Frequency-Based Contextual Classification and Gray-Level Vector Reduction for Land-Use Identification

*Peng Gong** and *Philip J. Howarth*
Earth-Observations Laboratory, Institute for Space and Terrestrial Science, Department of Geography, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

ABSTRACT: Attempts to map land use directly from higher spatial resolution satellite data with conventional computer classification techniques have proven to be ineffective. This is due to two facts. First, land use is a cultural concept. What we see on remote sensing imagery is only the physical evidence of land use as represented by combinations of land-cover types. Second, conventional classifiers employ only spectral information on a single-pixel basis. A large amount of spatial information is thus ignored.

In this research, a contextual classification method was developed to obtain land-use information. The number of gray-level vectors in multispectral space was reduced using a new data-reduction algorithm through rotating multispectral space into eigen space. As a result, the multispectral data were reduced to images of one feature dimension with the loss of relatively little information. Each gray-level vector-reduced image was then used in the frequency-based procedure to derive land-use information.

These land-use classification procedures were tested using SPOT HRV data obtained over part of the rural-urban fringe of Metropolitan Toronto, Canada. Best overall classification accuracies (measured by the Kappa coefficients) obtained using the three procedures were 0.616 when a classification scheme with 14 land-use classes was used. These accuracies are significantly better than an accuracy of 0.462 which was obtained using the maximum-likelihood classification method. The contextual classifier developed proved to be very efficient in terms of computation.

INTRODUCTION

THERE ARE SEVERAL CONTEXTUAL CLASSIFICATION METHODS which can be used to classify an image. They make use of additional information, as well as the multispectral information from a classification unit. Commonly used contextual classification methods use spatial features derived from spectral imagery in combination with the spectral bands of an image (e.g., Jensen, 1979; Dutra and Marscarenhas, 1984; Franklin and Peddle, 1990; Gong and Howarth, 1990b; Marceau *et al.*, 1990). Spatial information may be extracted either directly or indirectly from a pixel neighborhood (or pixel window) on a spectral image. For computational simplicity, a square pixel window is often used. Therefore, the first step in a contextual classification is the determination of a pixel window size. Once an appropriate pixel window size has been determined, a large number of statistical measures can be employed to obtain spatial features (Hsu, 1978; Haralick, 1979). This type of contextual classification approach usually requires a feature-selection procedure after the spatial features have been produced. There are several drawbacks to this method. First, only one spectral image can be used each time to extract spatial features. Second, the spatial feature derivation and feature selection require a considerable amount of computation and disk storage. In addition, there are no effective ways to conduct spatial feature selection.

Other types of contextual classification include the compound-decision method, and the relaxation-labeling method. The compound-decision method represents a large group of neighborhood-based classifiers which attempt to classify a pixel using not only its own values, but also those of its neighbors based on stochastic modeling. Following the development of a simple model based on the assumption of conditional independences in feature space for each pixel in a pixel neighborhood (Welch and Salter, 1971), there have been a variety of more complicated stochastic models proposed for modeling pixel neighborhoods (Landgrebe, 1980; Fu and Yu, 1980; Swain *et al.*, 1981; Chittineni, 1981; Tilton *et al.*, 1982; Owen, 1984; Haralick and Joo, 1986; Mohn *et al.*, 1987). Although some initial tests of both simulated data and small data sets of Landsat MSS imagery showing agricultural and forest environments have been made in some of the studies cited above, and improved results have been obtained in these studies, little research has been done to evaluate these approaches in an urban environment with higher spatial resolution data sets and higher spatial frequencies in the land-cover types. The usefulness of these compound-decision methods, which are frequently based on unrealistic assumptions, is suspect.

The relaxation-labeling method includes four different schemes: a discrete model, a fuzzy model, a linear probabilistic model, and a non-linear probabilistic model (Rosenfeld *et al.*, 1976). Different from the compound-decision method, the relaxation-labeling method is an empirical procedure which updates iteratively the class memberships (or class probabilities) for a pixel using class memberships from its neighbors. As the number of iterations increases, class memberships from more neighboring pixels will be incorporated, and therefore more contextual information will be involved in the classification of a particular pixel. Among the four schemes, the non-linear probabilistic model, known as the probabilistic-relaxation method, has been extensively modified to improve its performance with respect to both accuracy and computation time (Richards *et al.*, 1982; Haralick, 1983; Kalayeh *et al.*, 1984; Gong and Howarth, 1989a; 1989b; Duncan and Frei, 1989; Kittler and Hancock, 1989). Similar to the spatial feature methods, relaxation methods are computationally complicated. They are not very effective in improving classification accuracies when applied to classify urban land covers and land uses (Gong and Howarth, 1989a; 1989b).

In this paper, we focus on the development, implementation, and evaluation of a new contextual classification algorithm for

*Presently with the Department of Surveying Engineering, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

land-use information extraction. The method involves gray-level vector reduction and the use of gray-level vector-occurrence frequencies to characterize and classify land-use types. The objectives of this study are

- To develop a computationally efficient contextual classification algorithm which employs spatial information from a multispectral image; and
- To assess the contextual classification method with major emphasis on its ability to improve land-use classification accuracies of SPOT HRV multispectral (XS) data of a rural-urban fringe environment.

The study site is the Town of Markham which is situated on the northeastern rural-urban fringe of Metropolitan Toronto, Canada. The classification accuracy is measured by the Kappa coefficient.

FREQUENCY-BASED CLASSIFICATION

Occurrence Frequencies as a Surrogate for Spatial Features

Occurrence frequency, $f_l(i,j,v)$, is defined as the number of times that a pixel value v occurs in a pixel window centered at i,j. For computational simplicity, the pixel window has a square shape with a lateral length of l ($l>1$). For a single-band image, v represents a gray level. For multispectral images, v represents a gray-level vector. Within each pixel window, one can obtain an occurrence-frequency table containing all possible vs.

When a pixel window of a given size is moved all over an image(s), one can generate a frequency table for each pixel in the image(s), except for those pixels close to the image boundary. Those pixels within a distance to the image boundaries of half the lateral length, l, of the pixel window are called boundary pixels. Because full frequency tables cannot be obtained at boundary pixels, these pixel positions should be avoided in further analysis. To assure a small proportion of boundary pixels, the pixel window sizes used must be considerably smaller than the image size.

The number of occurrence frequencies in a frequency table increases linearly as the number of gray levels in an image increases, and exponentially as the number (or dimensionality) of spectral bands increases. For a single-band image quantized into n gray levels, one can produce gray-level occurrence frequency tables with a maximum number of n frequencies in each table. The maximum number of frequencies in a frequency table will increase to n^m when m spectral bands having the same number of gray levels are used. It requires a large amount of random access memory (RAM) in a computer to handle the n^m frequencies. For this reason, efficient gray-level vector-reduction algorithms are needed. One such algorithm will be introduced later in this paper. Frequency tables can be generated from gray-level vector-reduced images.

There are several advantages to using frequency tables when compared with the use of spatial statistical measures, as in spatial feature methods. First, a frequency table contains more spatial information than many statistical measures. For instance, the most commonly used statistical measures such as the mean, standard deviation, skewness, kurtosis, range, and entropy can all be derived from a gray-level frequency table. The following relationships show how the above mentioned spatial statistical measures are calculated from a frequency table, $f_l(i,j,v)$:

$$\text{mean} = \frac{1}{N_w} \sum_{v=0}^{N_v-1} f_l(i, j, v) \cdot v$$

$$\text{std} = \sqrt{\frac{1}{N_w - 1} \sum_{v=0}^{N_v-1} f_l(i, j, v)(v - \text{mean})^2}$$

$$\text{skewness} = \frac{1}{(N_w - 1)\,\text{STD}^3} \sum_{v=0}^{N_v-1} f_l(i, j, v)(v - \text{mean})^3$$

$$\text{kurtosis} = \frac{1}{(N_w - 1)\text{STD}^4} \sum_{v=0}^{N_v-1} f_l(i, j, v)(v - \text{mean})^4$$

$$\text{range} = \text{Maximum Gray Level} - \text{Minimum Gray Level}$$

$$\text{entropy} = \frac{1}{N_w} \sum_{v=0}^{N_v-1} f_l(i, j, v) \cdot \ln\left(\frac{f_l(i, j, v)}{N_w}\right)$$

where N_v is the number of gray levels or gray-level vectors; $N_w = l \times l$ is the size of the pixel window used. The above equations indicate that additional computation is required to obtain statistical parameters after the frequency tables are produced. Therefore, it becomes unnecessary to use statistical measures because frequency tables can be quickly computed, directly compared, and analyzed. The second advantage is that the feature-selection procedure, which is used to evaluate statistical parameters, is no longer needed because frequency tables contain more spatial information required for the classification than the above statistical parameters. Third, disk storage is not required by frequency tables due to the simplicity of their real-time creation.

Criterion for Pixel Window Size Selection

The success of the frequency-table method in land-use classification depends largely on the appropriate pixel window size being selected for frequency-table generation. If the window size is too small, sufficient spatial information cannot be extracted to a frequency table to characterize a land-use type. If the window size is too large, much spatial information from other land-use types could be included.

There seems to be no effective criterion for selecting pixe window sizes. Parametric feature-selection criteria such as various divergence measures (Thomas *et al.*, 1987) and the probability trend curve (Gong and Howarth, 1990d) do not work for frequency tables, simply because frequency tables are not parametric. Driscoll (1985) examined frequency means obtained from training samples using pixel windows with a range of successive sizes. The minimum window size was selected at which frequency means begin to stabilize in comparison to frequency means extracted from larger window sizes. Comparisons were made visually. This method is, however, based only on within-class variances. Because inter-class variances are as equally important as within-class variances for a classification, as in discriminant analysis, a separability measure which accounts for both within-class and inter-class variabilities is proposed for use in this study.

For a particular pixel window size $l \times l$, we can generate mean histograms for all c land-use classes using training samples. We may use $h_{lu} = (f_{lu}(1), f_{lu}(2), \ldots, f_{lu}(N_v))$, where $u = 1, 2, \ldots, c$, to denote a mean histogram for each class. Because the discriminating power of a vector comes from its capability to separate the two classes as far as possible, while permitting each class as little variation as possible, we define the separability between two classes s and t, $sep_l(s,t)$ in the following form:

$$\text{sep}_l(s, t) = \frac{1}{N_v - 1} \sum_{v=0}^{N_v-1} \frac{|f_{ls}(v) - f_{lt}(v)|}{\sigma_{ls}(v) + \sigma_{lt}(v)}$$

where σ represents the standard deviation of each corresponding frequency. The numerator on the right-hand side of the above equation reflects the inter-class deviation for a specific vector, while the denominator reflects the within-class variations for the vector. If the inter-class deviation remains un-

changed, the smaller the within-class variations and the more separable are the two classes for a given vector. On the other hand, if the within-class variations are constant, the higher the inter-class deviation and again the more separable the two classes are. The separability for all vectors is obtained by summing up all the divisions between inter- and within-class deviations.

In order to examine the separability power of a given pixel window size, an average is taken from separabilities for all possible class pairs. This is calculated from

$$\text{Sep}_l = \frac{1}{c\,(c-1)} \sum_{s=1}^{c-1} \sum_{t=s+1}^{c} \text{sep}_l\,(s,\ t)$$

where Sep_l is the average separability for all the classes with a given pixel window size. By comparing the average separabilities from different window sizes, a pixel window size is selected which has the greatest average separability.

The Classifier

The classifier used in this study is the minimum-distance classifier with the city-block metric (Gonzalez and Wintz, 1987). A city-block distance between two vectors is calculated by first obtaining a difference between every two corresponding vector elements, and then summing all the absolutes of these differences. There are two reasons for selecting the city-block distance. The first is that this distance is the simplest one in terms of computation, and therefore it could be used to handle occurrence frequencies extracted from an image with more gray-level vectors. Second, because we are comparing frequencies to make the classification decision, the use of Euclidian distance or other metrics is meaningless. In fact, some preliminary tests have been made in this study to compare the performances of the city-block metric and the Euclidian metric. Overall accuracies were on average 5 percent higher in favor of the city-block metric.

For given mean histograms of all c land-use classes, $h_u = (f_u(1), f_u(2), \ldots, f_u(N_v))$, $u = 1, 2, \ldots, c$, the city-block distance between a new histogram $h_l(i,j)$ and h_u is calculated from the following:

$$d_u = \sum_{v=0}^{N_v-1} \left| f_u\,(v) - f\,(i,\ j,\ v) \right|$$

The classifier compares all the c distances and assigns pixel (i,j) to the class which has a minimum distance to $h_l(i,j)$.

Performance Assessment

To evaluate the performance of the classification methods, two criteria are often used: final classification accuracy and the time consumed during the classification process. There are two types of time included in the "time consumed during the classification process:" the hours of human labor and the CPU times used by computers. While it is difficult to estimate accurately and compare the time consumed by human labor because of the different skill levels of different image analysts, it is relatively easy to determine and compare the CPU times required by computers. In this study, the CPU time was used as an index for the "time consumed during the classification process."

The most commonly used accuracy-assessment method is test-sample checking. It requires three steps: determination of sample size and sampling strategy, sample identification (ground confirmation) to generate reference data, and comparison of the reference data with classification results to derive classification accuracies. The first two steps are described in the experimental design section. The third step is discussed below.

For a classified image (or a map), a confusion matrix (also called an error matrix or a contingency table) can be made by comparing the classification results with reference data. In this matrix, the reference data are represented by the columns of the matrix while the classified data are represented by the rows, or *vice versa*. The major diagonal of the confusion matrix indicates the agreement between these two data sets. The confusion matrix allows various accuracy indices to be derived.

In this research, the Kappa coefficient K (Cohen, 1960) and its estimated variance $\hat{V}$ (Fleiss *et al.*, 1969) were calculated for each confusion matrix to evaluate the overall agreement between the classification results and the reference data. The Kappa coefficient was used as an overall accuracy index for each classification. It has been recommended as a suitable accuracy measure in thematic classification for representing the whole confusion matrix (Fung and LeDrew, 1988; Rosenfield and Fitzpatrick-Lins, 1986; Congalton and Mead, 1983). It takes all the elements in the confusion matrix into consideration, rather than just the diagonal elements, which occurs with the calculation of overall classification accuracy. The variance was used when significance tests were made.

For an m by m confusion matrix, let p_{ij} be the proportion of subjects placed in the i, j^{th} cell; let p_{i+} and p_{+j} be the proportions of subjects placed in the i^{th} row and j^{th} column respectively. Then, with

$$p_o = \sum_{i=1}^{m} p_{ii} \text{ and } p_c = \sum_{i=1}^{m} p_{i+}\ p_{+i},$$

the Kappa coefficient $\hat{K}$ is defined by

$$\hat{K} = \frac{p_o - p_c}{1 - p_c}$$

where p_o and p_c indicate the proportion of units which agree, and the proportion of units for expected chance agreement, respectively. With the above definition, Fleiss *et al.* (1969) showed that the most appropriate method to estimate the variance of $\hat{K}$ is

$$\hat{V} = \frac{1}{N(1 - p_c)^4} \left\{ \sum_{i=1}^{m} p_{ii}\,[(1 - p_c) - (p_{i+} + p_{+i})\,(1 - p_o)]^2 + (1 - p_o)^2 \sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij}\,(p_{i+} + p_{+j})^2 - (p_o p_c - 2\,p_c + p_o)^2 \right\}$$

To determine the difference between two $\hat{K}$ s, the significance test proposed by Cohen (1960) for comparing two classification results was adopted. With this method, the difference between two Kappa coefficients resulting from two classifications is first obtained. The square root of the sum of the variances Var between the two classifications is then calculated. A z-value can be determined by dividing the difference by the square root. A z-value greater than 2.58 indicates a significant improvement at the 0.99 probability confidence level.

In order to examine classification accuracies on a class-by-class basis, the conditional Kappa coefficient (Bishop *et al.*, 1975) was derived by comparing the classification results and the reference data. The conditional Kappa coefficient K_i is a class accuracy index which is derived from the agreement between the i^{th} row and i^{th} column in the confusion matrix for a particular land-use class i. The formula used to calculate the conditional Kappa coefficient is

$$K_i = \frac{p_{ii} - p_{i+}\ p_{+i}}{p_{i+} - p_{i+}\ p_{+i}}$$

Notations in this formula are the same as above.

EIGEN-BASED GRAY-LEVEL VECTOR REDUCTION

As explained in the above section, in order to make better use of the frequency-based classification technique, the number

of gray-level vectors in multispectral space has to be reduced. The simplest way of doing this is by compressing the number of gray levels in each band of the image. In this section, it is demonstrated that gray-level vector reduction in multispectral space is not appropriate. A more efficient method that is done in eigenvector space will be described.

Gray-Level Vector Reduction in Multispectral Space

The easiest way to reduce the number of gray-level vectors is to compress the number of gray levels in each individual spectral band. This is illustrated using two multispectral bands as an example in Figure 1. Band 1 and Band 3 in Figure 1 are hypothesized to be the green and near-infrared spectral bands, respectively. Each cell in Figure 1 corresponds to a two-dimensional gray-level vector. The shaded gray-level cells represent the magnitudes of various frequencies. The darker the shades, the higher are the occurrence frequencies in a particular cell. The white cells are empty (i.e., they have zero frequencies). The empty cells indicate that no pixel in the hypothesized two-dimensional image takes these gray-level vectors. In real multispectral space, there is a large proportion of empty cells. Should gray-level vectors be reduced, it is these empty cells that should be discarded first. Figure 1a shows the original multispectral space. Figure 1b shows the reduced multispectral space by compressing every successive two gray levels from the original bands into one gray level. It can be seen from Figure 1b that, although a number of empty cells have been removed, similar numbers of non-empty cells have also been reduced. The more bands that are used to construct multispectral space, the more empty gray-level vectors there will be. In general, it is impossible for the simple gray-level compression method to significantly reduce empty gray-level vectors in multispectral space without losing a similar amount of useful information provided by the non-empty cells.

More sophisticated gray-level reduction algorithms exist for each individual band. For instance, Sezan (1990) proposed an algorithm that locates peaks and valleys from a histogram of an image. Therefore, gray levels can be reduced without loss of much information. He tested the algorithm with some artificial images such as a face, a body, and a building. The histograms of these images have rugged appearances and therefore it is relatively easy to identify peaks and valleys on these histograms. However, the histograms of the SPOT HRV image used in this study are smooth and close to unimodal. Algorithms such as Sezan's may not prove very useful for the type of image used in this study. A more efficient gray-level vector-reduction method is therefore needed.

Gray-Level Vector Reduction in Eigen Space

Table 1 shows the covariance and correlation matrices for the SPOT HRV XS image of the study site (details of the data are presented later). It can be seen that the correlation between the two visible bands, Band 1 and Band 2, is over 0.98. This indicates that Band 1 and Band 2 are almost identical. Each of the two visible bands has a relatively lower negative correlation with the infrared band (Band 3).

To have a better understanding of the relationships between the three image bands in multispectral space, the scattergrams of each band-pair are shown in Figure 2. Each scattergram represents a projection of the gray-level vector distribution in three-

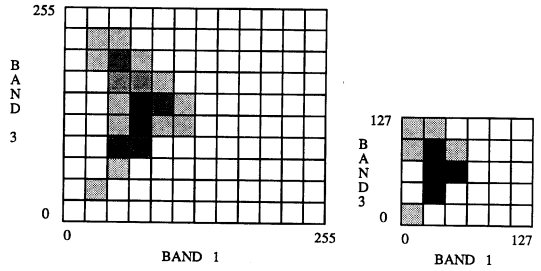

Fig. 1. Gray-level vector reduction by reducing the number of gray levels in each individual band. White cells indicate empty frequencies. The darker a cell, the higher the frequency it represents.

Table 1. Covariance, Correlation Matrices, and the Eigenvalues and Eigenvectors Derived from the Covariance Matrix

	Covariance Matrix		
	Band 1	Band 2	Band 3
Band 1	162.1904	205.1519	−70.3281
Band 2		268.9121	−112.1104
Band 3			270.7012
Means	55.6970	46.5744	81.4280
	Correlation Matrix		
		Band 2	Band 3
Band 1		0.9823	−0.3356
Band 2			−0.4155
	Eigenvectors		
Vector 1	−0.5214	−0.3484	0.7789
Vector 2	−0.6954	−0.3556	−0.6245
Vector 3	0.4946	−0.8673	−0.0569
Eigenvalues	502.4820	196.4858	2.8361

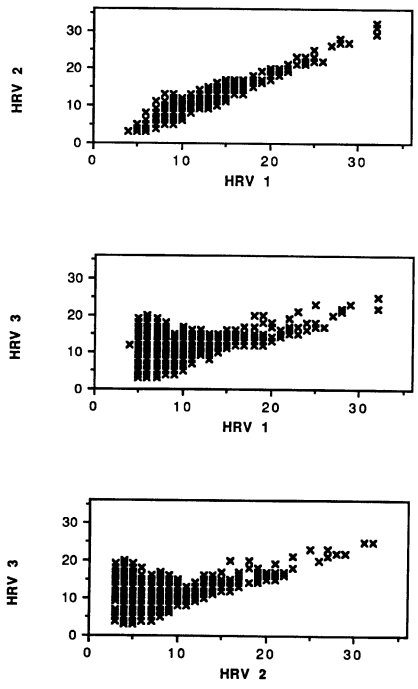

Fig. 2. Scattergrams of the SPOT HRV XS data. The original gray levels for each pair of image bands have been linearly compressed to 32 in order to reduce the amount of data for the graphics software.

dimensional multispectral space onto the plane constructed by the corresponding axes.

It can be seen in Figure 2a that the data projected on the Band 1 – Band 2 plane form in essence a thick line, while on the other planes all look like a tasselled-cap (Kauth and Thomas, 1976). These indicate that the gray-level vector distribution in three-dimensional space is approximately two-dimensional. Almost one dimension of multispectral space is wasted in the multispectral data.

Principal component (PC) transformations can reduce the data redundancy effectively through transforming the data from multispectral space to the eigenvector space of the data (e.g., Richards, 1986). From Table 1, it is noted that the first two components contain over 99.6 percent of the total variances. Because the data variability is to be preserved for discrimination purposes, only the first two principal components need to be kept. This results in a reduction of data dimension and therefore reduces the amount of data to be handled. However, simply reducing the dimensionality of the data is not sufficient if frequency tables are to be used in the classification. If the two principal components were to be quantized into 8-bit images and preserved for further analysis, the number of vectors resulting from the two 8-bit PC images would still be too big to work with. It is also inappropriate to quantize the two remaining PC images into the same number of gray levels because the proportions of variances contained in these two components are largely different. If they are quantized into the same number of gray levels, more variance in the first PC will be compressed than in the second PC. Gray-level reduction based on PC images equally quantized will cause more information loss in the first PC and a relative increase in the noise level in the second PC.

To overcome the problems caused by equal quantization, a new partition is proposed. It has the desirable performance of balancing the information loss in the eigen space and preserving the eigen structure of the original data while conducting gray-level vector reduction.

To illustrate the approach, Figure 3 shows the generalized eigen structure of the SPOT HRV data used in this study. In this figure, only the first two eigen vectors and the plane they constructed are shown. The variance in the third PC is too small to be considered. Our focus is on partitioning the preserved eigen plane. Figure 4 shows a partition of the preserved eigen plane with equal quantization of eight gray levels. It is obvious that along the second eigen axis there are too many partitions which make gray-level cells on the preserved eigen plane become rectangular. The partitioning method proposed in this study is shown in Figure 5. Because there is more variation along the first eigen axis than there is along the second eigen axis, there are more gray-level partitions on the first eigen axis than on the second axis.

The new gray-level vector-reduction scheme can now be formalized and generalized. Given the covariance matrix and the mean gray-level vector, $\mathbf{M} = (m_1, m_2, \ldots, m_k)^T$, calculated from k multispectral bands of the image, the multispectral coordinates from multispectral space S_k can be rotated into eigen coordinates in eigen space E_k (Richards, 1986). Let $V_1, V_2, \ldots, V_k$ represent the eigenvectors. A gray-level vector, $\mathbf{G} = (g_1, g_2, \ldots, g_k)^T$, in multispectral space can be transformed into a gray-level vector, $\mathbf{G}_e = (v_1, v_2, \ldots, v_k)^T$, in eigen space. This can be obtained from

$$\begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ \cdot \\ v_k \end{pmatrix} = (V_1, V_2, \ldots, V_k)^T \begin{pmatrix} g_1 \\ g_2 \\ \cdot \\ \cdot \\ \cdot \\ g_k \end{pmatrix}$$

Let $(S_{e1}^2, S_{e2}^2, \ldots, S_{ek}^2)$ represent the eigenvalues corresponding to each eigenvector. These eigenvalues are the variances along each eigen vector direction in eigen space. In order to keep the same signal to noise level between eigen axes (e.g., to make square cells on the eigen plane in Figure 5), our partition of eigen space is so designed that the number of gray levels along

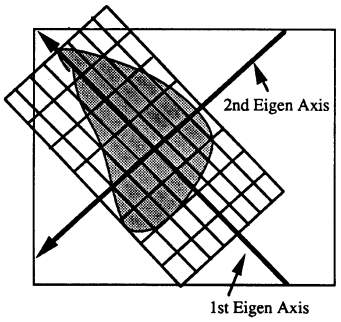

FIG. 4. Partition of the eigen space into equal gray levels along each eigen vector.

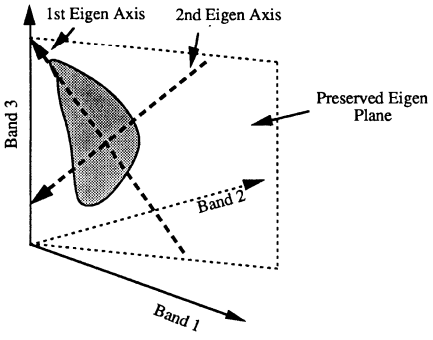

FIG. 3. Eigen structure of the SPOT HRV data used in this study.

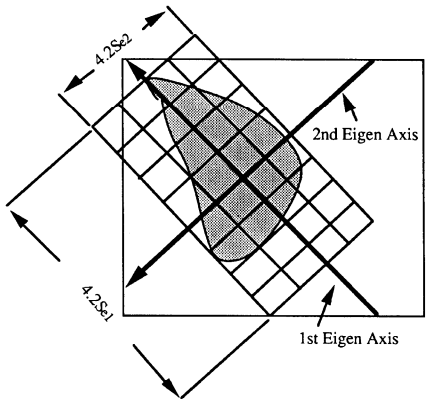

FIG. 5. Partition of the eigen space using the method proposed in this study.

each eigenvector is proportional to the square root of its corresponding eigenvalue. That is:

$$\frac{N_{e1}}{S_{e1}} = \frac{N_{e2}}{S_{e2}} = \ldots = \frac{N_{ek}}{S_{ek}}$$

where $N_{e1}, N_{e2}, \ldots, N_{ek}$ are the numbers of gray levels used for each corresponding eigenvector. As can be seen from the above equations, we have only k-1 equations but k unknown variables $N_{e1}, N_{e2}, \ldots, N_{ek}$. To determine all the k unknowns, one condition is added:

$$N_{e1} \cdot N_{e2} \cdot \ldots \cdot N_{ek} = N_E$$

where N_E is the total number of gray-level vectors to be expected for the partition of the eigen space.

To implement the eigen space partition, the origin of the eigen space (the same as the origin in multispectral space) needs to be shifted to the new origin $\mathbf{E}_o = (e_1, e_2, \ldots, e_k)^T$ which is the mean gray-level vector $\mathbf{M}$ in multispectral space. $\mathbf{E}_o$ is obtained through the following:

$$\begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_k \end{pmatrix} = (V_1, V_2, \ldots, V_k)^T \begin{pmatrix} m_1 \\ m_2 \\ \cdot \\ \cdot \\ \cdot \\ m_k \end{pmatrix}$$

Where the partition starts and how far apart each gray-level interval is along an eigen axis can now be determined. $2.1S_{ei}$ at each side of the origin along eigen axis i were selected as the starting and ending points for the gray-level partition. This was determined from the normal distribution curve by assuming data were normally distributed along each eigen axis. Based on this assumption, the use of 2.1 guarantees 97 percent of the gray-level vectors in multispectral space fall into the range $[-2.1\ S_{ei}, +\ 2.1\ S_{ei}]$ on eigen axis i. The remainder (less than 3 percent) will fall outside the range. Depending on the actual data distribution, the number "2.1" can be slightly adjusted to keep a majority of gray-level vectors falling into the specified range. The gray levels along each eigen axis are numbered in an ascending order from 0 with an increment of 1 to $N_{ei}-1$. Figure 6 illustrates the division of the i^{th} eigen axis into N_{ei} gray levels.

By dividing each eigen axis into the number of gray levels obtained above, the original multispectral space will be partitioned into N_E pieces or gray-level vectors in eigen space. The purpose, which is to reduce the large number of gray-level vectors in multispectral space, will therefore be achieved. From the transformed gray-level vector of a pixel, $\mathbf{G}_e = (v_1, v_2, \ldots, v_k)^T$, the reduced gray-level vector, $\mathbf{G}_r = (r_1, r_2, \ldots, r_k)^T$, can be obtained according to the division along each eigen axis, as described above. For example, r_1 is reduced from v_1 according to the following rule:

$$a = (v_1 - e_1 + 2.1S_{e1})(N_{e1} - 2)/4.2\ S_{e1} + 1$$

$$r_1 = \begin{cases} 0 & \text{if } a \le 1 \\ N_{e1} - 1 & \text{if } a > N_{e1} - 2 \\ a & \text{else} \end{cases}$$

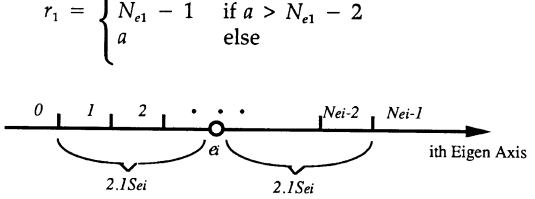

FIG. 6. Division of an eigen axis into a specific number of gray levels.

In order to allow the frequency-based classification algorithms easy access to the data after gray-level vector reduction, it was decided to use only one image to store the data. A labeling system was developed to assign a number to each gray-level vector created in eigen space. A number n_e for a particular gray-level vector, $\mathbf{G}_r = (r_1, r_2, \ldots, r_k)^T$, is calculated according to the following equation:

$$\begin{aligned} n_e = {} & r_k \cdot N_{e1} \cdot N_{e2} \cdot \ldots \cdot N_{e(k-1)} \\ & + r_{k-1} \cdot N_{e1} \cdot N_{e2} \cdot \ldots \cdot N_{e(k-2)} \\ & + \ldots \\ & + r_1 \end{aligned}$$

After this labeling, all the partitioned gray-level vectors in eigen space will range from 0 to $N_E - 1$.

In summary, it takes four steps to obtain reduced gray-level vectors using the new algorithm. In the first step, the algorithm generates the covariance matrix and mean gray-level vector from the original multispectral image by using either samples or the entire image. In the second step, the eigen values and their corresponding eigen vectors are derived from the covariance matrix. In the third step, the eigen space is partitioned into an expected number (N_E) of pieces. Finally, the gray-level values of every pixel in the multispectral image are transformed into the eigen space and each pixel is assigned a new gray-level vector number (n_e). The assignment is done according to the section (new gray-level vector) in the partition of the eigen space into which the transformed coordinates of each pixel fall.

EXPERIMENTAL DESIGN

The proposed algorithms have been implemented with the FORTRAN 77 programming language on a VAX 11/785 computer under the VMS operating system at the Faculty of Environmental Studies, University of Waterloo. In this section, the study site and data are introduced. The design of the land-use classification scheme, the training strategy, and the test sample selection will then be described.

STUDY SITE AND DATA

The Town of Markham is located at 43° 52′ N; 79° 15′ W. This site has been used for a variety of remote sensing studies of land-cover/land-use classification and rural-to-urban land conversion over a period of several years (Martin, 1975; Martin, 1986; Johnson and Howarth, 1987; Howarth *et al.*, 1988; Martin *et al.*, 1988; Martin, 1989; Gong and Howarth, 1989a; 1989b; 1990a; 1990b; 1990c; 1990d). It provides a good study site for this research as large tracts of natural and agricultural land are being rapidly converted to urban uses. The spatial structure within the study area presents a variety of patterns. These structural patterns are considered to be useful signatures for discriminating some of the land-use types whose spectral signatures are often difficult to differentiate. The second reason for selecting a rural-urban fringe environment is that the diversity of various land uses is well suited for studying the robustness of the classification algorithms proposed in this research.

The SPOT scene (K614 J262) had both panchromatic and multispectral (XS) images acquired on 4 June 1987. In this study, only the 20-m by 20-m spatial resolution SPOT HRV XS data were used. For this research, a cloud-free subscene of 512 by 512 pixels which covers a large portion of the Town of Markham (approximately 10 km by 10 km) was selected. A black-and-white reproduction of a standard false-color composite covering part of this subscene is shown in Figure 7.

Because the selected test site is small and relatively flat, both topographic and atmospheric conditions were assumed to be homogeneous throughout the image. Based on these assump-

FIG. 7. A black-and-white reproduction of a standard false-color composite of a portion of the study area covering approximately 300 by 400 pixels (6 km by 8 km). This image for picture center K614 J262 was acquired on 4 June 1987. Fields and a golf course occupy the upper left part of the image while urban uses, a tract of agricultural land, and cleared land cover the rest of this subscene.

tions, there is little or no topographic effect on the data and any atmospheric effects on the data are homogeneous in the study area. They can be considered as a contribution from atmospheric haze. To remove homogeneous haze effects from each band of the image requires only the subtraction of a constant from each pixel value. However, this would not change the radiometric structure of the data. In addition, the subsequent classifications used in this study are statistically invariant to linear transformations. Thus, no radiometric correction was made to the image.

In addition to the SPOT HRV data, various scales of topographic maps were available for the study area. While useful for field visits, they were five to six years old and were outdated for training and test-sample selection due to recent land-use changes. For these purposes, timely 1:8,000-scale aerial photographs were available which cover the entire study site. These aerial photographs were taken in April 1987, about two months prior to the acquisition of the SPOT HRV data.

LAND-USE CLASSIFICATION SCHEME

The land-use classification scheme (Table 2) was categorized by modifying the land-use classification schemes used by the local Planning Departments of both the Town of Markham and Metropolitan Toronto. Fourteen land-use classes were included. This represents an effort to map a range of land uses at a level suitable for the SPOT data used.

Some spatial-spectral characteristics of the 14 land-use classes are briefly described in Table 2 as well. These land-use classes can be divided into two groups:

- those which are dominated by an individual spectral class on the SPOT XS image; and
- those which are composed of more than one spectral class.

Five of the 14 classes are considered to be in the first group. These are mature crop, parks, cleared land, deciduous trees, and water. All the other classes belong to the second group. According to the definitions of land cover and land use in Gong and Howarth (1990c), the first group can also be called land-cover classes while the second group consists of land uses. Because the majority of the classes in Table 2 are land-use classes and the emphasis of this study was to demonstrate the effectiveness of the frequency-based classifier in the classification of land uses, hereafter in this paper no distinction between land cover and land use is made.

SUPERVISED TRAINING

The training procedure used in this research is straightforward. In order to achieve maximum flexibility, it was decided to use a block-training strategy. The advantage of this type of training is its ease in specifying training areas. By so doing, the image analyst also implicitly identifies the spatial structure for a particular class. The shape and the size of the training block contain important clues for selecting the appropriate pixel window size to be used in generating frequency tables.

Blocks of training samples were extracted using the Dipix ARIES III image analysis software package. A polygon was drawn to enclose image blocks representative of a specific land-use class. Training sample selection was assisted by referring to the 1:8,000-scale aerial photographs.

TEST-SAMPLE SELECTION

Test samples were selected using the stratified systematic unaligned sampling strategy (SUSS) (Jensen, 1983) as a guideline. In this procedure, a random sample is picked from every 16 by 16 stratum on the image. Consequently, for the data set used in this research a total of 1024 samples was identified. Because SUSS allocates samples according to the area of each class, classes with small areas will receive too few sample pixels while classes with large areas will have too many sample pixels. In order to maintain a similar confidence level for the accuracy estimate for each class, it was decided to select approximately 30 sample pixels for each class. Therefore, instead of using all the sample pixels selected by SUSS, only part of the sample pixels were used in the selection and identification of test pixels. For those classes

with areas that are not large enough for the SUSS to allocate the desired number of pixels, additional pixels located a few pixels away from existing sample pixels were randomly selected. Again, pixel identification was aided by using the 1:8,000-scale aerial photographs. In order not to bias the accuracy assessment, the test samples were selected without overlapping with training samples.

Generation of Gray-Level Vector-Reduced Images

With the gray-level vector-reduction algorithm, the original three bands of SPOT XS imagery can be reduced to one image.

Table 2. Land-Use Classification Scheme and Spatial-Structural Characteristics of Land-Use Classes when Observed on the SPOT HRV XS Image

Code	Land-Use Class	Color	Characteristics
1	Old Urban Residential	Red	Well-landscaped residential areas where trees, lawns, driveways, and roof tops dominate
2	New Urban Residential	Green	Fewer trees when compared to old residential areas; a very regular pattern
3	Rural Residential	Blue	A low density of roof tops which are surrounded by vegetation
4	Industrial/Commercial/Institutional	Yellow	Roof tops of large buildings and little vegetation
5	Idle land	Pink	No vigorous growth of vegetation
6	New Crop and Pasture	Turquoise	Fields where vegetation does not fully cover the soil or where the surface shows moderate vegetative growth
7	Mature Crop	Orange	Fields in which high-density vegetation is growing providing high spectral reflectance in the infrared band
8	Golf Course	Light Gray	Dominated by three types of land covers: well-maintained grass, normal grass, and trees. The spectral reflectance of the well-maintained grass is very high in the infrared band
9	Parks	Dark Green	Large areas of grassland in the urban area
10	Cleared Land	Dark Blue	Denuded of vegetation and top soil showing evenly high reflectance in every band (Martin *et al.*, 1988)
11	Land Under Construction	Purple	Land where construction is underway; varied reflectance is associated with building foundations and superstructures, construction materials, and partially installed roads (Martin *et al.*, 1988)
12	Deciduous Trees	Light Blue	A few patches of forest land dominated by deciduous trees; other trees such as those scattered through old urban residential areas, parks, and valleys do not belong to this class
13	Hazard Land	Dark Red	Valley land which is composed of rivers or streams, wet grass, and trees
14	Water	Pale Green	Several relatively large water surfaces such as reservoirs and ponds

Two factors affect the resultant gray-level vector-reduced images. They are the method used to calculate the covariance matrix for constructing the eigen space, and the number of gray-level vectors specified for the output image. The covariance matrix determines the orientations of the rotated coordinate axes from the multispectral space to the eigen space. In other words, the new coordinate axes in the eigen space change as the covariance matrix changes. When the covariance matrix is calculated using all pixels in an image as sample pixels, the eigen space built by this covariance matrix does not emphasize or favor any parts of the image. When the covariance matrix is obtained from only part of the image, the eigen space enhances the information provided by those pixels whose gray-level vectors are similar to the ones in the particular part of the image. Therefore, the eigen space can enhance some information while suppressing other information in the image by means of selective sampling from the image. Selective sampling places an emphasis on specific information from the imagery which is different from a systematic or a random sampling scheme that tends to be as representative as possible of the original imagery.

The training sample was used in this study to calculate the covariance matrix (Table 3). By comparing the eigenvectors and eigenvalues in Tables 1 and 4, it can be seen that the eigen vector directions are different in the two eigen spaces, which means that the eigen axes in the two eigen spaces are different. The total variance in the training sample is much smaller than that of the systematic sample. This means that the eigen structure built by the training sample enhances the differences among those pixels having gray-level vectors similar to those selected in the training sample. There is a slight difference between the means in these tables which corresponds to origins in the eigen spaces built by the two sampling strategies.

The number of gray-level vectors specified for the output image determines how much detail from the original image is to be preserved in the output image. Too few gray-level vectors will cause more inter-class confusion, while too many gray-level vectors will cause high within-class variation. In addition, the larger the number of gray-level vectors specified, the more computationally expensive it is for the frequency-based classifier. Five sets of gray-level vectors, 10, 20, 30, 40, and 50, were tested. For convenience, the resultant images are named IMG10T, IMG20T, IMG30T, IMG40T, and IMG50T, respectively. Figure 8 displays an example of a gray-level vector-reduced image (IMG40T). It should be noted that the gray levels in these gray-level vector-reduced images lose their original physical meaning after the gray-level vector reduction, and that these images were used as intermediate results for land-use classification only.

The CPU time required to produce a gray-level vector-reduced image in the eigen space was approximately 50 seconds. The

Table 3. Covariance Matrix Derived from the Training Sample and its Eigenvalues and Eigenvectors

	Covariance Matrix		
	Band 1	Band 2	Band 3
Band 1	55.7356	88.6321	−72.8320
Band 2		143.2875	−121.1418
Band 3		154.2422	
Means	51.1682	40.4893	87.0765
	Eigenvectors		
Vector 1	−0.3974	−0.3839	0.8335
Vector 2	−0.6475	−0.5264	−0.5511
Vector 3	0.6503	−0.7586	−0.0394
Eigenvalues	319.3556	33.3336	0.5762

TABLE 4. CONFUSION MATRIX AND CLASSIFICATION ACCURACIES OF THE LAND-USE MAP PRODUCED USING THE MLC*

		Classified Results														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	K_i (×100)
	1	26	8	1	1		1									65.3
	2	4	15		5	1										56.6
	3	8		5	3		3		2					3		12.2
R	4	2	2		26											84.5
e	5	2	5		6	30						1				64.9
f	6	12			16		8		3				6			14.5
e	7			1				16	3							78.8
r	8			2			2	1	18	1			2			64.1
e	9			8					17	1				1		3.2
n	10				1	1					13	17				38.4
c	11				3	3					1	12				60.1
e	12			1				4	3				7			45.7
	13	1		4		1	1	9						11		37.0
	14				8	1									6	39.1
K (×100)			46.2													
$\hat{V}$		0.000731														

*The numbers in the row and column headings indicate the classes listed in Table 2.

FIG. 8. Gray-level vector-reduced image with 40 specified gray-level vectors. The image shows exactly the same area as in Plate 1.

algorithm was designed in such a manner that it is independent of changes in the number of gray-level vectors specified.

LAND-USE CLASSIFICATION

For each gray-level vector-reduced image, ten pixel window sizes ranging from 3 by 3 to 21 by 21 were used in the land-use classification. At a specific pixel window size, a land-use classification map was created for any gray-level vector-reduced image. A land-use map was derived through

- applying the training procedure to obtain frequency averages for each land-use class; and
- classifying a gray-level vector-reduced image using the frequency-based classifier.

Therefore, ten land-use maps were produced for each gray-level vector-reduced image. This resulted in a total of 50 land-use maps.

RESULTS AND DISCUSSION

LAND-USE CLASSIFICATION RESULTS OBTAINED USING A MAXIMUM-LIKELIHOOD CLASSIFIER

For comparison purposes, a maximum-likelihood classifier (MLC) was applied to the SPOT HRV XS image; it resulted in the land-use map displayed in Plate 1. Training and test samples were the same as in the frequency-based classification. Kappa values, conditional Kappa values, and their corresponding variances have been calculated. The Kappa value is 0.462.

From Plate 1, it can be seen that all the urban land-use classes are inter-mixed. In agricultural areas, field boundaries have been classified into urban land-use classes by the MLC. The entire map looks rather fragmented. The greatest confusion is between golf courses (light gray) and parks (dark green). While the golf course class has been partly allocated to the parks and the mature crop (orange) classes, because of their similar spec-

tral signatures, the parks class has been omitted almost entirely by the MLC. Rural residential (blue) and new crop and pasture (turquoise) have also been partly allocated to the parks class by the MLC. Hazard land (dark red) has been identified as golf course. Other confusions occur between mature crops and deciduous trees (light blue), industrial land (yellow) and idle land (pink), rural residential and new crop and pasture, and between cleared land (dark blue) and land under construction (purple).

Table 4 shows the confusion matrix, conditional Kappa values, Kappa value and its variance for the land-use map displayed in Plate 1. The numbers in the row and column entries in Table 4 represent the land-use classes as listed in Table 2. From Table 4, additional confusion can be found between construction and industrial, construction and idle land, and between water and industrial. Conditional Kappa values for rural residential, new crop and pasture, parks, cleared land, hazard land, and water are low.

Land-Use Classification Results Obtained From Gray-Level Vector-Reduced Images

Table 5 shows Kappa values and their variances for land-use maps generated from gray-level vector-reduced images obtained using the training-sample method. The column entries in this table are the five gray-level vector-reduced images used in land-use classification. The row entries are the ten pixel window sizes used for each image. With ten gray-level vectors, the Kappa values at eight pixel window sizes are greater than the Kappa values derived from the MLC. At a pixel window size of 7, the classification accuracy improvement is significant at the 0.9 confidence level. At all the pixel window sizes, improved classification accuracies were achieved when compared with the results of the MLC, as 20, 30, 40, and 50 gray-level vectors were used, respectively. With IMG20T, classification accuracy improvements are significant at the 0.99 confidence level at four pixel window sizes. With both IMG40T and IMG50T, classification accuracy improvements are significant at the 0.99 confidence level at six pixel window sizes. The best classification result, 0.616, was achieved by the use of IMG40T at the pixel window size of 9 (Plate 2). This has made an improvement of 0.154 as compared with the MLC accuracy.

Table 5. Kappa Values and Their Variances Calculated from Land-Use Classification Results Derived from Gray-Level Vector-Reduced Images Using the Training Sample*

Pixel Window Size	IMG10T	IMG20T	IMG30T	IMG40T	IMG50T
3	0.491	0.487	0.558	0.586+	0.592+
	729	736	729	714	708
5	0.523	0.518	0.599+	0.610+	0.596+
	734	731	699	691	701
7	0.543	0.538	0.589+	0.594+	0.588+
	728	721	698	695	705
9	0.515	0.561+	0.600+	0.616+	0.597+
	736	710	691	679	696
11	0.523	0.539	0.589+	0.597+	0.582+
	729	713	693	691	699
13	0.521	0.530	0.550	0.574+	0.582+
	723	714	708	697	695
15	0.496	0.524	0.538	0.552	0.557
	726	721	711	707	707
17	0.468	0.518	0.524	0.544	0.535
	728	724	713	710	716
19	0.451	0.479	0.527	0.527	0.529
	726	734	715	719	716
21	0.449	0.493	0.507	0.524	0.521
	724	732	719	717	721

*Decimal numbers are Kappa values while integers are variances enlarged by 10^6.
+Indicates significance test passed at the 0.99 confidence level.
Underlines indicate column maxima which correspond to optimal pixel window sizes.

It can be seen in Plate 2 that the land-use classes appear to be very homogeneous. This land-use map looks more like product made by manual interpretation. The differences between this map and the one obtained by the MLC (Plate 1) are readily apparent. The "pepper and salt" effect of Plate 1 has been reduced dramatically. Confusion between rural and urban land-use classes has been reduced. Golf course (light gray) was classified without too much assignment of pixels to other vegetation classes. Table 6 shows the confusion matrix for this land-use map which summarizes the agreements and confusions of this map as compared with ground confirmation samples.

Average Separability Versus Classification Accuracies

For each gray-level vector-reduced image, an average separability was obtained at each pixel window size. To test the capability of the average separability in predicting optimal pixel window size, the average separabilities and their corresponding Kappa values have been plotted versus pixel window size (Figure 9).

From Figure 9(a), it can be seen that each curve has a single peak at the pixel window size of 5. Following a dip after each peak, the average separability curves gradually increase as the pixel window size becomes larger. Although the pixel window sizes selected in this research were not large enough to discover the trends of these curves as the pixel window size exceeded 21, our land-use classification results suggest that the use of even larger pixel window sizes is likely to result in lower classification accuracies. This can be seen from Figure 9(b) where the high classification accuracies mainly occur when relatively small pixel window sizes are used.

However, by comparing Figure 9(a) and Figure 9(b), there is not an exact correspondence between peaks in the separability curves and the greatest Kappa value curves. The optimal pixel window sizes are 7, 9, 9, 9, and 9 for IMG10T, IMG20T, IMG30T, IMG40T, and IMG50T, respectively. This is contrary to what was expected during the design of the separability measure. It was expected that there would be a close relationship between the separability and the classification result. This leads us to conclude that the proposed separability index is not effective in predicting the optimal pixel window size. Similarly to some suspicions voiced in other studies concerning the failure of transformed divergence as a feature selection criterion (Gong and Howarth, 1990d), it is suspected that the failure of the separability measure proposed in this study was due to the differences in composition between the training sample and the test sample. Only the training sample was involved in the calculation of the separability measures while the derivation of accuracies also involved the test sample.

The magnitudes of the separability curves seem to be related to the number of gray-level vectors in gray-level vector-reduced images. This is supported by all the curves in Figure 9. The higher the number of gray-level vectors, the lower is the average separability, given the same pixel window size. More importantly, as the number of vectors and clusters becomes higher, the separability curves become positioned closer to each other (i.e., the separability curves have an asymptotic behavior). When Kappa-value curves are examined in Figure 9(b), although not obvious, the differences between Kappa-value curves for higher numbers of gray-level vectors are smaller. Therefore, the separability curve may be used to approximately estimate the number of gray-level vectors required to obtain better classification accuracies. This is important because there is no *a priori* knowledge about the number of gray-level vectors required in a gray-level vector-reduced image which will result in a more accurate land-use map.

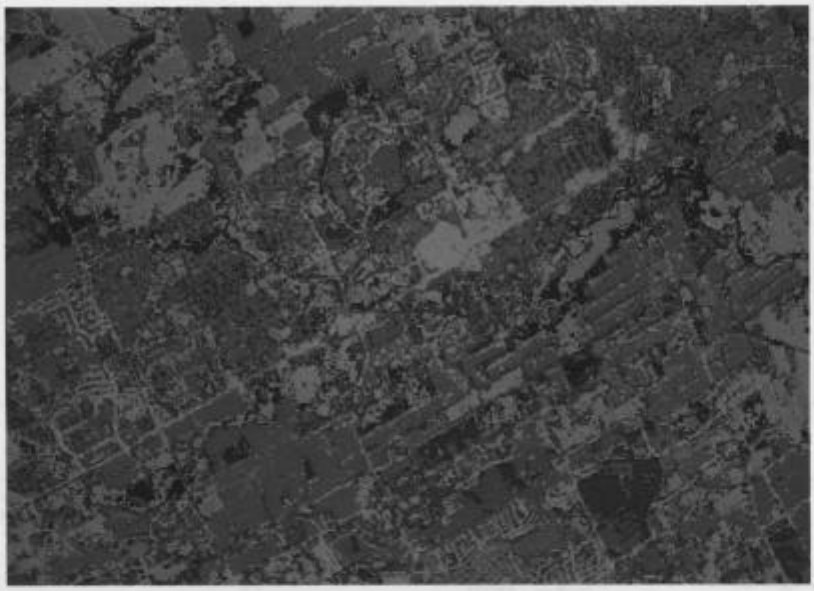

PLATE. 1. Land-use classification results obtained by applying the maximum-likelihood classifier to the SPOT HRV XS image. The image shows exactly the same area as in Figure 7. The classification accuracy measured by the Kappa coefficient is 0.462. Class names and colors are listed in Table 2.

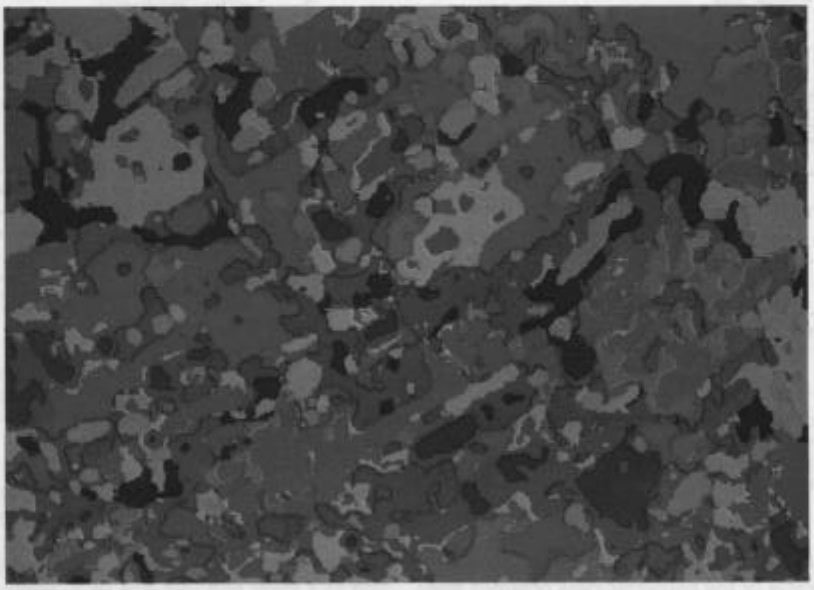

PLATE 2. Land-use map obtained from IMG40T using the frequency-based classifier with a pixel window size of 9 by 9 pixels. The image shows exactly the same area as in Plate 1. The classification accuracy measured by the Kappa coefficient is 0.616. Class names and colors are listed in Table 2.

CLASSIFICATION ACCURACY FOR EACH CLASS VERSUS PIXEL WINDOW SIZE

Figure 10 shows the conditional Kappa values obtained from IMG40T plotted against pixel window size. This allows a conditional Kappa value curve to be produced for each class. The 14 curves are grouped and displayed in three graphs. Figure 10(a) shows the three residential classes and the industrial/commercial/institutional class; Figure 10(b) shows curves for the three agricultural classes, plus deciduous trees, hazard land, and water; and Figure 10(c) shows curves for the two recreational land-use

TABLE 6. CONFUSION MATRIX FOR THE LAND-USE MAP PRODUCED FROM IMG40T WITH A PIXEL WINDOW SIZE OF 9*

		Classified Results													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
	1	30	4	1			2								
	2	1	20		1	3									
R	3	3		17			1		1					2	
e	4	2			24							4			
f	5	4	9	1		30									
e	6	2	3	13			13		3	3				8	
r	7							14	6						
e	8			1				2	22					1	
n	9	4		8					2	11				2	
c	10	1		1							24	6			
e	11				4						1	14			
s	12	1						3	3				6	2	
	13			6					3					18	
	14		2	1	1									5	6

*The numbers in the row and column entries represent the land-use classes in the same order as listed in Table 2.

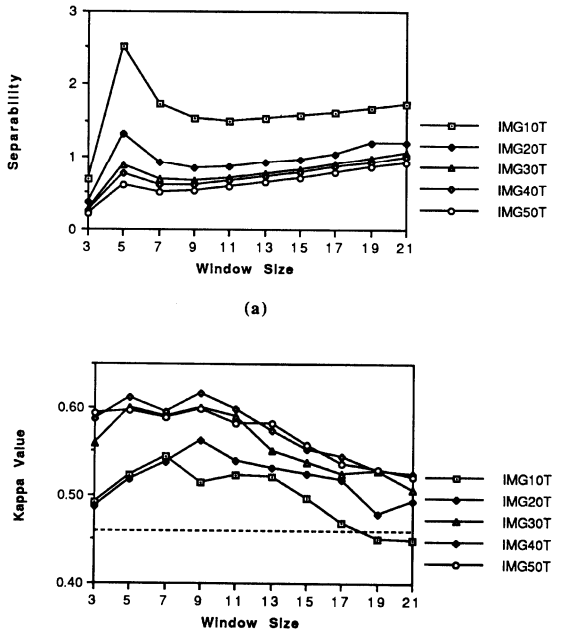

FIG. 9. Average separabilities and Kappa values for IMG10T, IMG20T, IMG30T, IMG40T, and IMG50T plotted against pixel window size; (a) separability curves, (b) Kappa-value curves. Dashed line is the Kappa value obtained with the MLC.

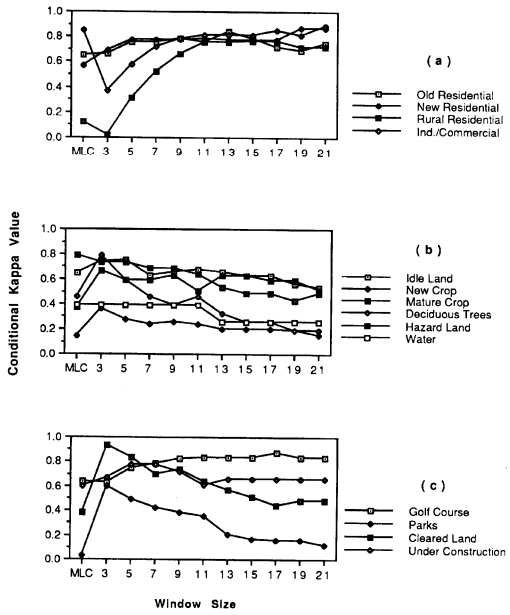

FIG. 10. Conditional Kappa values for land-use classification results obtained from IMG40T. Pixel window size "1" indicates classification accuracies obtained with the MLC.

classes and two transitional land uses. Each curve starts with the conditional Kappa value of the particular class obtained from the MLC, which serves as a reference for comparison and is plotted against the label MLC.

It can be seen from Figure 10 that different classes reach their maximum accuracies at different pixel window sizes. In Figure 10(a) maximum values are obtained at pixel window sizes of 13, 19, 15, and 17 for old residential, new residential, rural residential, and industrial, respectively, while in Figure 10(b) maxima are achieved at pixel window sizes of 3, 3, MLC, 3, 3, and MLC for idle land, new crop and pasture, mature crop, deciduous trees, hazard land, and water, respectively. It is interpreted that spatially heterogeneous classes reach their maxima at relatively larger pixel window sizes, while spatially homogeneous classes reach their maxima at smaller pixel window sizes. The contrast between the two types of land-use class can be clearly observed from Figure 10(c) in which the two spatially homogeneous classes, parks and cleared land, reach their maxima at a pixel window size of 3, while the more spatially heterogeneous classes golf course and land under construction reach their maxima at pixel window sizes of 17 and 7, respectively. The five land-use classes which achieve their maximum classification accuracies at the smallest pixel window size of 3 or 5 are those classes with only one dominant land-cover type, as mentioned above.

A trend can be observed that for spatially heterogeneous land-use classes such as urban land uses, a relatively large pixel window size is required. On the other hand, for spectrally homogeneous classes such as the five spectrally pure classes (see Table 2 and the section of the paper describing the "Land-Use Classification Scheme"), a small pixel window size is preferred.

It is interesting to note that with IMG40T the accuracies of mature crop and water have not been improved by the frequency-based classification method as compared with the accuracies obtained with the MLC. The relatively lower accuracies of the mature crop and the water achieved by the frequency-based classifier indicate that for spectrally pure classes it may be difficult for this type of classifier to obtain better classification accuracies than those of the MLC method.

The Pixel-Window Effects and Possible Solutions

In general, the frequency-based classification methods can significantly improve the land-use classification results, when compared with the conventional MLC. However, the accuracy is still relatively low. It is interesting to observe in Plate 2 that many land-use classes are partially or completely surrounded by a narrow zone of other classes. Most notably:

- agricultural fields are surrounded by old urban residential (red) and rural residential (blue);
- parts of the new urban residential (green) have been classified as idle land (pink) and industrial land (yellow) in a pattern with industrial surrounding the idle land;
- golf course (light gray) has been assigned to the boundaries between mature crop (orange) and new crop and pasture (turquoise); and
- cleared land (dark blue) surrounded by land under construction (purple).

Other errors include

- rural residential (blue) has been intermixed with old urban residential (red), new crop and pasture (turquoise), and hazard land (dark red);
- industrial (yellow) has been assigned to land under construction (purple);
- idle land (pink) has been classified as industrial land (yellow);
- new crop and pasture (turquoise) has been classified as hazard land (dark red); and
- parks (dark green) have been assigned to golf course (light gray) and new crop and pasture (turquoise).

In fact, a large portion of the confusion in the frequency-based classification results has a spatial pattern to it, and this is related to the spatial distributions of various land-use classes. As can be seen, most of the errors occur at boundaries between different land-use classes. This "boundary confusion" problem is caused by the pixel window effect (i.e., the effect of the pixel window-based frequency-extraction method). This is best illustrated by cleared land within an urban area. There should be only two resultant land-use classes: an urban land-use type neighboring a cleared-land type. However, as a pixel window is moved from the urban area across the boundary to the cleared land, the gray-level vector proportions contained in the histogram of this pixel window undergo a series of changes; from urban dominating, to similar proportions of urban surface and bare soil, to a high proportion of bare soil and a low proportion of paved surface, and finally to bare soil dominating the pixel window. The central two proportion configurations are transitional from urban area to cleared land, and may correspond to land under construction or new residential. Therefore, as the pixel window moves across the boundary between urban area and cleared land, more than two land-use classes will be obtained. This pattern can be found in Plate 2 at many locations. In most cases, these are errors.

Because the pixel window effect is a spatial process, it should be corrected spatially. It seems that two methods are promising. The simpler one is based on a morphological approach and will be called region-growing. The second method is a region-based contextual classification.

Region-growing is a process in which a region expands at its boundary until a specific target is met. In the region-growing method, one could set a distance threshold in the frequency-based classifier beyond which a pixel remains unclassified. In this manner, many transitional patterns will be thresholded and therefore remain unclassified. At the next step, a morphological erosion filter could be applied to the unclassified area until much or all of the entire unclassified area is eroded.

In the region-based contextual classification, one has to first correctly segment the image into a number of regions. Each of these regions can then be treated as an object. The frequency-based classifier can be applied to classify the image object-by-object. The difficult part in this method is the image segmentation. The regions have to contain different structure and cover components in order for the frequency-based classifier to be effectively applied.

Computational Efficiency of the Frequency-Based Classification Algorithm

The frequency-based classifier is coded so that, as the pixel window size increases, the required amount of CPU time remains the same. Therefore, when the image size is kept constant, the amount of CPU time required by the frequency-based classifier is determined only by the number of gray-level vectors in a gray-level vector-reduced image.

Table 7 lists the CPU times required by the frequency-based classifier on a VAX 11/785 computer when a pixel window size of 21 is used. For comparison purposes, the CPU time, 7 minutes and 47 seconds, used by the MLC method is also included.

The computational requirement of the frequency-based classifier is linear in relation to the number of gray-level vectors used. The classification time, plus the time for preparing a gray-level vector-reduced image, is comparable to that spent by the MLC method in this study. While the time required by the MLC is linearly related to the number of classes, it increases exponentially as the number of image bands is increased. As the number of image bands used in the MLC increases, the time required by the MLC is likely to be longer than the time needed by the frequency-based classifier.

SUMMARY AND CONCLUSIONS

A contextual method for land-use classification has been developed and evaluated using the SPOT HRV XS data obtained over the rural-urban fringe of northeastern Metropolitan Toronto. It involves two steps: gray-level vector reduction and frequency-based classification.

A new algorithm for gray-level vector reduction was developed and illustrated. The technical basis of the frequency-based contextual classifier and its advantages over commonly used contextual classification methods, such as those based on spatial feature extraction, have been introduced and discussed. A new criterion, average separability, for pixel window size selection was developed. The Kappa coefficient, the estimated variance of a Kappa value, and the conditional Kappa coefficient were used in the accuracy assessment of land-use classification results.

TABLE 7. CPU TIME REQUIRED ON A VAX 11/785 COMPUTER FOR LAND-USE CLASSIFICATION

Number of Elements	CPU Time	
	Minutes	Seconds
10	4	9
20	6	19
30	8	30
40	10	41
50	12	51
	147′ (required by the MLC)	47″

Some of the major findings from the experimental results are as follows:

- The frequency-based classification method can improve land-use classification accuracies obtained by the MLC. The overall classification accuracy measured by the Kappa coefficient was only 0.462 when the maximum likelihood classification (MLC) method was used in the classification of 14 land-use classes. This has been improved to 0.616 when a gray-level vector-reduced image with 40 gray-level vectors was classified using a frequency-based classifier with a pixel window size of 9 by 9 pixels.
- The significance of some accuracy improvements with the frequency-based method reached a confidence level of 0.99 for the land-use classification scheme and the SPOT HRV XS data used in this research.
- The contextual classification methods developed in this research are particularly effective in identifying spatially heterogeneous land-use classes, although in most cases these methods can also improve the classification accuracy of spatially homogeneous land-use classes.
- The gray-level vector-reduction algorithm is demonstrated to be very fast. The use of the frequency-based classification method requires similar or even less computation than the MLC method. To our knowledge, no other contextual classifiers are as computationally efficient as the method demonstrated in this study.
- Two parameters are important in the frequency-based classification. These are the pixel window size and the number of gray-level vectors to be used. However, as in other contextual classification algorithms, there is no effective indicator for the optimal pixel window size and the optimal number of gray-level vectors. The average separability measure proposed in this research provides an approximate measure to estimate the appropriate number of gray-level vectors and clusters.

In conclusion, the two general objectives of this research, (1) to develop computationally efficient contextual classification methods, and (2) to improve land-use classification accuracies of higher spatial resolution satellite data, have been achieved. Various aspects of these approaches have been evaluated. These methods are easy to implement and are computationally more efficient than commonly used contextual classification methods. They proved to be successful according to their capabilities for improving conventional computer-assisted land-use classification accuracies.

It is recommended that the algorithm developed in this research be used when higher spatial resolution remotely sensed data acquired over urban areas or rural-urban fringe areas are used for land-use classification. The potential of the methods proposed in this research need to be tested in other environments.

Further research is required to develop a procedure for optimal pixel window size prediction. It is recommended that, in addition to the training sample, the test sample has to be involved in such a procedure.

The pixel window effect is one of the major problems of the frequency-based classification method. It adds systematic spatial error patterns to the land-use classification results. Two methods (morphological region-growing and region-based contextual classification) have been proposed to reduce or overcome this effect. Tests of these methods or the development of new methods to overcome the pixel window effect will be important areas of research for further improving the performance of the frequency-based classification method. Research is also needed to experiment with frequency features other than the histogram features used in this study. For instance, co-occurrence frequencies may add more discriminating power to the frequency-based classification.

ACKNOWLEDGMENTS

We gratefully acknowledge the International Development Research Centre (IDRC), Ottawa, Canada, for the financial support to P. Gong while studying in Canada. The suggestions made by Dr. K. Zhang regarding the gray-level vector reduction in eigen space were very helpful. The assistance of SPOT Image Corporation of France and the Canada Centre for Remote Sensing is gratefully acknowledged in supplying the SPOT data used in this study as part of the Programme d'valuation Prliminair SPOT (PEPS), Project No. 229. This research was funded by a Centre of Excellence grant from the Province of Ontario to the Institute for Space and Terrestrial Science and NSERC Operating Grant A0766 awarded to P.J. Howarth. The reviewers' comments were most helpful.

REFERENCES

Bishop, Y. M. M., S. E. Feinberg, and P. W. Holland, 1975. *Discrete Multivariate Analysis - Theory and Practice*. The MIT Press, Cambridge, Mass.

Chittineni, C. B., 1981. Utilization of spectral-spatial information in the classification of imagery data. *Computer Graphics and Image Processing*, 16:305–340.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

Congalton, R. G., and R. A. Mead, 1983. A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering & Remote Sensing*, 49(1):69–74.

Driscoll, E. C., Jr., 1985. *Spatial Classification of Urban Landuses from Landcover Probabilities Derived through Remote Sensing*. Ph.D Dissertation, Stanford University, 158p.

Ducan, J. S., and W. Frei, 1990. Relaxation labeling using continuous label sets. *Pattern Recognition Letters*, 9(1):27–37.

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.

Fleiss, J. L., J. Cohen, and B. S. Everitt, 1969. Large sample standard errors of Kappa and weighted Kappa. *Psychological Bulletin*, 72(5):323–327.

Franklin, S. E., and D. R. Peddle, 1990. Classification of SPOT HRV imagery and texture features. *International Journal of Remote Sensing*, 11(3):551–556.

Fu, K. S., and T. S. Yu, 1980. *Spatial Pattern Classification Using Contextual Information*. Research Studies Press, Chichester, England.

Fung, T., and E. F. LeDrew, 1988. The determination of optimal threshold levels for change detection using various accuracy indices. *Photogrammetric Engineering & Remote Sensing*, 54(10):1449–1454.

Gong, P., and P. J. Howarth, 1989a. A modified probabilistic relaxation approach to land cover classification. *Proceedings of IGARSS'89/12th Canadian Symposium on Remote Sensing*, Vancouver, B. C., pp. 1621–1624.

———, 1989b. Performance analyses of probabilistic relaxation methods for land-cover classification. *Remote Sensing of Environment*, 30:33–42.

———, 1990a. Land cover to land use conversion: a knowledge-based approach. *Technical Papers, Annual Conference of American Society for Photogrammetry and Remote Sensing*, Denver, Colorado, 4:447–456.

———, 1990b. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogrammetric Engineering & Remote Sensing*, 56(1):67–73.

———, 1990c. An assessment of some factors influencing multispectral land-cover classification. *Photogrammetric Engineering & Remote Sensing*, 56(5):597–603.

———, 1990d. A graphical approach for evaluation of land-cover classification procedures. *International Journal of Remote Sensing*, 11(5):899–905.

Gonzalez, R. C., and P. Wintz, 1987. *Digital Image Processing*, Second Edition. Addison-Wesley Publishing Company, Reading, Mass.

Haralick, R. M., 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.

———, 1983. An interpretation for probabilistic relaxation. *Computer Vision, Graphics, and Image Processing*, 22:388–395.

Haralick, R. M., and H. Joo, 1986. A contextual classifier. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(6):997–1007.

Howarth, P. J., L.R.G. Martin, G. Holder, D. Johnson, and J. Wang, 1988. SPOT imagery for detecting residential expansion on the rural-urban fringe of Toronto, Canada. *SPOT-1 Image Utilization, Assessment, Results*, Cepadues-Editions, Toulouse, France, pp. 491–498.

Hsu, S., 1978. Texture-tone analysis for automated landuse mapping. *Photogrammetric Engineering & Remote Sensing*, 44(11):1393–1404.

Jensen, J. R., 1979. Spectral and textural features to classify elusive land cover at the urban fringe. *The Professional Geographer*, 31:400–409.

———, 1983. Urban/Suburban Land Use Analysis. *Manual of Remote Sensing*, Second Edition (R.N. Colwell, editor-in-chief), American Society of Photogrammetry, Falls Church, Virginia, pp. 1571–1666.

Johnson, D. D., and P. J. Howarth, 1987. The effects of spatial resolution on land cover/use theme extraction from airborne digital data. *Canadian Journal of Remote Sensing*, 13(2):68–74.

Kalayeh, H. M., and D. A. Landgrebe, 1984. Adaptive relaxation labeling. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-6(3):369–372.

Kauth, R. J., and G. S. Thomas, 1976. The Tasseled Cap, a graphic description of agricultural crops as seen by Landsat. *Symposium on Machine Processing of Remotely Sensed Data*, Purdue Univ., West Lafayette, Ind., pp.4B41–4B51.

Kittler, J., and E. R. Hancock, 1989. Combining evidence in probabilistic relaxation. *International Journal of Pattern Recognition and Artificial Intelligence*, 3(1):29–51.

Landgrebe, D. A., 1980. The development of a spectral-spatial classifier for earth observation data. *Pattern Recognition*, 12(3):165–175.

Marceau, D. J., P. J. Howarth, J-M. M. Dubois, and D. J. Gratton, 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.

Martin, L. R. G., 1975. *Land Use Dynamics on the Toronto Urban Fringe*. Lands Directorate Map Folio No. 3, Environment Canada, Ottawa.

———, 1986. Change detection in the urban fringe employing Landsat satellite imagery. *Plan Canada*, 26(7):182–190.

———, 1989. Accuracy assessment of Landsat-based visual change detection methods applied to the rural-urban fringe. *Photogrammetric Engineering & Remote Sensing*, 55(2):209–215.

Martin, L. R. G., P. J. Howarth, and G. Holder, 1988. Multispectral classification of land use at the rural-urban fringe using SPOT data. *Canadian Journal of Remote Sensing*, 14(2):72–79.

Mohn, E., N. L. Hjort, and G. D. Storvik, 1987. A simulation study of some contextual classification methods for remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(6):796–804.

Owen, A., 1984. A neighborhood-based classifier for Landsat data. *Canadian Journal of Statistics* 12(3):191–200.

Richards, J. A., 1986. *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag, Berlin.

Richards, J. A., D. A. Landgrebe, and P. H. Swain, 1982. A means for utilizing ancillary information in multispectral classification. *Remote Sensing of Environment* 12:463–477.

Rosenfeld, A., R. A. Hummel, and S. W. Zucker, 1976. Scene labeling by relaxation operations. *IEEE Transactions on System, Man and Cybernetics*, SMC-6(6):420–433.

Rosenfield, G. H., and K. Fitzpatrick-Lins, 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering & Remote Sensing*, 52(2):223–227.

Sezan, M. I., 1990. A peak detection algorithm and its application to histogram-based image data reduction. *Computer Vision, Graphics, and Image Processing*, 49:36–51.

Swain, P. H., S. B. Vardeman, and J. C. Tilton, 1981. Contextual classification of Multispectral image data. *Pattern Recognition*, 13(6):429–441.

Thomas, I. L., N. P. Ching, V.M. Benning, and J. A. D'Aguanno, 1987. A review of multichannel indices of class separability. *International Journal of Remote Sensing*, 8(2):331–350.

Tilton, J. C., S. B. Vardeman, and P. H. Swain, 1982. Estimation of contextual statistical classification of multispectral image data. *IEEE Transactions on Geoscience and Remote Sensing*, GE-20(4):445–452.

Welch, J. R., and K. G. Salter, 1971. A context algorithm for pattern recognition and image interpretation. *IEEE Transactions on System, Man and Cybernetics*, SMC-1(1):24–30.