkinson, and P.J. Curran, 1999. Fine spatial resolution simulated satellite sensor imagery for land cover mapping in the United Kingdom, *Remote Sensing of Environment*, 68:206–216.

Barr, S., and M. Barnsley, 2000. Reducing structural clutter in land cover classification of high spatial resolution remotely-sensed images for urban land use mapping, *Computers and Geosciences*, 26:433–449.

Biswas, S., and N.R. Pal, 2000. On hierarchical segmentation for image compression, *Pattern Recognition Letters*, 21:131–144.

Burnett, C., and T. Blaschke, 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis, *Ecological Modelling*, 168:233–249.

Campbell, J.B., 1996. *Introduction to Remote Sensing*, pp. 622, Taylor & Francis, London.

Canny, J., 1986. A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, 6:679–698.

Carleer, A.P., O. Debeir, and E. Wolff, 2004. Comparison of very high spatial resolution satellite image segmentations, *Proceedings of SPIE Image and Signal Processing for Remote Sensing IX*, (Lorenzo Bruzzone, editor), Vol. 5238:532–542.

Carleer, A.P., and E. Wolff, 2004. Exploitation of very high resolution satellite data for tree species identification, *Photogrammetric Engineering & Remote Sensing*, 70(1):135–140.

Cocquerez, J.P., and S Philipp, 1995. *Analyse d'images: filtrage et segmentation*, Masson, Paris, pp. 457.

Cushnie, J.L., 1987. The interactive affect of spatial resolution and degree of internal variability within land-cover types on classification accuracies, *International Journal of Remote Sensing*, 8(1):15–29.

Debeir, O., 2001. *Segmentation Supervisée d'Images*, Ph.D., Faculté des Sciences Appliquées, Université Libre de Bruxelles.

Definiens Imaging, Inc., 2005. *eCognition Professional User Guide*, URL: http://www.definiens-imaging.com/ (last date accessed: 03 August 2005).

Deravi, F., and S.K. Pal, 1983. Grey level thresholding using second-order statistics, *Pattern Recognition Letters*, 1:417–422.

Ding, L., and A. Goshtasby, 2001. On the Canny edge detector, *Pattern Recognition*, 34:721–725.

Ferro, C.J.S., and T.A. Warner, 2002. Scale and texture in digital image classification, *Photogrammetric Engineering & Remote Sensing*, 68(1):51–63.

Guindon, B., 1997. Computer-based aerial image understanding: A review and assessment of its application to planimetric information extraction from very high resolution satellite images, *Canadian Journal of Remote Sensing*, 23(1):38–47.

Guindon, B., 2000. Combining diverse spectral, spatial and contextual attributes in segment-based image classification, *Proceedings of the ASPRS Annual Conference*, Washington D.C., 22–26 May, unpaginated CD-ROM.

Haralick, R.M., and L.G. Shapiro, 1985. Survey: Image segmentation techniques, *Computer Vision, Graphics, and Image Understanding*, 29:100–132.

Herold, M., J. Scepan, A. Müller, and S. Günter, 2002a. Object-oriented mapping and analysis of urban land use/cover using IKONOS data, *Proceedings of 22nd EARSEL Symposium "Geoinformation for European-wide Integration"*, Prague.

Herold, M., M. Gardner, B. Hadley, and D. Roberts, 2002b. The spectral dimension in urban land cover mapping from high-resolution optical remote sensing data, *Proceedings of the 3rd Symposium on Remote Sensing of Urban Areas*, Istanbul, Turkey.

Herold, M., X. Liu, and K.C. Clarke, 2003a. Spatial metrics and image texture for mapping urban land use, *Photogrammetric Engineering & Remote Sensing*, 69(9):991–1001.

Herold, M., M.E. Gardner, and D.A. Roberts, 2003b. Spectral resolution requirement for mapping urban areas, *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1907–1919.

Hill, R.A., 1999. Image segmentation for humid tropical forest classification in Landsat TM data, *International Journal of Remote Sensing*, 20(5):1039–1044.

Hou, Z., and T.S. Koh, 2003. Robust edge detection, *Pattern Recognition*, 36(9):2083–2091.

Irons, J.R., B.L. Markham, R.F. Nelson, D.L. Toll, D.L. Williams, R.S. Latty, and M.L. Stauffer, 1985. The effects of spatial resolution on the classification of Thematic Mapper data, *International Journal of Remote Sensing*, 6(8):1385–1403.

Janssen, L.L.F., and M. Molenaar, 1995. Terrain objects, their dynamics and their monitoring by the integration of GIS and remote sensing, *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):749–758.

Johnsson, K., 1994. Segment-based land-use classification from SPOT satellite data, *Photogrammetric Engineering & Remote Sensing*, 60(1):47–53.

Karantzalos, K.G., and D.P. Argialas, 2003. Evaluation of selected edge detection techniques in remotely sensing images, *Proceedings of SPIE Image and Signal Processing for Remote Sensing VIII*, (Sebastiano B. Serpico, editor), Vol. 4885:102–110.

Key, T., T.A. Warner, J.B. McGraw, and M.A. Fajvan, 2001. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest, *Remote Sensing of Environment*, 75:100–112.

Lillesand, T.M., and R.W. Kiefer, 1994. *Remote Sensing and Image Interpretation*, John Wiley & Sons, Inc., New York, 750 p.

Lee, S.U., S.Y. Chung, and R.H. Park, 1990. A comparative performance study of several global thresholding techniques for segmentation, *Computer Vision, Graphics, and Image Processing*, 52:171–199.

Neubert, M., and G. Meinel, 2003. Evaluation of segmentation programs for high resolution remote sensing applications, *Proceedings of the Joint ISPRS/EARSel Workshop "High Resolution Mapping from Space 2003"*, Hanover, Germany, 06–08 October, unpaginated CD-ROM.

Nicolin, B., and R. Gabler, 1987. A knowledge-based system for the analysis of aerial images, *IEEE Transactions on Geoscience and Remote Sensing*, 25(3):317–329.

Pal, N.R., and D. Bhandari, 1993. Image thresholding: some new techniques, *Signal Processing*, 33:139–158.

Thomas, N., C. Hendrix, and R.G. Congalton, 2003. A comparison of urban mapping methods using high-resolution digital imagery, *Photogrammetric Engineering & Remote Sensing*, 69(9):963–972.

van der Sande, C.J., S.M. de Jong, and A.P.J. de Roo, 2003. A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood risk and flood damage assessment, *International Journal of Applied Earth Observation and Geoinformation*, 4:217–229.

Vincent, L., and P. Soille, 1991. Watershed in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598.

Welch, R., 1982. Spatial resolution requirements for urban studies, *International Journal of Remote Sensing*, 3(2):139–146.

Woodcock, C.E., and A.H. Strahler, 1987. The factor of scale in remote sensing, *Remote Sensing of Environment*, 21:311–332.

Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation, *Pattern Recognition*, 29(8):1335–1346.

Zhang, Y.J., 1997. Evaluation and comparison of different segmentation algorithms, *Pattern Recognition Letters*, 18:963–974.

Zhang, Y.J., 2001. Texture-Integrated classification of urban treed areas in high-resolution color-infrared imagery, *Photogrammetric Engineering & Remote Sensing*, 67(12):1359–1365.

(Received 15 April 2004; accepted 29 September 2004; revised 25 October 2004)