

# Scale and Texture in Digital Image Classification

Christopher J.S. Ferro and Timothy A. Warner

## Abstract

Classification errors using texture are most likely associated with class edges, but investigators often avoid edges when evaluating texture for classification. The large windows needed to produce a stable texture measure produce large edge effects. Small windows minimize edge effects, but often do not provide stable texture measures. Simulated data experiments showed that class separability increased when texture was used in addition to spectral information. Texture separability improved with larger windows. This improvement was over estimated when pixels were chosen away from class edges. Airborne Data Acquisition and Registration (ADAR) data showed that separability of class interiors improved with the addition of texture, but that, for the whole class, separability fell. Maximum-likelihood classification of the ADAR data demonstrated the effect of edges and multiple scales in reducing the accuracy of classification incorporating texture.

## Introduction

Incorporating spatial information in image processing is a recurrent theme in the remote sensing literature. Spatial properties are so clearly an important part of human vision (Marr, 1982) that it seems intuitive that texture and context should also be important in image processing. Nevertheless, despite decades of research, a robust method for the incorporation of texture into image classification remains an elusive goal. This study is an empirical analysis of the problems of utilizing texture in classification. Major themes of this work include the issue of the scale over which texture is measured, and the context dependence of the scale of texture, particularly the problem known as the edge effect. This latter issue arises because in most texture images the major cause of variation in texture is the difference between the typically high values associated with between-class texture, and the lower values associated with within-class texture. An understanding of the problems of using texture is an important first step in developing strategies to overcome the barriers that have made utilizing texture so difficult.

This paper is organized in six sections. Following this introduction, the next section gives a short review of texture and scale. This is followed by a conceptual development, in which the terminology and main issues discussed in this paper are defined in more detail. After the description of the methods, the results are presented for simulated and then real data. Simulated data are used because imagery of known and specified spectral and spatial parameters is particularly useful in running experiments. Aircraft data are used to show the real-world

application of the relationships observed with the simulated data. Finally, the conclusions summarize the major findings of the work, and point towards solutions that may overcome the problems described.

## Image Texture

Texture, the expression of the local spatial structure in digital images, has been described in many ways. Wang and He (1990) defined texture as the tonal or gray level variation of an image. Hsu (1978) defined it as the spatial distribution of tones of the pixels in remotely sensed images. Haralick (1979) described texture as a scale-dependent phenomenon that arises from the spatial interrelationship between tonal primitives, or groups of similar pixels, which comprise the scene.

Texture is usually quantified using the statistics of image digital numbers (DN) within a window of interest, such as a 3-by 3- or 5- by 5-pixel matrix. First-order statistics, especially variance, are common measures of texture. The variance filter provides a measure of local homogeneity in an image and can be regarded as a non-linear, non-directional edge detector (Wilson, 1997). Second-order statistical measures, such as the gray-level co-occurrence matrix (GLCM) of Haralick and Shanmugam (1974), are also used as texture measures. Weszka *et al.* (1976), Connors and Harlow (1980), and Marceau *et al.* (1990) found that the GLCM and other similar measures of texture did not necessarily result in improved classification accuracy compared to first-order statistical texture measures. Therefore, a variance filter was chosen as the texture measure for this research. Variance was calculated as (Jensen, 1996)

$$VAR = \frac{1}{W} \sum_{i=0}^{quant_k} (i - AVE)^2 \times f_i$$

where

$$AVE = \frac{1}{W} \sum_{i=0}^{quant_k} i \times f_i$$

and  $f_i$  is frequency of gray level  $i$  occurring in a pixel window,  $quant_k$  is the quantization level of band  $k$  (e.g.,  $2^8 = 0$  to 255), and  $W$  is total number of pixels in a window.

Although much research has focused on developing and comparing different texture measures, the question of scale over which texture is measured (Hay *et al.*, 1997) may be a more important issue. For example, Marceau *et al.* (1990) found that,

C.J.S. Ferro is with the Center for Educational Technologies, Wheeling Jesuit University, Wheeling, WV 26003 (cferro@cet.edu).

T.A. Warner is with the Department of Geology and Geography, West Virginia University, Morgantown, WV 26506-6300 (Tim.Warner@mail.wvu.edu).

Photogrammetric Engineering & Remote Sensing  
Vol. 68, No. 1, January 2002, pp. 51-63.

0099-1112/02/6800-051\$3.00/0

© 2002 American Society for Photogrammetry  
and Remote Sensing

for classifications incorporating texture, 90 percent of the variability in the accuracy is accounted for by the size of the texture window, compared to only ten percent for the particular texture algorithm.

A number of studies have pointed out the importance of scale in texture analysis. Hodgson (1998) found that a minimum size window was needed for accurate cognitive (i.e., visual) classification, and that the window size needed increased as spatial resolution increased. Accuracy tended to increase until a threshold was reached, after which no further advantage was gained from larger windows (Hodgson, 1998). In an early attempt to develop an automated method of determination of the appropriate window size for texture calculation, Chavez and Bauer (1982) proposed a technique based on the horizontal first difference, or derivative, of an image. Curran (1988) and Franklin and McDermid (1993) used semivariograms for the same purpose. Woodcock and Strahler (1987) demonstrated the use of graphs of local variance plotted against spatial resolution (pixel size) as a tool for understanding the scale of elements in an image. In a later study, Marceau *et al.* (1994) used the same approach to derive optimal spatial resolutions for studying forested environments.

A number of contradictory conclusions have been reached about the optimum window size in texture calculation. Hsu (1978) and Dutra *et al.* (1984) suggest that small window sizes are a better choice than larger windows, due to the contaminating effect of edge pixels in classification. Dutra *et al.* (1984) also suggest that smaller window sizes preserve "microtextures." Nellis and Briggs (1989), in a study of tallgrass prairie landscape units, suggested that larger window sizes were preferable for homogeneous landscapes. Marceau *et al.* (1990) found that maximum classification accuracy was achieved with window sizes that were class specific. Relatively large windows (17 by 17 and 25 by 25) worked well for most cover types.

## Conceptual Framework

### Scenes and Images

One of the difficulties in discussing texture is that the same terms are used in many different ways in the literature, and often the underlying cause of texture is not described adequately. Therefore, in this section, a terminology for describing texture, as well as a conceptual framework for investigating its use, is presented.

For this paper, we define the *image* as an abstraction of the real world *scene* and the objects within it. One of the most important controls on the relationship between the scene and the image is the *scale of capture*, as defined by the sensor's effective resolution (normally approximated by the instantaneous field of view, or IFOV). The scale of capture determines whether a scene object is resolved in the image as a discernible element. The sensor predetermines the scale of capture, and the investigator generally has little or no control over it (Atkinson and Curran, 1997).

Informational classes, usually determined by the investigator, exist at one of the various hierarchical scales in an image and, thus, there is an *a priori* optimal scale of capture. For example, if Residential is an informational class, the scales of capture that resolve both Rooftops and Lawn as spectral classes will provide unwanted spectral variability. An optimal scale of capture for identification and mapping the boundaries of an object, ignoring image collection costs, is one that has an IFOV *smaller* than the informational class, but *larger* than the size of objects at the next finest hierarchical scale (Curran, 1988).

Strahler *et al.* (1986) define two resolution models: the H-resolution model and the L-resolution model. When the spatial arrangement of ground features can be detected directly by the sensor as individual elements, the image is said to conform to

the H-resolution model. When the sensor is unable to distinguish individual elements, because the scale of capture is larger than the elements, the image is said to have an L-resolution. The H- and L-resolution model is analogous to the Shannon sampling theorem commonly used in geophysics, in which it is shown that the rate at which a continuous function is sampled determines the maximum spatial frequency that can be discerned (Milman, 1999). While Strahler *et al.* (1986) state that H- and L-resolutions are image properties, real scenes have objects that exist at various scales, and therefore, in this paper, it is the image elements that are referred to as having H- through L-resolution characteristics. It is expected that all images have both H- and L-resolution elements. In a single image the H- through L-resolution elements may vary in size along a continuum; however, generally image elements will cluster at specific scales. Pioneering work by Marr (1982) suggested that human vision is particularly adept at generalizing the overall pattern from the multiple scales of information.

### Texture and Classification

Image textures result from an interface between the scale of capture and the scales of image elements and classes. Because there is a hierarchy of scales of image elements, a hierarchy of texture scales must exist. Investigators are generally interested only with those texture scales *within* the informational class of interest. However, the spectral differences between the classes create a *between-class* texture, which is often more distinctive than the within-class texture. Although adaptive filters that minimize the between-class texture have been proposed, the results tend to be somewhat blocky, and therefore not entirely satisfactory (Ryherd and Woodcock, 1996).

In order for image texture measures to be useful for classification, the within-class texture of each class must be relatively uniform within each class, and distinct from other classes. This requires window sizes that are sufficiently large to encompass the variability of the size and spacing of the texture elements. However, between-class texture, which tends to be highest on the interface of two or more classes, and decays with distance away from the interface, will generally degrade the performance of the classification. Thus, there is a trade-off between the large window sizes that give stable measures of within-class texture, and the increasing proportion of between-class texture that such large windows incorporate. Figure 1 shows the relationship between window size, the percentage of pixels not influenced by between-class variance, and the size of the region covered by a single class. As window size increases,

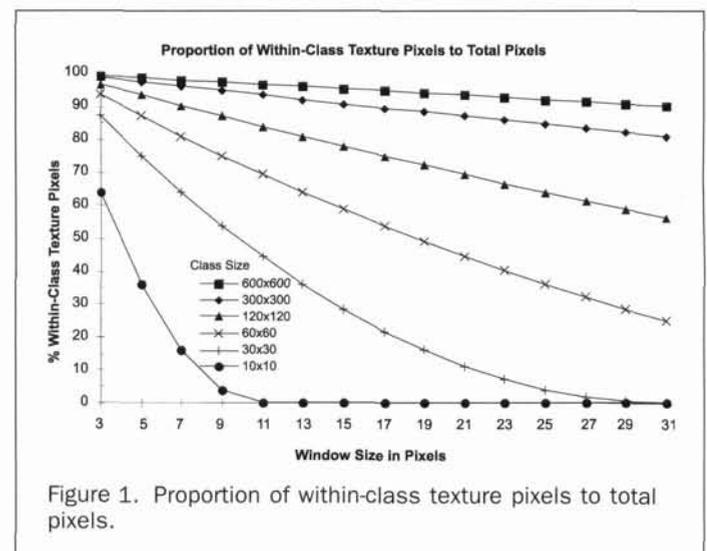


Figure 1. Proportion of within-class texture pixels to total pixels.

the proportion of within-class texture pixels to total pixels decreases. For small areas, the rate of decline is most rapid. With a window size of 5 by 5 pixels, only 36 percent of the texture pixels of a 10- by 10-pixel region are interior pixels, unaffected by between-class variance. Relatively large areas of a single class can have large window sizes but still be dominated by interior pixels. For example, even with texture calculated with a window size of 31 by 31 pixels, a 600- by 600-pixel area has 95 percent interior pixels.

However, not all class boundaries are likely to be associated with zones of misclassification caused by between-class texture. The edge effect is dependant on the difference between the class DN values across the class border. To understand this better, consider Figure 2, a cross section of a high-resolution image. Water is a relatively dark class with a smooth texture, for example. The neighboring commercial class is relatively bright, with highly variable pixel values, and a coarse texture.

Texture values calculated over a small window size remain low for the smooth class along the transect length until near the edge of the neighboring class (Figure 2). This small window, however, produces a very unstable measure of within-class variability for the commercial class because the scale is insufficient to capture the variation in the classes with coarse texture. By contrast, the large window size gives a more uniform texture value for the commercial class than do small window sizes. This comes at the cost of a broad "edge effect" for the water, which would tend to cause those pixels to be misclassified as commercial. There is no equivalent edge effect for the commercial class, because the within-class variability of this class is similar to that of the between-class variability of the two classes.

Thus, when texture is used in a standard automated classification, the scale of the window size will determine the relative location of misclassified pixels. Coarse classes have errors distributed throughout the class, but these will be progressively eliminated with larger window sizes, and little edge effect will be noticed. Smooth classes will give high accuracies for relatively small window sizes; however, even at the small window size there will be an edge effect, and thus errors will be most common on the edge of the class. As the window size increases, this edge effect will increase commensurately. Therefore, the standard, though generally unstated, practice of avoiding edges in selecting both training and testing pixels is

misleading when texture is used in a classification. Edges of classes may be poorly defined and confused by mixed pixels; however, evaluations of classification using texture will significantly underestimate the errors, particularly for relatively smooth classes unless such regions are proportionately represented in evaluation data.

The concepts of between- and within-class textures are also useful in defining an objective, context-dependent method for defining the subjective terms of smooth and coarse textures. A *smooth* texture has lower within-class variability than the associated between-class variability. By contrast, a *coarse* texture has a within-class variability that is similar to or greater than the between-class variability.

## Methods

The preceding discussion has shown that texture in a single class may vary between interior and edge pixels, and that the severity of the edge effect will also vary depending on the average DN values of the neighboring classes. These issues are investigated in the remainder of the paper through an empirical analysis of scale and texture. Both simulated and real data were resampled to give a total of six spatial resolutions. Texture was calculated from a variance filter applied to a square moving window of ten sizes ranging from 3 by 3 pixels to 21 by 21 pixels for each image. In this way, the effect of image scale and texture scale was evaluated simultaneously.

## Simulated Data

The simulated data were based on a scene model of two classes separated from each other by a background class (Figure 3). Each class was initially created as a separate sub-image with pixel values drawn from a Gaussian random distribution, with a specified mean and standard deviation (Table 1). The sub-images were expanded by pixel replication so that the original random pixel distribution was expanded to represent scene objects of specified sizes. Class A has an average DN value similar to that of the background, but is distinguishable from the background by a difference in texture. Class B has a texture similar to that of the background, but it is distinguishable by a difference in average DN value. The original simulated image, which was assumed to have a 1-meter pixel size, was resampled by pixel aggregation (Hay *et al.*, 1997) to create a set of five additional images at resolutions of 2, 5, 10, 20, and 30 meters. A

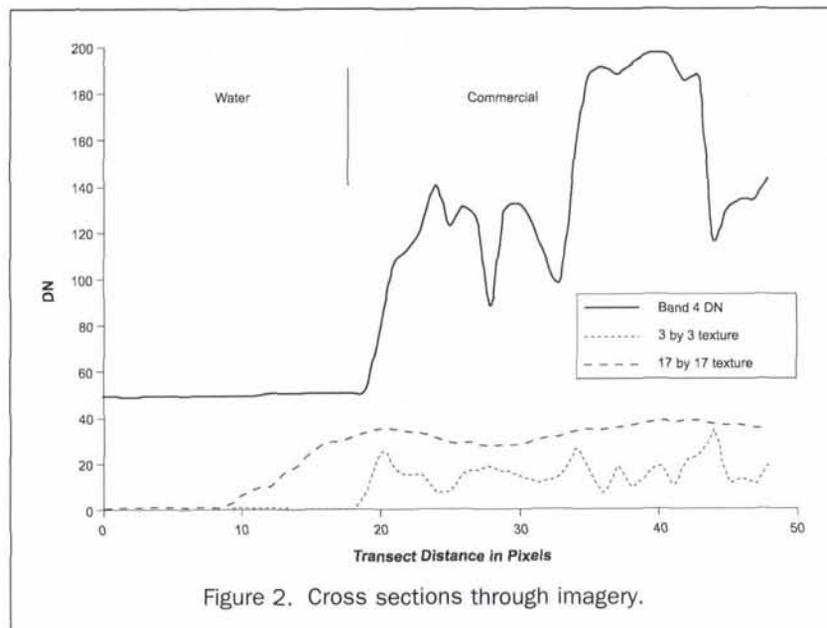


Figure 2. Cross sections through imagery.

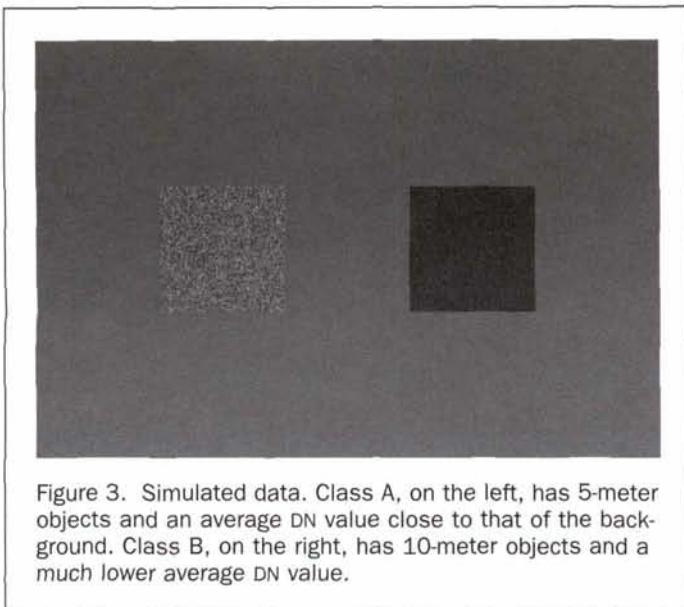


Figure 3. Simulated data. Class A, on the left, has 5-meter objects and an average DN value close to that of the background. Class B, on the right, has 10-meter objects and a much lower average DN value.

TABLE 1. CHARACTERISTICS OF THE SIMULATED DATA SET

Class	Size of Objects within Class	Mean Pixel Value	Standard Deviation	Narrowest Dimension of Class
Background	1 meter	159	4.7	600 meters
Class A	5 meter	162	28.3	600 meters
Class B	10 meter	106	5.3	600 meters

small random noise element with a mean of 15 and a standard deviation of 1 was added to each image after resampling to represent dark current noise. Because the object sizes within the background and the two classes were 1, 5, and 10 meters, respectively, resampling produced simulated images with elements ranging from H- through L-resolution. Texture analysis was performed on each re-sampled image at ten window sizes.

Class separation was expressed by a general separability index, to facilitate the application of the experiments to actual image data. The Jefferies-Matusita (J-M) Distance is an index that takes into account both the measure of distance between class means and the differences between the covariance matrices of the classes (Richards, 1993). J-M Distances range from 0, which indicates no separability, to 1.414, which indicates complete separability.

The J-M Distance,  $J_{ij}$ , is calculated (Swain and Davis, 1978) as the distance between the probability distributions of two classes,  $i$  and  $j$ ; i.e.,

$$J_{ij} = \sqrt{2(1 - e^{-B})}$$

where

$$B = \frac{1}{8} (\mu_i - \mu_j)^t \left\{ \frac{\Sigma_i + \Sigma_j}{2} \right\}^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left\{ \frac{|\Sigma_i + \Sigma_j|/2}{|\Sigma_i|^{1/2} |\Sigma_j|^{1/2}} \right\}$$

and  $\Sigma_i$  is the covariance matrix of class  $i$ ,  $\mu_i$  is the mean vector of class  $i$ , and  $|\Sigma_i|$  is the determinant of the covariance matrix of class  $i$ .

#### Real Data

To demonstrate the importance of the context-dependence of texture with real data, 1-meter Airborne Data Acquisition and

Registration (ADAR) data of the Morgantown, West Virginia area were used in a second series of experiments. The ADAR images were captured with four digital cameras (Stow *et al.*, 1996) mounted in a fixed wing aircraft. The four cameras were fitted with band pass filters to produce blue (0.44 to 0.54  $\mu\text{m}$ ), green (0.52 to 0.60  $\mu\text{m}$ ), red (0.61 to 0.69  $\mu\text{m}$ ), and infrared (0.78 to 1.00  $\mu\text{m}$ ) bands, each 1500 by 1000 pixels. After acquisition, the individual bands were automatically co-registered to produce a single, four-band image (Stow *et al.*, 1996).

The ADAR data were acquired on four dates in late March and early April 1997, prior to forest leaf out. However, lawns and pastures were already green by this time. Flight altitude was approximately 2,620 meters above mean sea level. This altitude resulted in a nominal pixel size of approximately one meter and a scene field of view of approximately 1.5 by 1.0 kilometers. The individual frames were digitally mosaicked to create a coverage equivalent to the United States Geologic Survey (USGS) 7<sup>1/2</sup>-minute Morgantown North topographic quadrangle.

Errors in band registration and sensor performance necessitated the exclusion of band 1 (blue) from the data used in the analysis. This was not expected to have a deleterious effect on the analysis, because this band is susceptible to atmospheric haze and is often highly correlated to the two other visible bands (Stow *et al.*, 1996).

Band 2 (green) was chosen for the texture calculation because of its relatively superior spectral response and contrast properties. An added benefit of using the green band is that pilot tests showed that the effect of shadows and changing illumination between image acquisition was minimal in this band. The ADAR data were resampled to five spatial resolutions ranging from 2 meters to 30 meters. This yielded a total of six images at 1-, 2-, 5-, 10-, 20-, and 30-meter resolutions. Texture was calculated for each image at ten window sizes.

In addition to the calculation of J-M Distances for evaluating class separability, various band and texture combinations were classified using a maximum-likelihood decision rule. An error analysis was performed on each classification to test for accuracy against ground reference data. Training data were extracted from the interior of classes and from the pixels adjacent to class boundaries. Ground reference data included field checking and a National High Altitude Photography (NHAP) program color infrared photograph enlarged to a scale of 1:14,500. For the accuracy testing, polygons were selected from areas that were not used in training the classifier.

## Experimental Results and Discussion

### Simulated Data

#### Texture Variability

Figure 4 displays the change in texture variability, calculated over the entire simulated image, as a function of window size and spatial resolution. The window size in pixels has been converted from pixels to meters so that window sizes can be compared between different image resolutions on the same graph. Figure 4 shows that, as images are aggregated to a coarser spatial resolution, the texture variability decreases for a window of a specified size in real world units. Exceptions to this are the 5-meter resolution data and the 30-meter resolution data. The anomalous pattern for the 5-meter resolution data may result from the fact that this window size corresponds with that of class elements in the simulated scene model.

Figure 4 also shows that texture variability tends to increase as the window size increases. This relationship occurs because with larger windows a greater proportion of the texture calculation window includes between-class texture. The effects of the between-class texture can be seen in the example texture images shown in Figure 5. The edge effect is much

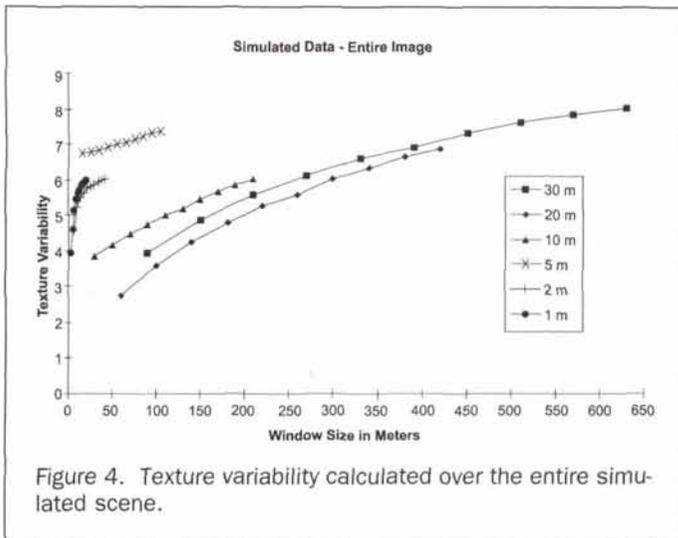


Figure 4. Texture variability calculated over the entire simulated scene.

more obvious for Class B, especially for the larger window size, because Class B has a large difference in average DN value compared to the background. Class A, however, is separable from the background only by a difference in texture, and therefore

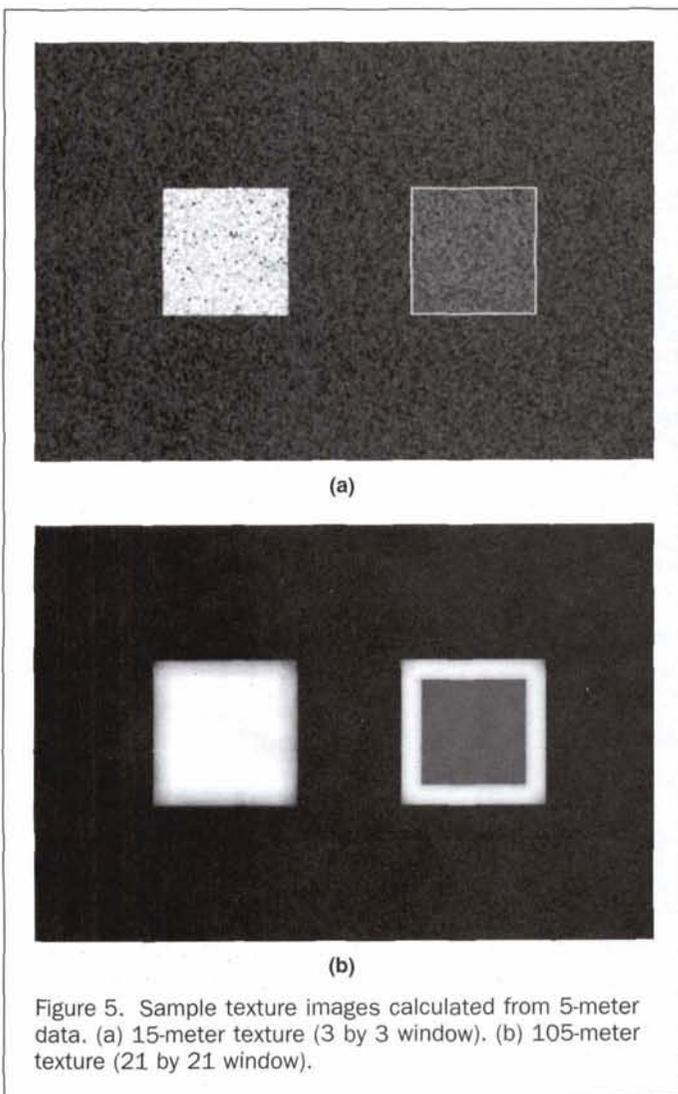


Figure 5. Sample texture images calculated from 5-meter data. (a) 15-meter texture (3 by 3 window). (b) 105-meter texture (21 by 21 window).

has no noticeable edge effect (though there is an edge effect for the background class, adjacent to Class A).

Figure 4 might be interpreted to imply that the smallest possible window size is preferable because it produces the lowest variability in texture. However, plots of average texture variability against window size can be misleading because the texture is not uniform across the scene, or even within a single class. Therefore, it is important to examine classes individually, and to compare texture from interior pixels to that of the entire class.

Figure 6a shows that texture variability for interior and edge pixels for Class A tends to be very similar. As mentioned above, this is because Class A is on average spectrally similar to the background; thus, the between-class variance is similar to the within-class variance. The initial high texture variability at relatively small window sizes occurs when the class elements are in H-resolution and the texture window is similar in size to that of the class elements. Variability declines rapidly as window size increases, leveling off in the 50- to 150-meter range. As with Figure 4, the 5-meter resolution data show anomalously high values of texture variability.

Class B has larger within-class objects but a smaller spectral variability compared to Class A. Furthermore, Class B has a mean DN value that contrasts strongly with that of the background. Consequently, the texture variability for the interior of Class B is significantly different from the variability for the entire class (Figure 6b). The variability of the texture for the entire class (including edges) increases as window size increases due to the high proportion of pixels dominated by between-class texture. At the largest window sizes there is a slight decline, because the proportion of within-class texture declines to such an extent that the overall variability in texture

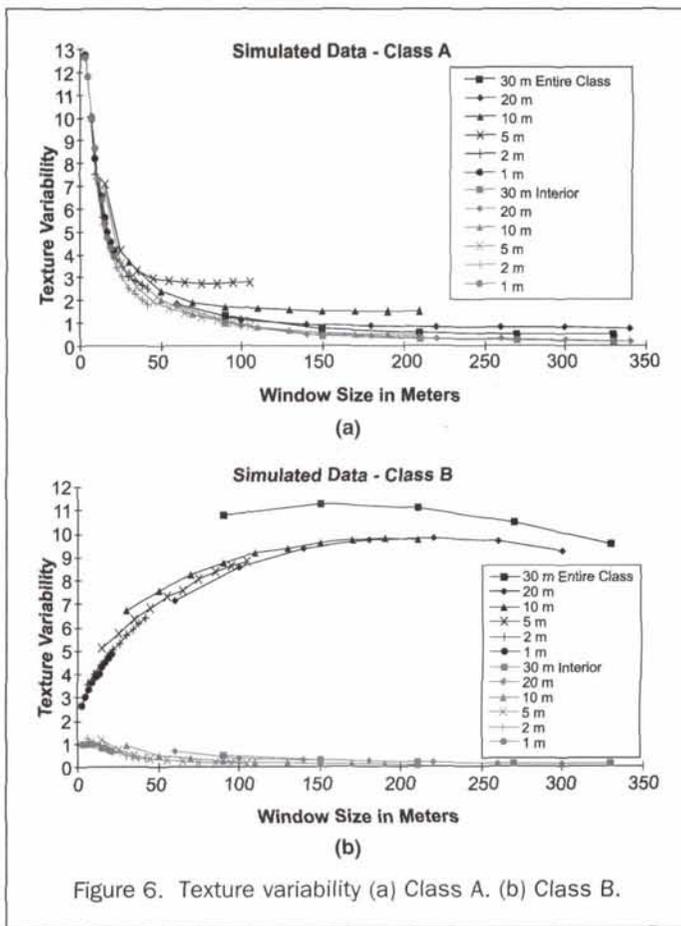


Figure 6. Texture variability (a) Class A. (b) Class B.

is reduced. However, when only interior pixels are examined, a very different pattern emerges. Without the effect of the high-variance between-class texture pixels, the texture variability for interior pixels is much lower for all window sizes. The curve shows a decline as window size increases beyond the 10-meter size of the image elements, and begins to flatten out for windows of approximately 50 meters. Variability continues to decline slowly but does not reach a constant minimum even at the largest scale that was evaluated (350 meters).

The results presented in Figure 6 are important to investigators who wish to use texture-derived data in classification schemes. First, the effect of window size is more important than image resolution in calculating texture, although image resolution determines the minimum window size possible (normally a 3-by 3-pixel window). Second, while an evaluation of pixels in the interior of the class might indicate a useful scale for calculating texture, the effect of the surrounding classes is large. In some cases, such a large portion of the class polygon may be affected by between-class variance that the incorporation of texture might not produce fruitful results in classification. This suggests that evaluations of classifier accuracy should include data from class edges; otherwise, an overly optimistic result may be obtained.

### Class Separability

The J-M separability measure, which provides a metric for predicting classification accuracy (Richards, 1993), gives a slightly more complex view of the value of texture for class A. The J-M distances were plotted for texture calculated from the simulated 10-meter data at ten window sizes (Figure 7). Ten-meter texture was chosen because, as is seen in Figures 6a and 6b, this resolution provides a good range in spatial properties for these images.

Figure 7 shows that the addition of 30-meter texture improves the separability of Class A from the background, with a J-M distance increasing from 0.880 to 1.250. After a peak at a 50-meter window size, separability falls steadily as the window size increases. However, if only interior pixels are evaluated, Class A and the background are completely separable over the scale range of 50 to 170 meters. Larger window sizes give slightly reduced separability.

The analysis so far has focused on simulated data with simple, known spatial parameters. In the next section, the evaluations of scale and texture are applied to real data to evaluate how well these methods and results can be applied to real data.

### ADAR Data

#### Texture Variability for Selected Cover Types

The Central Business District (CBD) of Morgantown, an example of the Commercial/Industrial class (Figure 8), has a within-class texture similar to the associated between-class texture, and consequently meets the criterion of a coarse textured class. Texture variability is therefore very similar for both the entire class and the interior of the class for all window sizes (Figure 9). The results match well with the trends found with Class A in the simulated data experiment (Figure 6a), except that the CBD has a pronounced increase in variability of texture with increasing window size at scales up to 20 meters. This is due to the texture window incorporating the maximum relative diversity from objects in the next finer class in the spatial hierarchy, in this case, buildings, roads, and other urban structures. This is similar to Woodcock and Strahler's (1987) finding that, when pixel size reached one-third to three-quarters the size of the scene objects, a peak in local variance occurred.

The relatively low texture variability minimum at the 3-meter texture scale occurs because the texture measure window is smaller than the elements of the next finest scale in the hierarchy. The elements within the CBD are therefore also at H-resolution at this scale of capture and texture scale. Although the average texture variability is quite low, the texture data gathered at that scale might not be useful for classification of elements at the next level of the hierarchy.

The Agricultural/Grass class (Figure 10) has a much smoother texture than does the CBD. The average texture variability of the interior pixels is much lower than the texture variability of the entire class (Figure 11), because of the pronounced edge effect for this class. Thus, the Agricultural/Grass class is similar to Class B of the simulated data (Figures 3, 5, and 6b). The peak at 45 meters for the entire class is probably related to the changing properties of interior and edge pixels.

As with the CBD data, a relatively low level of texture variability occurs at the 3-meter texture scale. Texture variability increases to a maximum at approximately 15 meters, and then falls slowly as the window size increases. It seems that, for the Agricultural/Grass class, large window sizes are needed to capture the texture, probably because of variations such as soil and topographic situation, as well as the presence of dirt roads throughout the area.

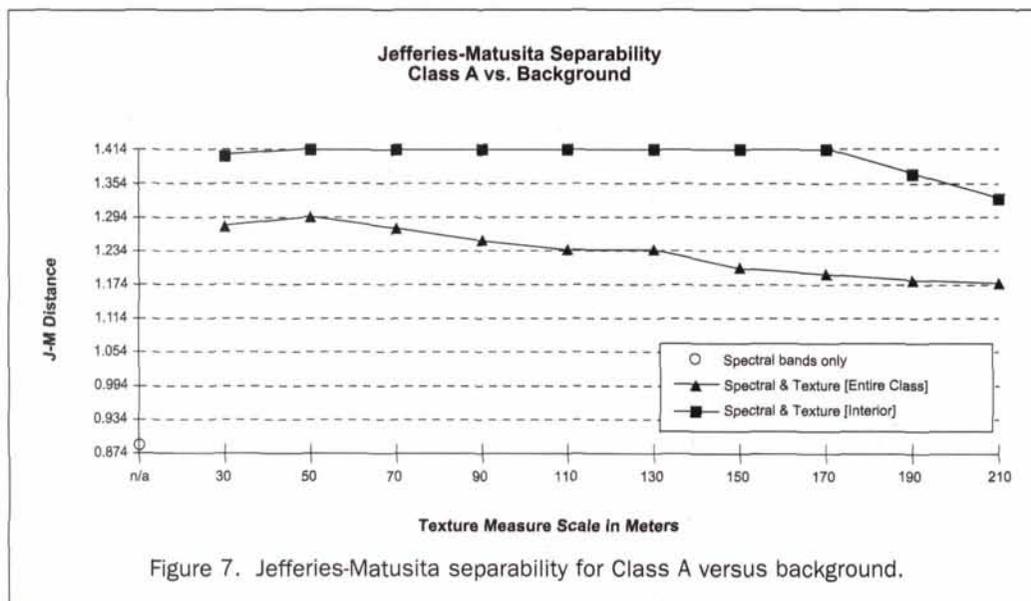
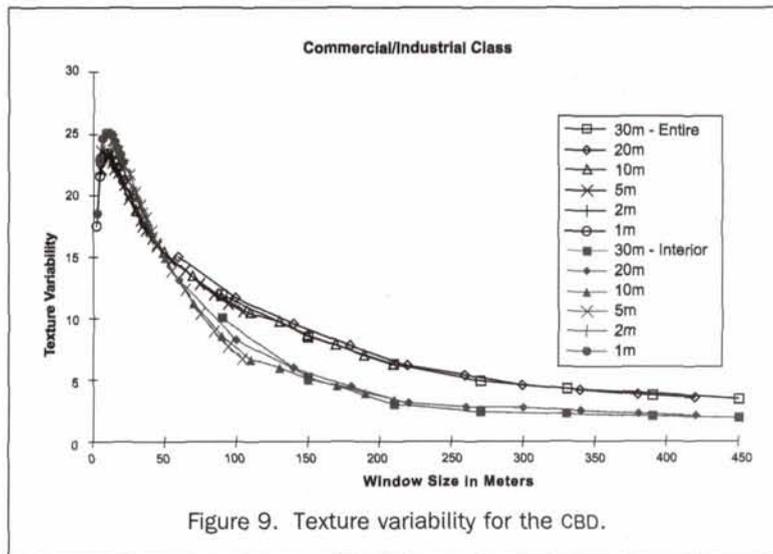
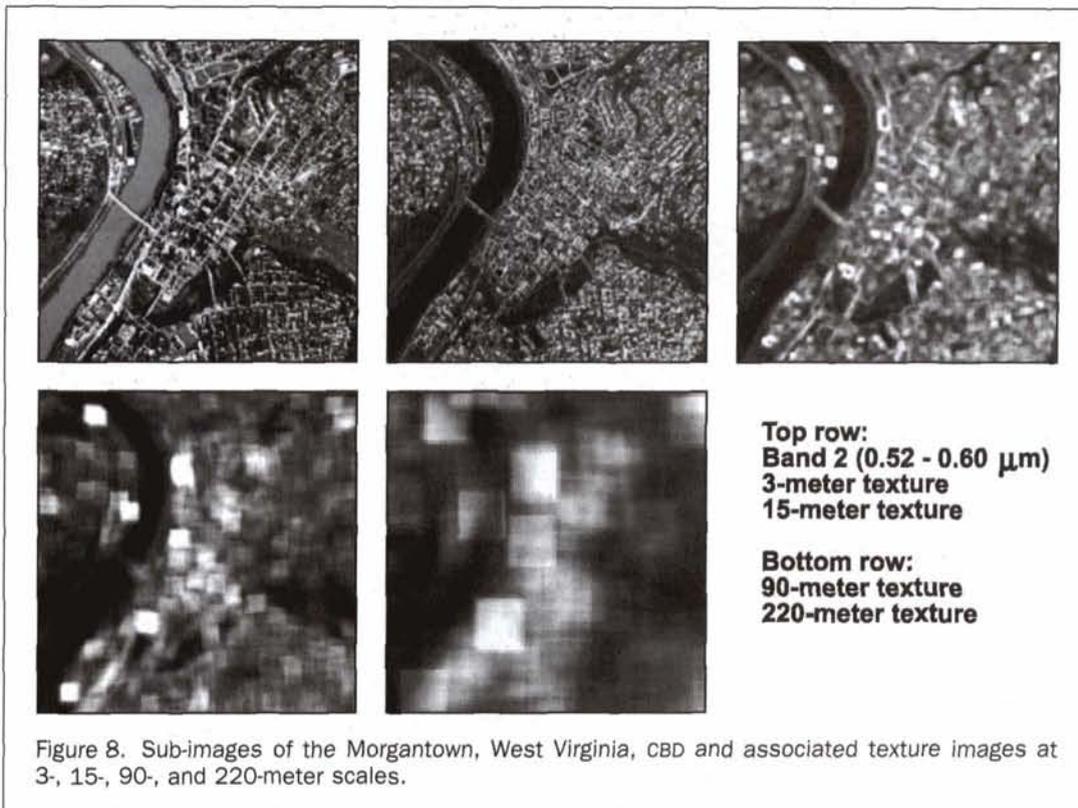


Figure 7. Jefferies-Matusita separability for Class A versus background.



### Class Separability

J-M separability graphs for the real data differ from those for the simulated data in that, with the inclusion of the edge-affected pixels, the average separability of the classes was less than that of the spectral data alone. The Forest and Commercial/Industrial classes are completely separable without the inclusion of texture data (Figure 12a) due to their highly different spectral characteristics. The inclusion of texture results in a small degradation in separability in this instance.

The Residential and Commercial/Industrial classes are potentially difficult to separate because both include similar cover types, including paved areas, buildings, and landscaped vegetation. For the interiors of these classes texture improved

separability but, when the entire classes including the edges were used to evaluate separability, it was found that texture did not improve potential classification (Figure 12b). Increasing the size of the window produced a near linear increase in separability over the 30-meter to 210-meter range studied, and thus once again larger window sizes were found to be preferable.

An even stronger example of the potentially deleterious effects of texture is illustrated by the Residential and Agricultural/Grass separability graph (Figure 12c). Without textural data, a J-M distance of 1.288 is obtained. When texture is included, separability for the class interiors increases immediately to 1.360 and continues to improve to complete separability at the 110-meter texture scale. However, if the between-class texture pixels on the class edges are included, a much

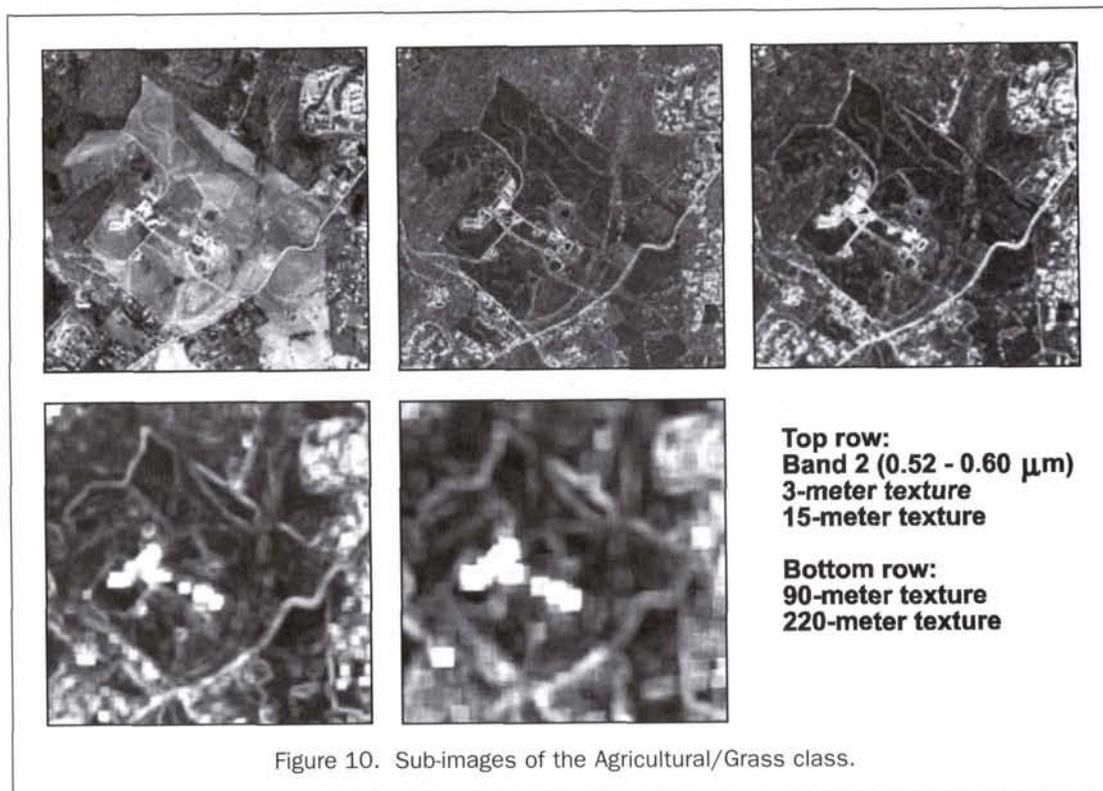


Figure 10. Sub-images of the Agricultural/Grass class.

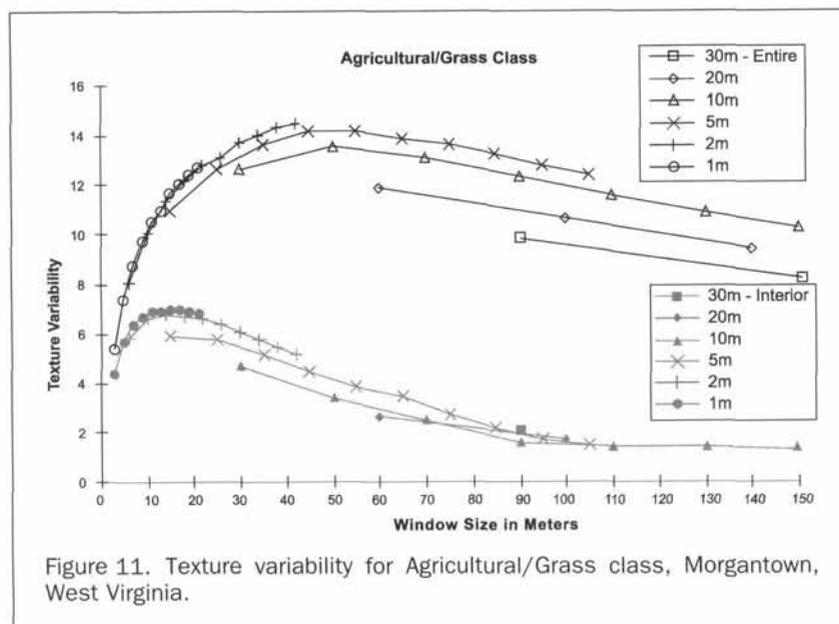


Figure 11. Texture variability for Agricultural/Grass class, Morgantown, West Virginia.

lower separability is observed. Even at the greatest texture scales, the J-M distances did not exceed 1.114.

#### Classification Results

Classification was performed to provide a more comprehensive evaluation of the effect of scale on texture for these data. Six combinations of spectral and textural data were used, with pixels assigned to one of six cover classes using a maximum-likelihood classifier (Table 2). The classification was performed on

10-meter spectral data, because it falls within the range of resolutions used by current remote sensing platforms, and a range of texture scales from 10 meters through 220 meters. This was also the scale used in the spectral separability evaluations.

The average classification accuracy was evaluated using error matrices and the Kappa statistic (Figure 13). Kappa is a measure of the difference between the observed accuracy and the random possibility of chance agreement between the reference data and the classification (Jensen, 1996). The class interiors, consisting of pixels unaffected by between-class variance,

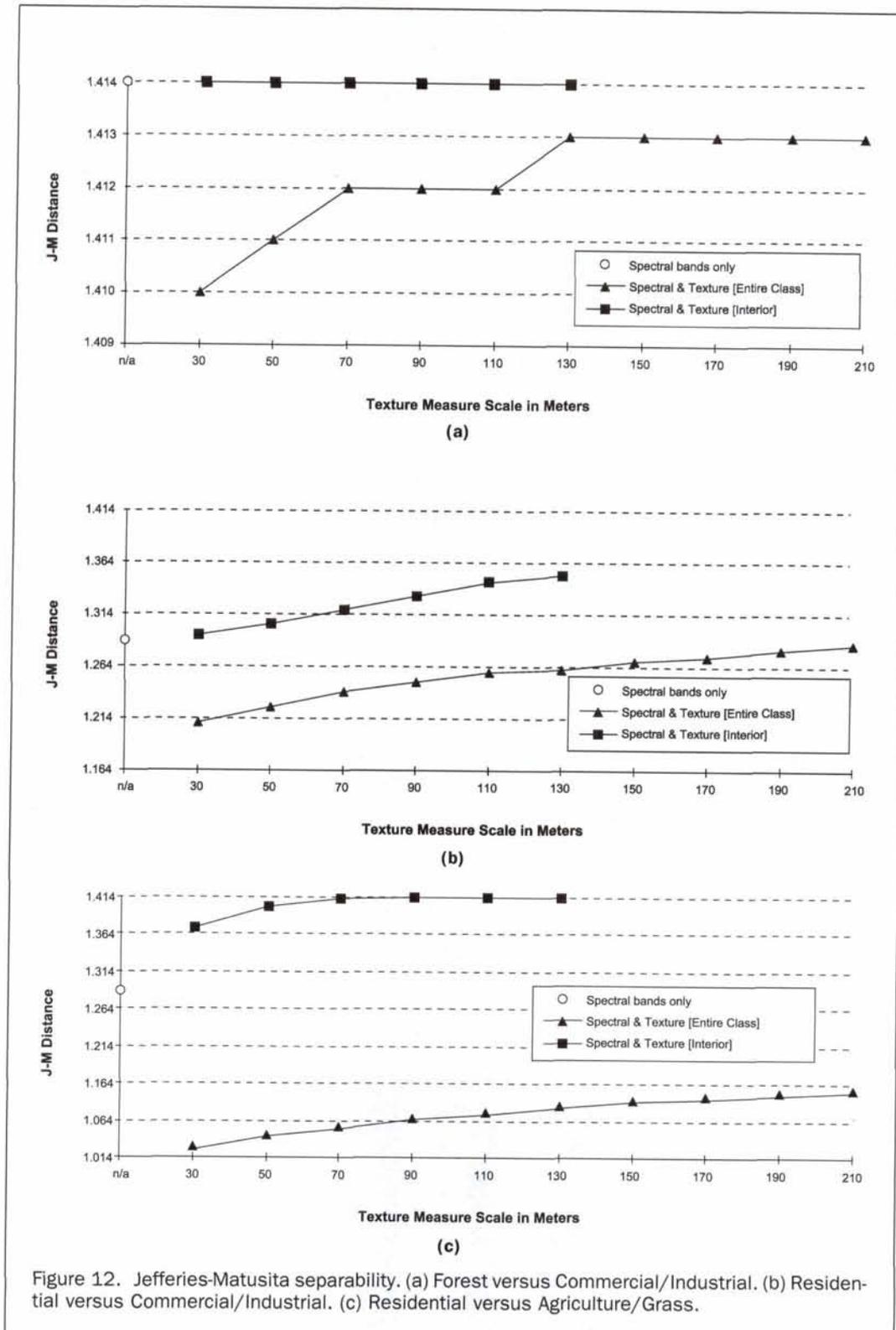


Figure 12. Jefferies-Matusita separability. (a) Forest versus Commercial/Industrial. (b) Residential versus Commercial/Industrial. (c) Residential versus Agriculture/Grass.

shrink as the window size increases. The rate at which this occurs is a function of the dimensions of the class polygon (Figure 1). For the Morgantown ADAR data at window sizes greater than 110 meters, there was an insufficient number of interior pixels to evaluate classification accuracy for pixels unaffected by between-class variance, because the texture calculation window approached the size of the class polygons.

Band combinations that included texture resulted in slightly higher classification accuracies than did those that did not, with the exception of the combination using 110-meter texture. The greatest level of accuracy including pixels near the edges of classes was found with bands 4, 3, and 2 with 210-meter texture data. However, an increase in just ten meters in window size resulted in a decrease in Kappa to 0.789 for class

TABLE 2. MORGANTOWN LAND-COVER TYPES CLASSIFIED FROM THE ADAR DATA

Class	Range of cover types
Water	Rivers, ponds
Forest	Deciduous forest, evergreen forest
Agricultural/Grass	Crops, pasture, fields
Residential	Homes, lawns, trees, roads, trailer parks
Commercial/Industrial	Businesses, factories, roads, parking lots
Surface Mining Related	Strip mines, coal piles/storage

interiors and 0.715 for interior and edge pixels. This decrease in accuracy at 220 meters suggests that the scale at which texture is calculated should be chosen very carefully, because even with large windows a small change in window size can result in large differences in classification accuracy.

Though useful for understanding how texture scale relates to the error assessment of an entire image, a comparison of selected classes at various texture scales provides added insight when determining optimal texture scales for image classification. In addition, *producer's accuracy* and *user's accuracy* (Story and Congalton, 1986) were also calculated for the individual classes. Producer's accuracy measures the likelihood an area of known cover type will be assigned to the correct class in the classification. User's accuracy is the likelihood a classified pixel represents the correct class.

*User's Accuracy*

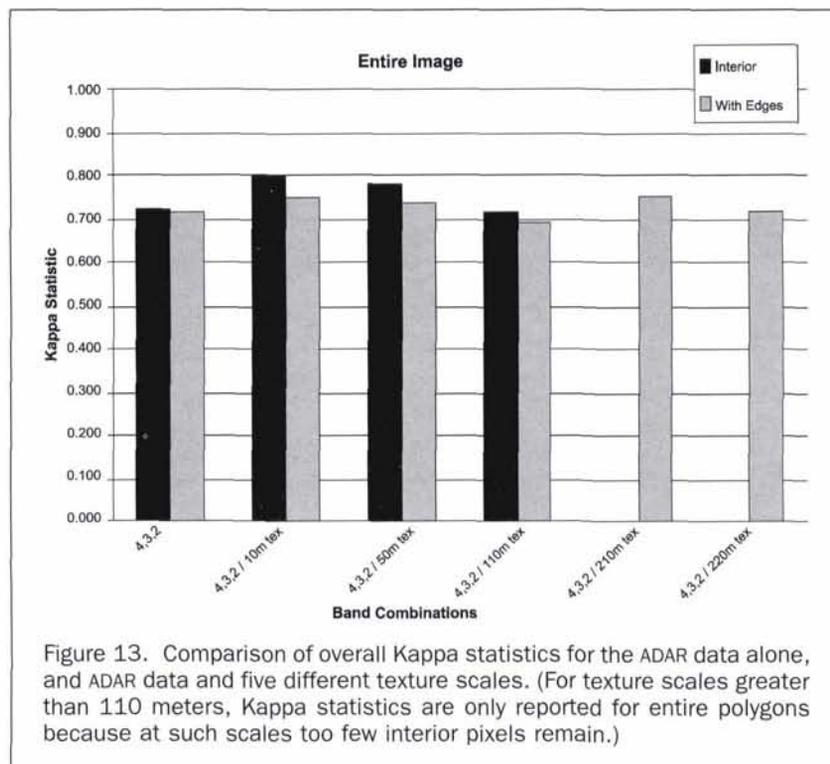
As suggested by the texture variability plots, classes that have a low within-class texture compared to the associated between-class texture, and thus a relatively smooth texture, should show degradation in classification accuracy when the test pixels include areas near the edges of the class. This is confirmed in Figure 14, where interior pixels generally have higher Kappa statistics than do the values calculated from entire class polygons. In Figure 14a, the user's accuracy for the

Agricultural/Grass class increases from the 10-meter to the 50-meter texture scales and then decreases at 110 meters. For the interior of the class, this is somewhat unexpected, given the texture variability minimum at 90 meters in Figure 11. At window sizes greater than 110 meters, there is an insufficient number of pixels that are unaffected by between-class variance to evaluate the class interiors. The accuracy for the entire polygon increases at the 210-meter scale, before declining again at 220 meters. The 210-meter texture scale of maximum classification accuracy for the entire polygon is large, considering the relative smoothness of this land-cover type and the relatively small size of the class dimensions. At these large scales, most of the measured texture arises from the between-class variance. The pattern of peaks and troughs in this graph suggests the complexity of multiple scales in real land-cover classes.

The Commercial/Industrial class, an example of a relatively coarse-textured class, shows little change in Kappa with the inclusion of pixels near the edges of the test polygon (Figure 14b). Classification accuracy generally increases as window size increases across the entire graph, though the increased accuracy from the inclusion of texture is small for texture scales up to 110 meters. The 220-meter texture scale is the maximum scale used for evaluating classification accuracy; thus, it is unknown if accuracy would improve further with the addition of texture at larger scales. All band combinations that include texture resulted in higher Kappa statistics than did those with spectral data alone. The between-class texture seems to be more valuable than within-class texture for very large window sizes (210- and 220-meter scales).

*Producer's Accuracy*

The producer's accuracy graphs for the Agricultural/Grass class (Figure 15a) more clearly illustrate the effect of edge pixels on the Kappa statistic. As with the graphs of user's accuracy, there were insufficient pixels unaffected by between-class variance



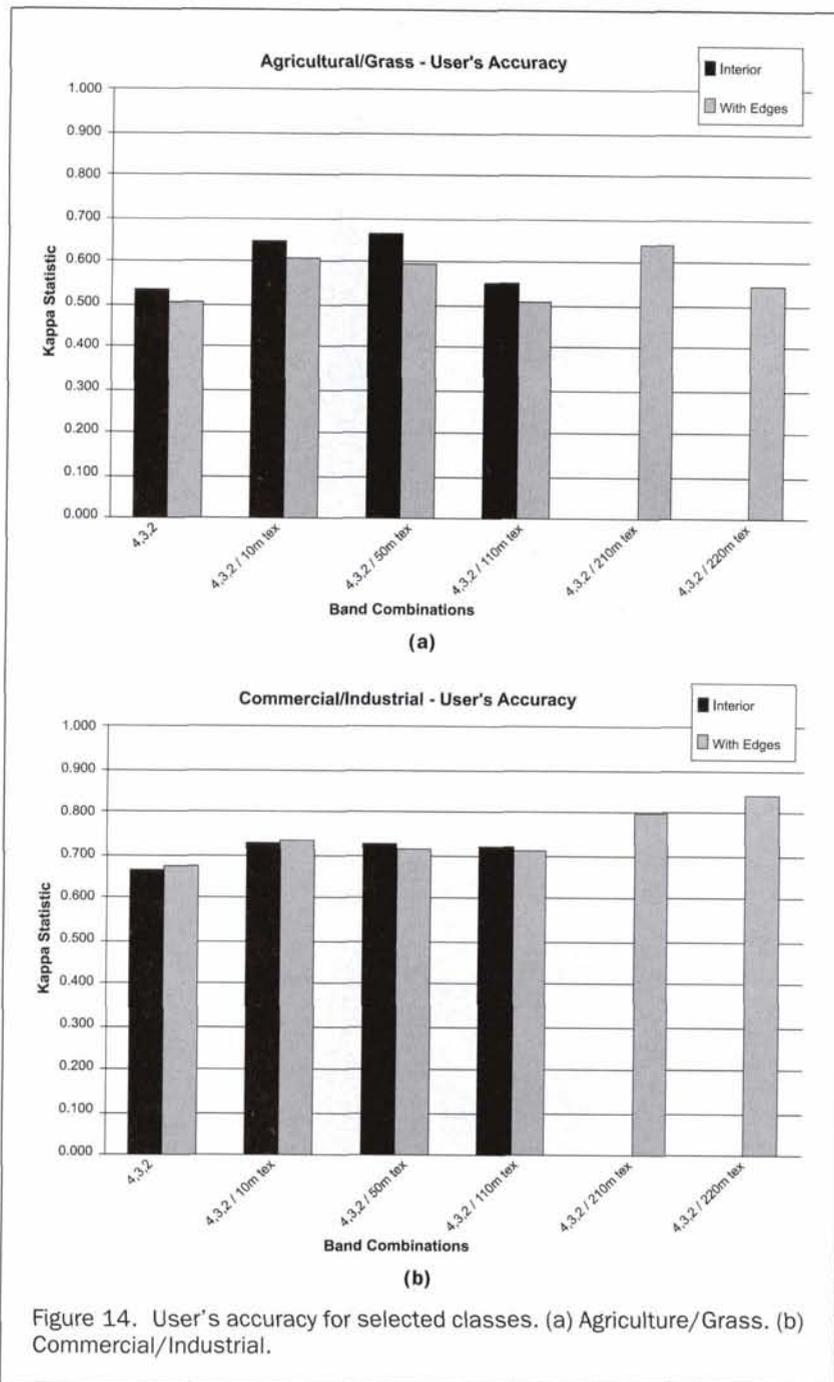


Figure 14. User's accuracy for selected classes. (a) Agriculture/Grass. (b) Commercial/Industrial.

to evaluate class interiors at window sizes greater than 110 meters. The inclusion of texture data improved classification results at the 10-, 50-, and 110-meter scales for the interior of Agricultural/Grass classes (Figure 15a). For the entire polygon, however, the value of texture was marginal for small to medium size windows, and decreased sharply for windows greater than 110 meters. This provides strong evidence for the edge effect, and supports the notion that this phenomenon is expected to be most clear in smooth classes adjacent to more coarse textured classes.

The graph for the Commercial/Industrial class producer's accuracy (Figure 15b) shows little differentiation between interior and entire class accuracies. This supports the earlier discussion in which it was argued that larger relatively coarse classes do not develop noticeable edge effects within the

classes. The value of texture for the commercial/Industrial class for the producer's accuracy (Figure 15b) is highly scale dependant. Kappa increases with the addition of texture at the 10-meter scale then drops for the 50- and 110-meter scales, increases at 210 meters, before finally dropping slightly at the 220-meter scale. It is unexpected that the interior pixels show a decline in accuracy with larger scales of texture, because Figure 9 suggests that texture variability for the CBD is at its greatest at the 10-meter scale. The multiple-scale objects found in the Commercial/Industrial class might explain the apparent trough in Kappa at the 110-meter textural scale.

### Conclusions

The theoretical framework presented here provides a basis for discussion and explanation of the concepts and relationships

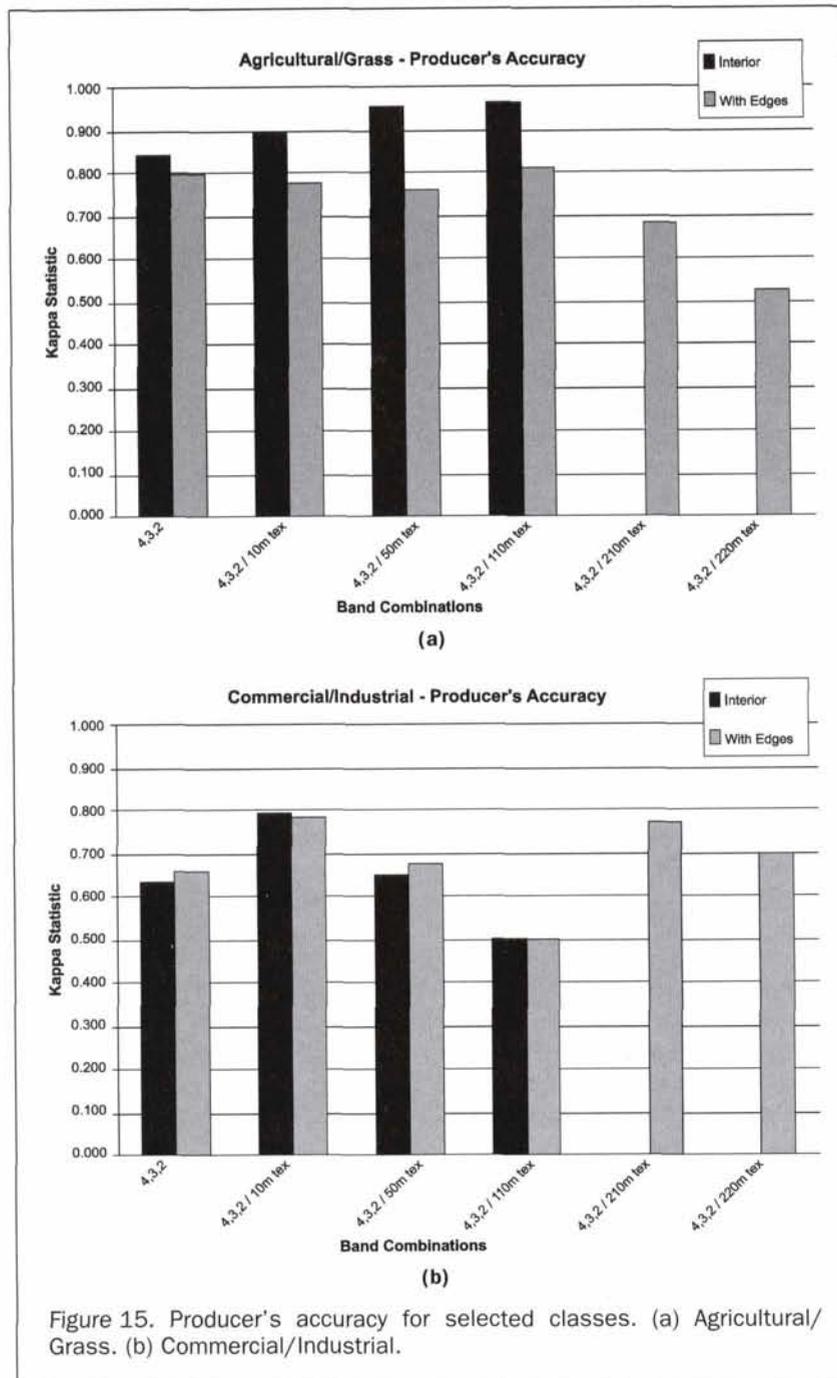


Figure 15. Producer's accuracy for selected classes. (a) Agricultural/Grass. (b) Commercial/Industrial.

involved in image texture. Through these concepts, it is possible to clarify the role of scale in digital image classification and to evaluate the usefulness of texture. In conceptualizing scene objects as H- and L-resolution elements representing those objects, the multi-scale aspects of images can more adequately be investigated and analyzed. Because the use of image texture for classification depends on the informational classes chosen by the investigator, special emphasis must be given to hierarchies of scales and how the image elements relate to the classes. A relatively small texture window would be inappropriate for texture feature extraction in commercial land-use classes. The texture such a window would capture would not relate to the size of the class elements (large buildings and parking lots) and would not aid in class separability. Small texture windows would aid in separating members of the next finer classes in the

hierarchy, the buildings and parking lots themselves. Conversely, a relatively large texture window would be inappropriate for texture feature extraction in a suburban area if houses were the land-class type. Such a window would capture the texture of the residential land-use class and would not assist in the separation of houses, lawns, and driveways. The concepts in this framework were used in developing the simulated and real image data experiments.

The results of the simulated data experiment show that per-class data are more useful for texture evaluation than are averages for an entire scene. A clear relationship was seen between window size for texture calculation and class separability. Class A had a coarse texture with a high within-class variance compared to the background; thus, the variability of texture dropped to a minimum value for both interior pixels

and the entire class polygon. This is contrasted by Class B, which had a smooth texture with a high between-class variance. When pixels near the edges were included in the analysis, the texture variability remained high and only dropped once the number of edge-affected pixels dominated the class statistic.

While plots of texture variability and tests on simulated data show the promise of finding the appropriate window size for texture feature extraction, the results of the real data experiment reveal the inherent complexities involved in interpreting and analyzing actual image data. Multiple peaks in classification accuracy help to illustrate this complexity. Although objects within classes exist at multiple scales, shapes, and orientations, it is possible to make some general statements about the role scale plays in image texture.

The window sizes necessary to achieve a stable, relatively homogeneous measure of texture are often quite large even for relatively smooth informational classes such as Agricultural/Grass. For relatively coarse classes, such as Commercial/Industrial, the window size required might well be larger than the dimensions of the class polygon. While classification often results in a large increase in accuracy for pixels far from class edges, the accuracy tends to fall for pixels near the edges of classes due to between-class variance. Because a useful application of textural information requires the inclusion of these edge pixels, the large window sizes necessary for achieving the stable texture measures result in a large number of pixel values being influenced by between-class variance. The investigator must weigh the potential usefulness of texture against the added confusion of the between-class variance pixels. The size of the image elements, the class polygon dimensions (Figure 4), and the spectral properties of the classes and elements adjacent to the class of interest must be considered.

Measures of producer's accuracy and user's accuracy showed the effect of high between-class variability in image classification. Large window sizes for some classes resulted in improved producer's accuracy for the same class, because no pixels were omitted from the class. The same windows resulted in lower user's accuracy for the same classes because the effect of between-class variance on the edge pixels caused many of these pixels to be placed in an incorrect category due to their high calculated texture values.

## References

- Atkinson, P.M., and P.J. Curran, 1997. Choosing an appropriate spatial resolution for remote sensing investigations, *Photogrammetric Engineering & Remote Sensing*, 63(12):1345-1351.
- Chavez, P., Jr., and B. Bauer, 1982. An automatic optimum kernel-size selection technique for edge enhancement, *Remote Sensing of Environment*, 12:23-38.
- Connors, R.W., and C.A. Harlow, 1980. A theoretical comparison of texture algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(3):204-222.
- Curran, P.J., 1988. The semivariogram in remote sensing: An introduction, *Remote Sensing of Environment*, 24:493-507.
- Dutra, L.V., and N.D.A. Mascarenhas, 1984. Some experiments with spatial feature extraction methods in multispectral classification, *International Journal of Remote Sensing*, 5(2):303-313.
- Franklin, S.E., and G.J. McDermid, 1993. Empirical relations between digital SPOT HRV and CASI spectral response and lodgepole pine (*Pinus contorta*) forest stand parameters, *International Journal of Remote Sensing*, 14(12):2331-2348.
- Haralick, R.M., 1979. Statistical and structural approaches to texture, *Proceedings of the IEEE*, 67(5):786-804.
- Haralick, R.M., and K.S. Shanmugam, 1974. Combined spectral and spatial processing of ERTS imagery data, *Remote Sensing of Environment*, 3:3-13.
- Hay, G.J., K.O. Niemann, and D.G. Goodenough, 1997. Spatial thresholds, image-objects, and upscaling: A multiscale evaluation, *Remote Sensing of Environment*, 62:1-9.
- Hodgson, M.E., 1998. What size window for image classification? A cognitive perspective, *Photogrammetric Engineering & Remote Sensing*, 64(8):797-807.
- Hsu, S., 1978. Texture-tone analysis for automated land-use mapping, *Photogrammetric Engineering & Remote Sensing*, 44(11):1393-1404.
- Kramer, H.J., 1996. *Observation of the Earth and its Environment. Survey of Missions and Sensors*, Springer Verlag, Berlin, 960 p.
- Jensen, J.R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice Hall, Upper Saddle River, New Jersey, 316 p.
- Marceau, D.J., P.J. Howarth, J.M. Dubois, and D.J. Grattton, 1990. Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):513-519.
- Marceau, D.J., D.J. Grattton, R.A. Fournier, and J. Fortin, 1994. Remote sensing and the measurement of geographical entities in a forested environment, 2., The Optimal Spatial Resolution, *Remote Sensing of Environment*, 49:105-117.
- Marr, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman and Company, San Francisco, California, 397 p.
- Milman, A.S., 1999. *Mathematical Principles of Remote Sensing: Making Inferences from Noisy Data*, Sleeping Bear Press, Chelsea, Michigan, 406 p.
- Nellis, M.D., and J.M. Briggs, 1989. The effect of spatial scale on Konza landscape classification using textural analysis, *Landscape Ecology*, 2(2):93-100.
- Richards, J., 1993. *Remote Sensing Image Data Analysis*, Springer Verlag, New York, N.Y., 340 p.
- Ryherd, R., and C. Woodcock, 1996. Combining spectral and texture data in the segmentation of remotely sensed images, *Photogrammetric Engineering & Remote Sensing*, 62(2):181-194.
- Story, M., and R.G. Congleton, 1986. Accuracy assessment: A user's perspective, *Photogrammetric Engineering & Remote Sensing*, 52(3):397-399.
- Stow, D., A. Hope, A.T. Nguyen, S. Phinn, and C.A. Benkelman, 1996. Monitoring detailed land surface changes using an airborne multi-spectral digital camera system, *IEEE Transactions on Geoscience and Remote Sensing*, 34(5):1191-1203.
- Strahler, A.H., C.E. Woodcock, and J.A. Smith, 1986. On the nature of models in remote sensing, *Remote Sensing of Environment*, 20:121-139.
- Swain, P.H., and S.M. Davis, 1978. *Remote Sensing: The Quantitative Approach*, McGraw Hill, New York, N.Y., 369 p.
- Wang, L., and D.C. He, 1990. A new statistical approach for texture analysis, *Photogrammetric Engineering & Remote Sensing*, 56(1):61-66.
- Weszka, J.S., C.R. Dyer, and A. Rosenfeld, 1976. A comparative study of texture measures for terrain classification, *IEEE Transactions on Systems, Man, and Cybernetics*, 6(4):269-285.
- Wilson, P.A., 1997. Rule-based classification of water in Landsat MSS images using the variance filter, *Photogrammetric Engineering & Remote Sensing*, 63(5):485-491.
- Woodcock, C.E., and A.H. Strahler, 1987. The factor of scale in remote sensing, *Remote Sensing of Environment*, 21:311-332.
- Woodcock, C.E., and S.L. Ryherd, 1989. Generation of texture images using adaptive windows, *Technical Papers, ASPRS/ACSM Annual Convention*, 02-07 April, Baltimore, Maryland, 2:11-22.
- Woodcock, C.E., A.H. Strahler, and D.L.B. Jupp, 1988a. The use of variograms in remote sensing, I: Scene models and simulated images, *Remote Sensing of Environment*, 25:323-348.
- , 1988b. The use of variograms in remote sensing, II: Real digital images, *Remote Sensing of Environment*, 25:349-379.

(Received 06 November 2000; accepted 27 April 2001; revised 31 July 2001)