

On the Classification of Classes with Nearly Equal Spectral Response in Remote Sensing Hyperspectral Image Data

Victor Haertel, *Member, IEEE* and David Landgrebe, *Life Fellow, IEEE*¹

© 1999 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. Reprinted from IEEE Transactions on Geoscience and Remote Sensing, Vol. 37, No. 5, Part 2, September, 1999.

Abstract

It is well known that high dimensional image data allows for the separation of classes that are spectrally very similar, i.e., possess nearly equal first order-statistics, provided that their second-order statistics differ significantly. The aim of this study is to contribute to a better understanding, from a more geometrically-oriented point of view, of the role played by the second-order statistics in remote sensing digital image classification of natural scenes when the classes of interest are spectrally very similar and high dimensional multispectral image data is available. A number of the investigations that have been developed in this area, deal with the fact that as the data dimensionality increases, so does the difficulty in obtaining a reasonably accurate estimate of the within-class covariance matrices from the usually limited number of available labeled samples. Several approaches have been proposed to deal with this problem. This study aims toward a complementary goal. Assuming that reasonably accurate estimates for the within-class covariance matrices have been obtained, we seek to better understand what kind of geometrically-oriented interpretation can be given to them as the data dimensionality increases and how this knowledge can help the design of a classifier. In order to achieve this goal, the covariance matrix is decomposed into a number of parameters that are then analyzed separately with respect to their ability to separate the classes. Methods for image classification based on these parameters are investigated. Results of tests using data provided by the sensor system AVIRIS are presented and discussed.

Index terms: AVIRIS sensor, digital image classification, high dimensional data, remote sensing, second-order statistics.

¹ Manuscript Received August 12, 1997; revised June 8, 1998. This work was supported in part by the Brazilian National Council for Scientific and Technological Development (Conselho Nacional para o Desenvolvimento Científico e Tecnológico) CNPq, Grant # 200066/96-0 (NV).

V. Haertel is with the Federal University at Rio Grande do Sul, C.P. 15044, Porto Alegre RS, 91501-970 Brazil (e-mail: haertel@if.ufrgs.br). D. Landgrebe is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285 USA (e-mail: landgreb@ecn.purdue.edu).

I. Introduction

The remote sensing analysis of natural scenes, has been relying mostly on data collected by sensors that provide a relatively small number of spectral bands. Sensor systems such as Landsat-TM and Spot have been widely used to gather the information required in areas such as forestry, agriculture, geology and many others. In almost all of the cases, the distribution of the individual spectral classes which are present in the image data, can be approximated by a multivariate normal distribution and the classification procedure can be performed by the well known quadratic classifier

$$G_i(\mathbf{X}) = -(\mathbf{X} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{X} - \boldsymbol{\mu}_i) - \ln|\boldsymbol{\Sigma}_i| + 2\ln P(\omega_i)$$

where ω_i represents a particular class, \mathbf{X} an unlabeled pixel, $\boldsymbol{\mu}_i$ the class mean vector, $\boldsymbol{\Sigma}_i$ the class covariance matrix, $P(\omega_i)$ the corresponding *a priori* probability for class ω_i and $G_i(\mathbf{X})$ the discriminant function associated with class ω_i .

In most of the cases this general approach has proved capable of performing image data classification in an acceptable way. There are some cases however in which some or all of the classes involved are very similar spectrally, i.e., their first-order statistics are nearly identical. In these cases, the more traditional sensor systems and image classification methods either yield a very low accuracy or fail completely. The answer to this problem is provided by a new generation of sensors which allow for a much larger number of spectral bands. The AVIRIS system, which possess the capability of sensing natural scenes in 220 spectral bands is one example of these new high dimensional sensor systems. As a consequence, the development of new methods to analyze this high dimensional image data have become necessary.

In this context, it is also well known that in a higher dimensional space, samples drawn from normally distributed data tend to fall towards the tails of the density function with virtually no samples falling in the central region (where the value of the density function is largest [1]). The result of this phenomenon is that in a high-dimensional space, different classes sharing the same expected values can become separable with very little error, provided that their covariance matrices are sufficiently distinct [1]. Thus, the second-order statistics can assume a preponderant role in remote sensing image data classification, possibly allowing for the separation of classes spectrally close to each other and therefore not separable by the current low-dimensionality sensor systems nor by analysis methods that do not take into account second-order statistics.

The classification of high dimensional image data, however, poses some new problems. One of the most difficult problems in dealing with high dimensional data resides in the estimation of the classes' covariance matrices. As the dimensionality increases, so do the number of samples required for a reliable estimation of the covariance matrix. Unfortunately, the number of samples in remote sensing applications is, in almost all practical cases, very limited. Methods to solve or to mitigate this problem have been receiving a considerable attention from the scientific community. Two main approaches have been pursued:

- 1) development of methods that allow the reduction of the data dimensionality with small loss of information;
- 2) development of new approaches for dealing with the covariance matrix when the sample size is small.

Several new approaches have been reported in the literature [2]-[7].

The aim of this study is, however, somewhat different. High dimensional spaces present a number of geometrical properties that contradict our usual

perception of space [4]. The aim of this study is to contribute to a better understanding (from a more geometrically-oriented point of view) of data distribution in high dimensional spaces. This in turn may lead to more adequate classification algorithms when the classes involved possess nearly equal first-order statistics and class separation is only possible via the second-order statistics. More specifically, we seek to take some initial steps that may lead to the development of a classification algorithm based solely on the within-class second order statistics, in high dimensional spaces. The first step required to implement such an algorithm consists in the segmentation of the image into homogeneous areas, each one supposedly belonging to one of the existing classes. The covariance matrix associated to each image segment is then decomposed into three parameters, which describe the size, shape and orientation of the image segment in the multispectral space. These parameters are then tested for their ability to separate the classes, and a classifier is then proposed based on these three parameters.

II. The Covariance Matrix

In multispectral image data, each class can be seen as a cluster of points in a multidimensional space. The position of the cluster can be measured by its central point, as defined by the class mean vector (first-order statistic). The cluster size, shape and orientation are measured by its covariance matrix (second-order statistics).

In order to better understand the meaning of the class covariance matrix in this context, we break it down into components that lend themselves to a more geometrical meaning. This can be accomplished by the spectral decomposition of the covariance matrix [8], [9].

$$\Sigma = \lambda \mathbf{D} \mathbf{A} \mathbf{D}^T \tag{1}$$

where λ is the largest eigenvalue of Σ , \mathbf{A} is a diagonal matrix whose elements are the

ratio of eigenvalues of Σ with respect to λ and \mathbf{D} is a matrix whose columns are the eigenvectors of Σ .

This particular form for the spectral decomposition is particularly useful for the purpose of this study, i.e., to understand the geometry of clusters representing different classes with nearly equal first-order statistics and further, to understand their separability in high dimensional spaces. A clear geometrical interpretation can be assigned to three components in (1). The largest eigenvalue λ represents the size of the cluster, the diagonal matrix \mathbf{A} defines the shape of the cluster and the matrix \mathbf{D} the orientation of the cluster in the multidimensional space.

Assuming that the original image data is multivariate normally distributed, then the same data measured along the rotated axis also presents a multivariate normal distribution [10]. The eigenvalues represent the variances along the rotated axis. In practice, we work with the sample estimate (\mathbf{S}) of the covariance matrix Σ . Thus all eigenvalues of \mathbf{S} (including the largest one λ) are random variables with Gamma distribution [11]. The Gamma density function parameters can be estimated from the available samples for later use in image data classification. A more difficult situation is presented by matrix \mathbf{A} . Since here we are dealing with a matrix, a statistical approach to its implementation in image data classification is more complex. In this study we propose to estimate the information carried by \mathbf{A} , by its trace. The cluster orientation can be estimated from the eigenvectors. In this study the eigenvector associated with the largest eigenvalue is used to characterize each cluster orientation. It is also assumed that the sample-estimated eigenvectors are random variables whose components are normally distributed.

III. Design of a Statistically Based Classifier

In order to understand better the importance of each component of the covariance matrix in multispectral image classification, the probability density function for each one of the three components must be derived.

A. Cluster size

Since it is basically a sample variance, the parameter size has a Gamma density

$$f_X(x) = \frac{1}{b^c \Gamma(c)} x^{c-1} \exp\left[-\frac{x}{b}\right] \quad (2)$$

Parameters b and c are associated, respectively, with the scale and shape of the density function (2) and can be estimated by the method of the moments. The first and second order-moments of the Gamma density are given by:

$$\begin{aligned} \mu_{(1)} &= E[X^1] = bc \\ \mu_{(2)} &= E[X^2] = b^2 c(c+1) \end{aligned}$$

B. Cluster Shape

The shape of the cluster can be defined by the relative length of its axes. The diagonal elements in A provide this data. One remaining question is how to make use of A in a classifier. One possible approach is by using the trace of A

$$T = \text{trace}(A) = \frac{\lambda_1}{\lambda_p} + \frac{\lambda_2}{\lambda_p} + \dots + \frac{\lambda_{p-1}}{\lambda_p} + 1$$

The eigenvalues of \mathbf{S} can be understood as the variances along the orthogonal directions defined by the eigenvectors, i.e., after rotating the original coordinate system in order to eliminate the correlation among them. Therefore, the sample eigenvalues are independent random variables. The elements in \mathbf{A} are the eigenvalues normalized with respect to the largest one, and they also form a set of independent

random variables. Therefore, the trace of \mathbf{A} lies within the interval $[1, p]$, p being the dimensionality of \mathbf{S} . A perfectly spherical cluster, i.e., the one with equal eigenvalues along all p axes will result in the trace of \mathbf{A} being equal to p . As the cluster becomes increasingly ellipsoidal, the trace of \mathbf{A} starts decreasing, becoming equal to the unit in the extreme case of only one eigenvalue being nonzero. The actual use of the trace of \mathbf{A} requires the knowledge of its density function.

Representing by λ_p the largest eigenvalue of Σ and by λ_i the remaining $(p-1)$ eigenvalues ($i = 1, 2, \dots, p-1$), it can be proved that the density function for the ratio $Z = \lambda_i / \lambda_p$ is given by

$$f_Z(z) = k \frac{z^{c_i-1}}{(z\mathbf{b}_p + \mathbf{b}_i)^{c_i+c_p}} (\mathbf{b}_i \mathbf{b}_p)^{c_i+c_p} \Gamma(c_i + c_p) \quad z > 0 \quad (3)$$

with

$$k = \frac{1}{\mathbf{b}_i^{c_i} \mathbf{b}_p^{c_p} \Gamma(c_i) \Gamma(c_p)}$$

The density function for the sum

$$R = \frac{\lambda_1}{\lambda_p} + \frac{\lambda_2}{\lambda_p} + \dots + \frac{\lambda_{p-1}}{\lambda_p}$$

is then given by the convolution of the $(p-1)$ densities [12]

$$f_R(r) = f_Z\left(\frac{\lambda_1}{\lambda_p}\right) * f_Z\left(\frac{\lambda_2}{\lambda_p}\right) * \dots * f_Z\left(\frac{\lambda_{p-1}}{\lambda_p}\right) \quad (4)$$

where each individual density function is given by (3). Finally, the density function $f_T(t)$ for the trace T of matrix \mathbf{A} can be obtained from (4) by a simple transformation of variables

$$f_T(t) = f_R(t-1) \quad (5)$$

The full derivation of the density function for the trace of matrix \mathbf{A} (5) is presented in the Appendix.

C. Cluster Orientation

The third parameter, orientation, can be characterized by the eigenvectors of Σ . Since the eigenvectors are orthogonal, one will suffice to describe the orientation of the cluster. In this study, the eigenvector associated with the largest eigenvalue is selected, and the normal distribution is assumed for the length of its components.

IV. Evaluating the Adequacy of the Parameters

As stated earlier, the aim of this study is to investigate the geometrical behavior of the second order statistics in remote sensing image data classification as a function of the data dimensionality. Primarily, we seek to investigate the importance of the three components (size, shape and orientation) for the classes that most often appear in satellite image data of natural scenes, particularly the ones that show equal or nearly equal first-order statistics and therefore are not separable by these alone.

Several approaches to measure the separability among classes are described in the literature. In this study, the Bhattacharyya distance is applied. The general expression for this distance is given by [13]

$$B = -\ln \left[\int_{-\infty}^{\infty} \sqrt{f_1(Z) f_2(Z)} dZ \right] \quad (6)$$

Bhattacharyya distance must be implemented to the three density functions and to the pairs of classes to be used in the test procedure to determine the efficiency of each of the three proposed random variables as the data dimensionality increases.

Size is a one-dimensional random variable (irrespective of the number of spectral bands in use) and has a Gamma density. In this case, the Bhattacharyya distance (6) between two classes, characterized by the Gamma parameters b_1, c_1 and b_2, c_2 , respectively, becomes

$$B = -\ln \int_{-\infty}^{\infty} \left[\frac{1}{b_1^{c_1} \Gamma(c_1)} x^{c_1-1} \exp\left[-\frac{x}{b_1}\right] \frac{1}{b_2^{c_2} \Gamma(c_2)} x^{c_2-1} \exp\left[-\frac{x}{b_2}\right] \right]^{\frac{1}{2}} dx$$

Performing the above integration, we obtain

$$B = -\ln \left[\frac{\Gamma\left(\frac{c_1+c_2}{2}\right) b_1^{c_2/2} b_2^{c_1/2} 2^{(c_1+c_2)/2}}{[\Gamma(c_1) \Gamma(c_2)]^{1/2} (b_1+b_2)^{(c_1+c_2)/2}} \right] \quad (7)$$

The estimation of Bhattacharyya distance for the random variable shape is a far more difficult problem. The general form for the density function [as in (4)], which involves the convolution among the component density functions, renders extremely difficult or even precludes a closed-form expression for this density function, even when one tries to perform this operation in the Fourier domain. In this study, we tried to estimate the Bhattacharyya distance by implementing (4) into (6) in a numerical fashion. This approach, however, presented the drawbacks that are inherent in many numerical procedures, namely, a large computational time and numerical instability at points where the density functions vary rapidly. Due to these problems, this approach proved to be unsatisfactory in the experiments performed in this study. A more convenient approach to this problem remains a topic for future research.

The Bhattacharyya distance for the cluster orientation parameter presents the well-known form for multivariate normal data

$$B = \frac{1}{8} (\mathbf{M}_1 - \mathbf{M}_2)^T \left(\frac{\mathbf{S}_1 + \mathbf{S}_2}{2} \right)^{-1} (\mathbf{M}_1 - \mathbf{M}_2) + \frac{1}{2} \ln \left[\frac{|(\mathbf{S}_1 + \mathbf{S}_2)/2|}{|\mathbf{S}_1|^{1/2} |\mathbf{S}_2|^{1/2}} \right] \quad (8)$$

where \mathbf{M} and \mathbf{S} represent respectively the mean vector and the covariance matrix for parameter orientation, for each pair of classes.

V. Testing the Parameters Size, Shape and Orientation

In order to investigate how the parameters size, shape and orientation perform in remote sensing image data classification, tests were carried out using multispectral image data obtained by the sensor AVIRIS. A 30 channel subset of an image obtained in June 1992 and covering an agricultural area in Indiana was used. The 30 channels were selected from the original 220 channels in a uniformly spaced fashion, in order to cover the entire range of wavelengths sensed by the AVIRIS system. The area covered by this image shows agricultural fields with corn and soybeans. Different agricultural techniques are present across these fields, allowing for the following user defined classes:

- 1) corn no till;
- 2) corn minimum-till (corn minimum);
- 3) corn clean (corn);
- 4) soybeans no till;
- 5) soybeans minimum-till (soybean minimum);
- 6) soybeans clean (soybean).

As the AVIRIS image was collected at the beginning of the growing season, only a small fraction ($\approx 5\%$) of the ground was actually covered by the crop. The largest contribution to the spectral response comes from the background, i.e., exposed soil plus debris. This situation results in first-order statistics that are nearly equal for all six classes. Under these circumstances, the traditional data classification algorithms based on low dimensionality data such as Landsat-TM, either yield very inaccurate results or fail completely. A successful procedure must be based on the class second-order statistics, employing higher dimensional data. Thus, the AVIRIS data provides ideal conditions to test image classification methods based on second-order statistics.

Also, it provides the right conditions to investigate the contribution of the three parameters originating from the spectral decomposition of the covariance matrix.

To test the ability of each of the three parameters in discriminating between the data classes as a function of the data dimensionality, the 30 channel AVIRIS image data were divided into six sub-sets of 5, 10, 15, 20, 25 and 30 channels, respectively. The Bhattacharyya distances for the 15 pairs of classes were then computed. Figs. 1 and 2 illustrate typical behavior of Bhattacharyya distance for parameters size (7) and orientation (8), respectively, as a function of the data dimensionality. As for the parameter shape, only a few values for Bhattacharyya distance were actually computed. Due to the reasons reported in Section IV, we did not succeed in obtaining a closed-form for Bhattacharyya distance for this parameter, and the numerical approach tested proved to be unsatisfactory. Therefore, Bhattacharyya distance for the parameter shape was not fully estimated in the experiments involving AVIRIS data. However, the histograms for this parameter, involving the six classes and the six sub-sets of channels, tend to suggest a comparatively low capability of discriminating the image classes under consideration (Figs. 3 and 4).

The analysis of the data depicted in Figs. 1-4 allows one to draw the following conclusions.

- 1) The parameter orientation presents a very consistent pattern of increasing Bhattacharyya distance as the dimensionality of the data increases.
- 2) The parameter size provides smaller values for Bhattacharyya distance.
- 3) The histograms for parameter shape tend to suggest a lower discriminant power in separating the image classes involved.

The analysis of these experiments suggests that the parameter orientation is the most promising one to successfully classify the image data being tested. The Bhattacharyya distance among the pairs of classes increases consistently as the number of channels

increases. The other two parameters show a less promising behavior.

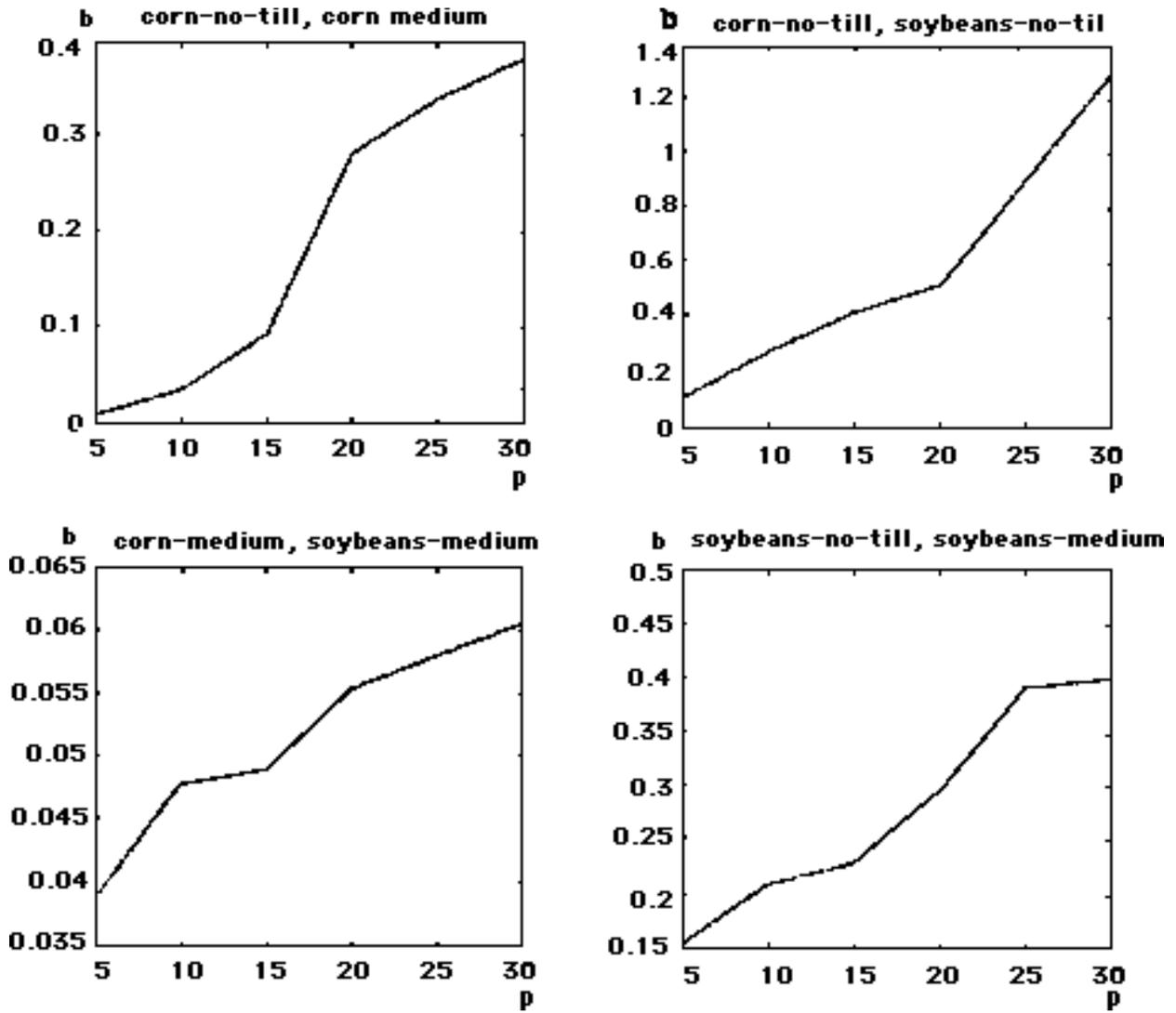


Fig. 1.

Bhattacharyya Distance (b) for parameter size as a function of the number of spectral bands (p), for four different pairs of classes

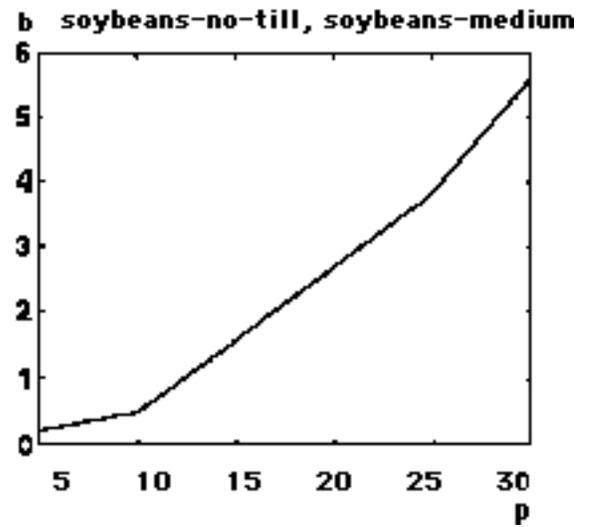
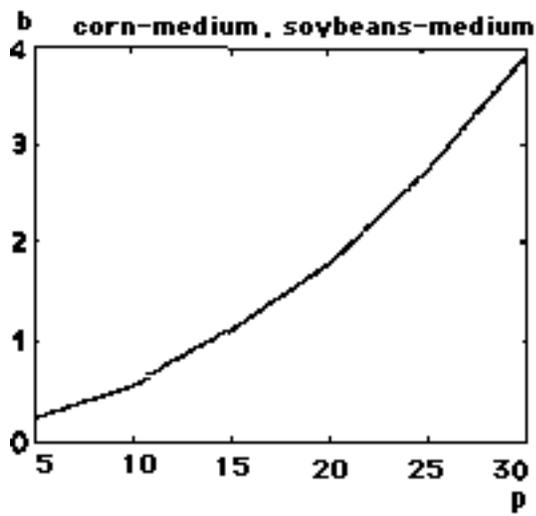
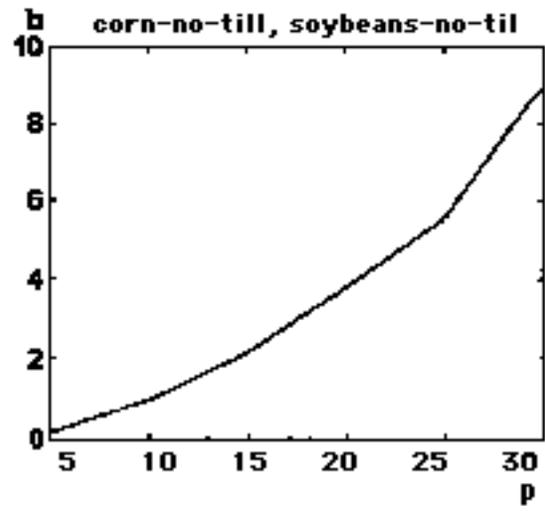
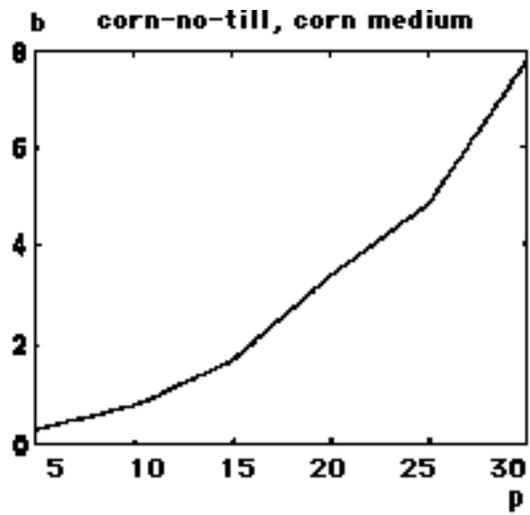


Fig. 2.

Bhattacharyya Distance (**b**) for parameter orientation as a function of the number of spectral bands (**p**), for four different pairs of classes

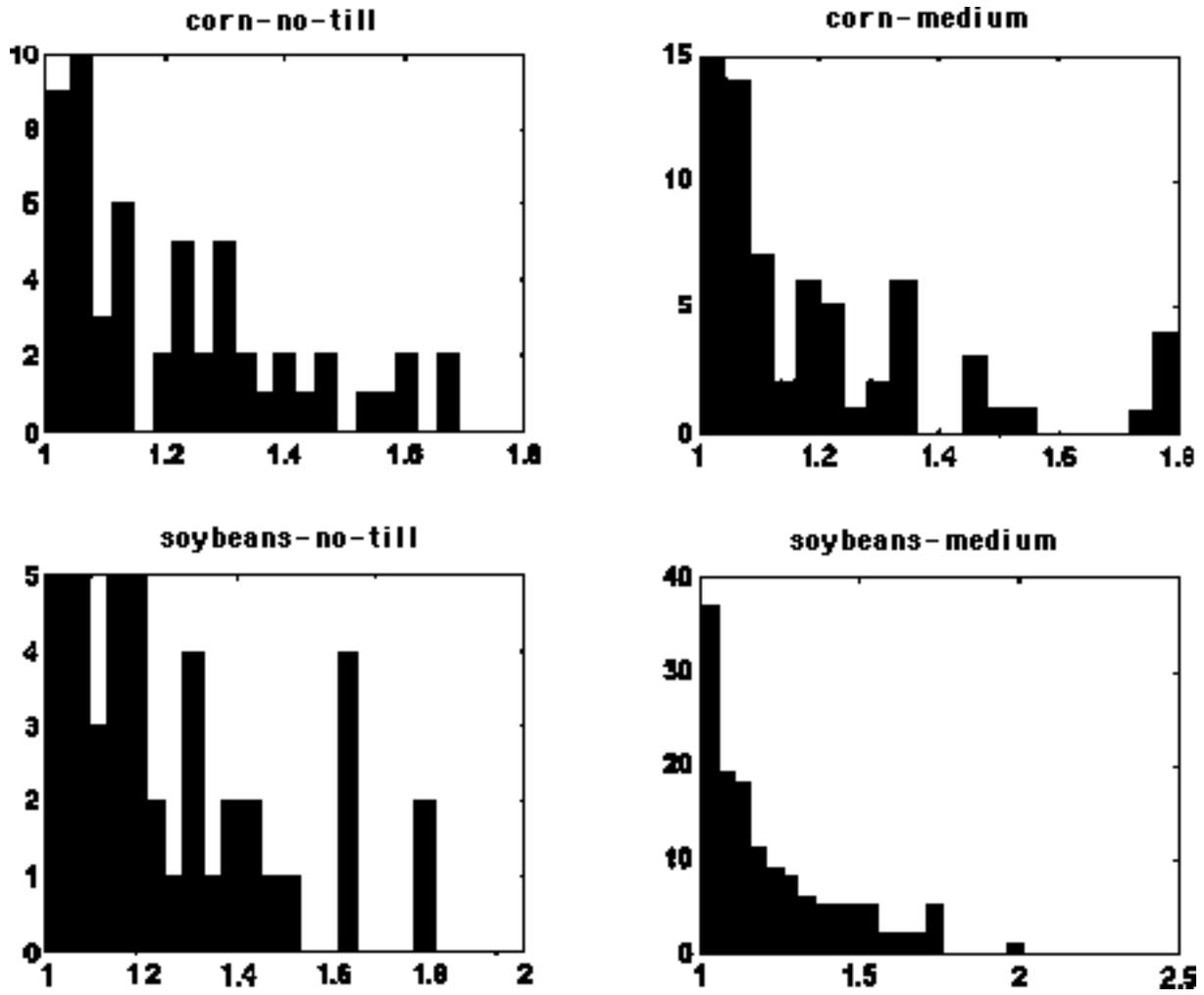


Fig. 3. Histogram for parameter shape, for four different classes. Experiment using 5 spectral bands. Parameter shape is represented along the horizontal axis and the corresponding number of occurrences along the vertical axis.

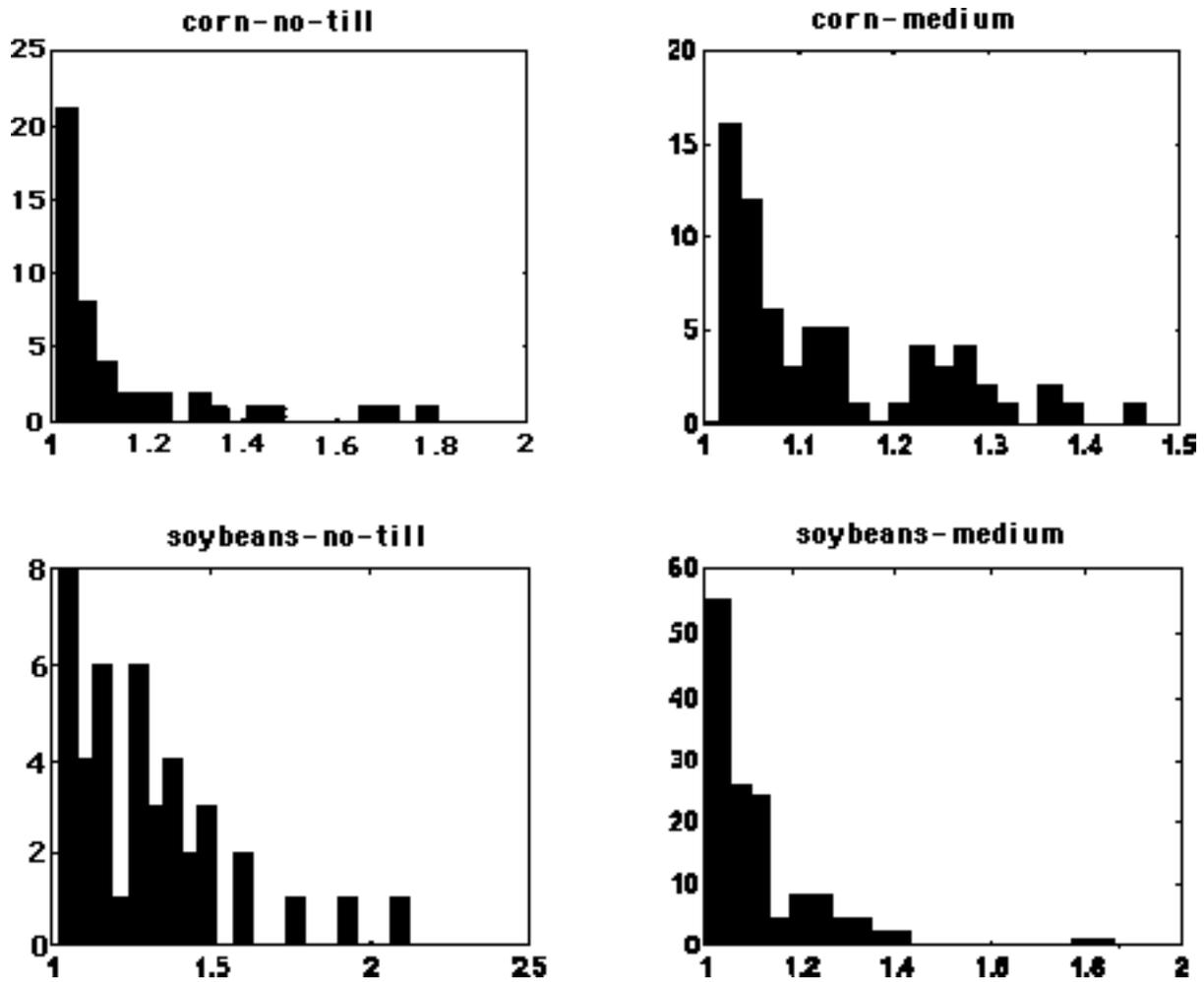


Fig. 4.
Histogram for parameter shape, for four different classes. Experiment using 30 spectral bands. Parameter shape is represented along the horizontal axis and the corresponding number of occurrences along the vertical axis.

VI. Use of the Parameters Size, Shape and Orientation as a basis for an image classification algorithm

In this section we investigate the development of an algorithm based on the parameters size, shape, and orientation for image classification purposes. In this case, the classification procedure would be based on the second-order statistics only. Thus, the unit elements to be classified are unlabeled clusters of pixels, not individual pixels.

At this point, we have to deal with three problems:

- 1) how to segment a given multispectral image into homogeneous areas (each segmented area would then be treated as an individual cluster and would be assumed to belong to one of the existing classes);
- 2) which parameter(s) is (are) more efficient for classification purposes;
- 3) which method is the most adequate for this particular approach to image classification.

This study deals basically with problems 2 and 3. As for problem 1, some initial tests were performed making use of the first part of the ECHO algorithm as proposed by Kettig and Landgrebe [14], [15]. Encouraging results were obtained, but additional work is needed on this topic when high dimensional data is involved. In the experiments performed, the number and particularly the size of the resulting fields showed a tendency to decrease as the data dimensionality increased, leaving many blanks across the image. This item is therefore left as a topic for future investigation.

In order to make possible the investigation into problems 2 and 3, clusters are needed for both training and testing purposes for every class. In order to obtain these clusters, the entire data set available for each class was uniformly divided into a training set and a test set. A number of clusters were then drawn randomly from both

the training and the test sets. Therefore, a number of labeled clusters were obtained, allowing for the estimation of parameters and for the testing of the methods for classification. The results obtained in section V suggest the idea of carrying out image data classification based on parameter orientation. In this case, each homogeneous image region or cluster can be represented by a point defined by the eigenvector associated with its largest eigenvalue, in the p -dimensional multispectral space. The space defined in this way, is herewith designated by "eigenvector-orientation" space.

As for the classification method in the eigenvector-orientation space, a conventional minimum distance approach was initially tested. The mean value for each class in the eigenvector-orientation space was then estimated from the corresponding training set. Next, the eigenvectors were calculated for all individual test clusters, and in the eigenvector-orientation space each test cluster was assigned to the class presenting the closest mean. Different sizes for the test clusters were used: 1.2, 1.3, 2.0, 4.0 and 6.0 times the number of spectral bands. This approach however, did not provide accurate results. The reason can be better understood by examining the two terms of Bhattacharyya distance between the classes in the eigenvector-orientation space as shown in Table I. As is well known, the first term of Bhattacharyya distance estimates the separability caused by the first-order statistics whereas the second term estimates the contribution due to the second-order statistics. At a low dimensionality, both terms are small. As the dimensionality increases, the first term in Bhattacharyya distance shows a small increase, whereas the second one shows a much larger increase (Table I). Thus, the separability among classes in the eigenvector-orientation space is caused primarily by the second-order variation on the orientation of the eigenvectors, not by the first-order variation. This fact explains the failure of a minimum distance algorithm approach in yielding accurate results in image classification.

To overcome this problem, two different approaches were pursued and later combined into a single one. First, additional information was added to each cluster, i.e., the number of parameters as defined in the aforementioned problem 2 was increased from one (orientation) to all the three parameters. The rationale for this approach is provided by the experiments reported in section V. The parameter orientation proved to be the most promising one for classification purposes (larger Bhattacharyya distance between classes). However, the contribution of the remaining two parameters (size and shape of the cluster), although smaller, is not entirely negligible. Thus, adding parameter size and shape might improve the accuracy. One way of implementing this idea is by taking into consideration all p eigenvectors associated with each cluster, instead of the single one associated with the largest cluster's eigenvalue. One possible way of achieving this intent is by weighting the p eigenvectors by their associated eigenvalues and calculating their sum. The resulting vector, herewith called "weighted-eigenvector", defines a point that is then supposed to represent a cluster in a p -dimensional space called the "weighted-eigenvector space".

A minimum distance classifier, similar to the one previously applied to the orientation-eigenvector space, was tested on the weighted-eigenvector space. Tables II-VII depict the accuracies achieved for the six classes involved. The lines on these tables show the resulting accuracies as a function of the number of spectral bands used (5, 10, 15, 20, 25 and 30), and the columns display the accuracies as a function of the size of the test clusters, here expressed as a multiple of the data dimensionality (1.2, 1.5, 2.0, 4.0 and 6.0 times the number of spectral bands used).

The results conform with the theoretical predictions. The accuracy increases both with the data dimensionality and with the size of the cluster. Higher dimensionality enhances the separability based on the second-order statistics, whereas

a larger sample size allows for a more reliable estimation of the covariance matrix. It also shows that a reasonably high accuracy can be achieved even for classes that possess mean values that are nearly equal. Some variations on the accuracy can be seen across these tables and in the case of smaller clusters, are most likely caused by inaccuracies in the estimation of covariance matrix in the multispectral space.

However promising, the minimum distance classifier as applied to the weighted-eigenvector space presents some drawbacks. The results shown in Tables II-VII were obtained for clusters formed by drawing pixels randomly from the entire test set. These clusters are therefore formed by pixels spread across the entire test set. Thus, they are likely to yield more accurate estimates for the second-order statistics, unaffected by local variations in the image data. Subsequent tests were performed on clusters formed by pixels drawn sequentially from the test set. Thus, these clusters correspond approximately to segmented image regions. In this case, the minimum distance classifier on the weighted-eigenvector space yielded a much lower accuracy. These poorer results can be explained by the local variations across the image, which in turn cause larger variations on the second-order statistics in clusters belonging to the same class.

This set of experiments demonstrate the necessity of introducing the contribution of the second-order statistics in the weighted-eigenvector space, as made clear by the experiments summarized in Table I. One simple way of introducing this contribution, is by replacing, in the weighted-eigenvectorspace, the conventional minimum distance classifier which makes use of the Euclidean distance, by a classifier based on the Mahalanobis distance

$$G_i(\mathbf{X}) = -(\mathbf{X} - \mu_i)^T \Sigma_i^{-1} (\mathbf{X} - \mu_i) \quad (9)$$

where \mathbf{X} represents the cluster-weighted eigenvector, μ_i represents the mean-weighted eigenvector for class ω_i and Σ_i represents the weighted-eigenvector

covariance matrix for class ω_i . Estimates for both μ_i and Σ_i were obtained from each class training-clusters. It must be emphasized that the classification approach implementing the Mahalanobis distance implies two distinct natures in the estimation of the covariance matrices: 1) the covariance matrix for every segmented image area or cluster, in the multispectral space, to estimate the corresponding weighted-eigenvector; 2) the covariance matrix associated with each class in the weighted-eigenvector space, as required by the Mahalanobis distance classifier (9). The reliability of the former basically depends upon the size of the cluster (the number of pixels) whereas the latter depends upon the number of training clusters (each cluster provides one sample). Experiments were then carried out by creating clusters from contiguous pixels in the test set rather than from pixels drawn on a random fashion as done in the previous experiment. These clusters are then likely to be formed by pixels that are spatially close to each other in the image. Two experiments were carried out, using 30 spectral bands. In the first case, the size of each test cluster was taken equal to 180 pixels, and in the second case it was taken as equal to 100 pixels. In both cases, a sufficiently large number of training clusters were provided in order to allow a reasonably accurate estimation of the class covariance matrices in the weighted-eigenvector space. Four classes were tested and the results are shown in Table VIII. As one might have expected, the experiment involving a larger number of pixels per cluster allowed a more accurate estimation of the initial covariance matrices, resulting in a more accurate estimation of the parameters orientation, size, and shape, and therefore, in more accurate final classification results. However, in digital image classification, smaller clusters are quite often present and a method must be found to deal with them accurately. Here, we come to the well-known and extensively-discussed problem of parameter estimation from a sample of limited size. This problem becomes of crucial importance in the present

case where the following applies.

- 1) The algorithm is based solely on the second order statistics;
- 2) We are dealing with high dimensional data;
- 3) As in most remote sensing applications, the number of available labeled samples is limited.

Methods for the estimation of the covariance matrix when the number of available samples is small, have been extensively studied, and many approaches to mitigate this problem have been reported in the literature. In his paper, Friedman [5] discusses methods of regularization to improve the estimates of the covariance matrices. Hoffback and Landgrebe [7], among several others, also proposed methods for the estimation of the covariance matrix when the sample size is small.

Therefore, in order to make the weighted-eigenvector space concept fully operational, further investigation is still required into two main topics:

- 1) Incorporation into the algorithm of more adequate methods to estimate the class covariance matrices when small clusters are involved;
- 2) The development of methods for image segmentation in high dimensional spaces, i.e., methods to form the initial clusters.

Pair of Classes	Bhattacharyya distance terms 5 spectral bands	Bhattacharyya distance terms 30 spectral bands
corn_notill and corn_minimum	0.1079 0.3562	2.2183 69.0899
corn_notill and corn	0.1251 0.4727	1.8476 65.5503
corn_notill and soybean_notill	0.0564 0.6545	7.0056 137.8793
corn_notill and soybean_minimum	0.0199 0.3124	0.7894 65.4253
corn_notill and soy_clean	0.0298 0.3665	1.9428 65.1725
corn_minimum and corn	0.0483 0.1774	0.8256 9.8284
corn_minimum and soybean_notill	0.0325 0.3803	1.2763 83.3849
corn_minimum and soybean_minimum	0.1067 0.1257	0.8116 8.3095
corn_minimum and soybean_clean	0.0909 0.2200	0.8941 9.2650
corn and soybean_notill	0.0565 0.5304	2.9888 79.2144
corn and soybean_minimum	0.0745 0.3090	0.7033 6.7028
corn and soybean_clean	0.0832 0.2763	0.6025 6.5457
soybean_notill and soybean_minimum	0.0461 0.1752	1.5873 79.5270
soybean_notill and soybean_clean	0.0519 0.4746	2.3716 79.7257
soybean_minimum and soybean_clean	0.0357 0.2938	0.4289 6.5080

Table I

First and second term of Bhattacharyya distance for parameter Orientation, for different pairs of classes and for two sets of spectral bands. The first term (upper entry on every box) measures the contribution due to the classes mean vectors whereas the second term (lower entry on every box) measures the contribution due to the classes covariance matrices.

Cluster Size Dimension	1.2	1.5	2.0	4.0	6.0
5	21%	20%	39%	36%	46%
10	32%	41%	58%	73%	82%
15	43%	52%	61%	85%	87%
20	51%	63%	64%	86%	94%
25	70%	67%	78%	91%	95%
30	68%	67%	88%	94%	97%

Table II
Class Corn No Till

Classification accuracy, yielded by the Minimum Distance Classifier based on parameter "orientation"
The results are shown for different numbers of spectral bands (5,10,15,20,25 and 30) and for different cluster sizes, i.e., number of pixels per cluster (1.2, 1.5, 2.0, 4.0 and 6.0 times the number of spectral bands).

Cluster Size Dimension	1.2	1.5	2.0	4.0	6.0
5	26%	24%	27%	40%	48%
10	38%	44%	41%	63%	70%
15	50%	50%	54%	73%	82%
20	54%	61%	66%	82%	92%
25	60%	63%	68%	90%	91%
30	59%	64%	74%	93%	96%

Table III
Class Corn Minimum

Classification accuracy, yielded by the Minimum Distance Classifier based on parameter "orientation"
The results are shown for different numbers of spectral bands (5,10,15,20,25 and 30) and for different cluster sizes, i.e., number of pixels per cluster (1.2, 1.5, 2.0, 4.0 and 6.0 times the number of spectral bands).

Cluster Size Dimension	1.2	1.5	2.0	4.0	6.0
5	53%	63%	52%	60%	56%
10	58%	65%	62%	72%	75%
15	71%	62%	60%	72%	75%
20	64%	64%	63%	72%	80%
25	66%	70%	70%	71%	82%
30	70%	73%	73%	79%	86%

Table IV
Class Corn

Classification accuracy, yielded by the Minimum Distance Classifier based on parameter "orientation"
The results are shown for different numbers of spectral bands (5,10,15,20,25 and 30) and for different cluster sizes, i.e., number of pixels per cluster (1.2, 1.5, 2.0, 4.0 and 6.0 times the number of spectral bands).

Cluster Size Dimension	1.2	1.5	2.0	4.0	6.0
5	80%	86%	87%	97%	99%
10	87%	91%	96%	99%	100%
15	92%	95%	98%	99%	100%
20	93%	96%	99%	100%	100%
25	98%	97%	99%	100%	100%
30	97%	100%	100%	100%	100%

Table V
Class Soybean No Till

Classification accuracy, yielded by the Minimum Distance Classifier based on parameter "orientation"
The results are shown for different numbers of spectral bands (5,10,15,20,25 and 30) and for different cluster sizes, i.e., number of pixels per cluster (1.2, 1.5, 2.0, 4.0 and 6.0 times the number of spectral bands).

Cluster Size Dimension	1.2	1.5	2.0	4.0	6.0
5	22%	37%	38%	54%	68%
10	48%	63%	57%	73%	83%
15	55%	57%	70%	83%	93%
20	73%	77%	79%	94%	96%
25	69%	75%	82%	95%	99%
30	84%	83%	83%	98%	99%

Table VI
Class Soybean Minimum

Classification accuracy, yielded by the Minimum Distance Classifier based on parameter "orientation"
The results are shown for different numbers of spectral bands (5,10,15,20,25 and 30) and for different cluster sizes, i.e., number of pixels per cluster (1.2, 1.5, 2.0, 4.0 and 6.0 times the number of spectral bands).

Cluster Size Dimension	1.2	1.4	2.0	4.0	6.0
5	35%	30%	29%	42%	59%
10	44%	46%	48%	68%	70%
15	49%	52%	55%	68%	84%
20	54%	59%	71%	81%	89%
25	56%	63%	76%	89%	90%
30	73%	65%	76%	90%	95%

Table VII
Class Soybean Clean

Classification accuracy, yielded by the Minimum Distance Classifier based on parameter "orientation"
The results are shown for different numbers of spectral bands (5,10,15,20,25 and 30) and for different cluster sizes, i.e., number of pixels per cluster (1.2, 1.5, 2.0, 4.0 and 6.0 times the number of spectral bands).

Cluster Size Class	180 pixels	100 pixels
corn_notill	100%	93.4%
corn_minimum	100%	100%
soy_notill	100%	100%
soy_minimum	93%	72%

Table VIII
Classification Accuracy, Obtained by Applying the Mahalanobis Distance on the Weighted-eigenvector Space
Results are Shown for 30 Spectral Bands and Two Cluster Sizes (180 and 100 pixels)

VII. Conclusions

This study deals with some aspects of the problem of classifying remote sensing image data in which classes with equal or nearly equal first-order statistics are present. In this case, the conventionally-used multispectral image data with a small number of spectral bands, as provided by systems such as Landsat-TM and Spot, either yield low accuracy results or fail completely. The way to deal with this problem involves the use of classification procedures relying solely on the second-order statistics estimated from high-dimensional image data as provided by hyperspectral sensors.

In order to better understand the role played by the second-order statistics in a higher dimensional space, the technique known as the spectral decomposition of the covariance matrix was used. This decomposition allows for an easier, more geometrically-oriented interpretation of the covariance matrix components and thus for a better insight into the way it separates classes in a higher dimensional space. These components represent the size, shape and orientation of clusters of pixels in the multispectral space. As these three component parameters are estimated from the sample covariance matrix, they are random variables themselves. Except for the

parameter shape, which presents a rather complex form for its density function, the ability of each parameter to separate the image classes under consideration was estimated by the Bhattacharyya distance. For the parameter shape, only a few values for the Bhattacharyya distance were actually computed. Histograms were then used to provide some insight into its behavior.

Based on this idea, classification methods were proposed and tested. Tests were carried out on a 30 channel sub-set of an AVIRIS image, displaying a number of classes having first-order statistics that were nearly equal. The results suggested the orientation parameter as the most promising one to separate the classes involved, and to a lesser degree, they suggest the other two parameters.

Based on these initial findings, an attempt was made to perform the classification based only on the parameter orientation. Following this approach, each image segment or cluster is represented by a single point, defined by the eigenvector associated with its largest eigenvalue. The mean value for every class is represented exactly in the same way, the eigenvectors being estimated from the corresponding training sets. These points lie on the surface of an hypersphere of unit radius. The minimum Euclidean distance classifier was then applied. The resulting classification accuracy was low, however. Experiments have shown that this low performance is caused by the fact that it is the second-order variation of the parameter orientation, not the first-order variation that carries the discriminant power among the classes. When the euclidean distance was replaced by the Mahalanobis distance, and the contribution of the other two parameters (size and shape) were introduced, a very significant improvement in the accuracy was obtained.

The experiments performed in this study proved that the weighted-eigenvector space based on the parameters size, shape and orientation provide a promising approach for image data classification, based solely on the second order statistics.

This approach also provides a more geometrically-oriented insight into the separability of classes possessing equal or nearly equal first-order statistics.

Crucial to this image data classification approach is the estimation of the class covariance matrix in high dimensional spaces from a limited number of training samples. Some approaches proposed in the literature must be introduced here.

Also, the problem related to the initial image segmentation still remains to be solved. As this proposed image data classification method is solely based on the second-order statistics, the basic unit to be classified has to be a cluster of pixels, not individual pixels as in more traditional data classification methods. Thus, the first step needs to be an image segmentation procedure to segment the image into homogeneous regions, each of which is assumed to belong to a given class. In other words, we need to start by forming clusters. The proposed classification method can then be applied, taking these clusters as basic units. Some tests were performed, implementing the first part of the ECHO algorithm as proposed by Kettig and Landgrebe [14]. Additional work on this topic, however, is still needed when dealing with high dimensional image data.

In this study, we have investigated a more geometrically-oriented view of the behavior of the second-order statistics for the case of image data in high dimensional spaces, and we have seen how this approach can be used for classification purposes when the classes involved possess an equal or nearly equal first order-statistics. The concept of weighted-eigenvector space was proposed, and tests were performed. This approach can be used to develop a method for image data classification when the classes involved are spectrally very similar and therefore not separable by the more traditional methods based on low dimensional data. In order to fully attain this objective, two additional topics need to be further investigated and added to the procedure:

- 1) methods to deal with the estimation of the covariance matrix when the number of training samples is small;
- 2) methods for image segmentation in high dimensional multispectral spaces, to form the clusters required by the approach proposed in this study.

It is also possible that the weighted-eigenvector space may prove to be adequate to implement clustering methods based solely on the second-order statistics. These items, however, remain as topics for future research.

Appendix

In this appendix we provide the full derivation of the density function for the trace of matrix \mathbf{A} .

Since the elements in \mathbf{A} are given by ratios (Z) among two independent random variables (X, Y), i.e., $Z=X/Y$, and also recalling that $Z \geq 0$, we have

$$P(Z < z) = \int_0^{\infty} dy \int_0^{yz} f_{XY}(x,y) dx \quad x,y \geq 0$$

and

$$P(Z < z) = \int_{-\infty}^0 dy \int_0^{yz} f_{XY}(x,y) dx \quad x,y \leq 0$$

The distribution function can thus be calculated by

$$F_Z(z) = P(Z < z) = \int_0^{\infty} dy \int_0^{yz} f_{XY}(x,y) dx - \int_{-\infty}^0 dy \int_0^{yz} f_{XY}(x,y) dx \quad z \geq 0$$

The density function can be obtained by

$$f_Z(z) = \frac{\partial}{\partial z} (F_Z(z))$$

Recalling that in a general case we have

$$\frac{\partial}{\partial q} \int_p^q f(x) dx = f(q) \quad (p = \text{constant})$$

and

$$\frac{\partial f_{XY}(x,y)}{\partial z} = \frac{\partial f_{XY}(x,y)}{\partial(yz)} \frac{\partial(yz)}{\partial z} = y \frac{\partial f_{XY}(x,y)}{\partial(yz)}$$

we can obtain the general expression for the density function of the ratio of two independent random variables, $Z = X/Y$

$$f_Z(z) = \int_0^{\infty} y f_{XY}(yz,y) dy - \int_{-\infty}^0 y f_{XY}(yz,y) dy \quad (\text{A1})$$

For the specific case under consideration the densities are given by the Gamma density function

$$X \sim \gamma(\mathbf{b}_1, c_1) = \frac{1}{\mathbf{b}_1^{c_1} \Gamma(c_1)} x^{c_1-1} \exp\left[-\frac{x}{\mathbf{b}_1}\right]$$

$$Y \sim \gamma(\mathbf{b}_2, c_2) = \frac{1}{\mathbf{b}_2^{c_2} \Gamma(c_2)} y^{c_2-1} \exp\left[-\frac{y}{\mathbf{b}_2}\right]$$

Since X and Y are independent

$$f_{XY}(x,y) = f_X(x)f_Y(y) = kx^{c_1-1}y^{c_2-1} \exp\left[-\frac{x}{\mathbf{b}_1}\right] \exp\left[-\frac{y}{\mathbf{b}_2}\right] \quad (\text{A2})$$

with

$$k = \frac{1}{\mathbf{b}_1^{c_1} \mathbf{b}_2^{c_2} \Gamma(c_1) \Gamma(c_2)}$$

Making use of this result in (A1) and recalling that Y being an eigenvalue of S cannot take negative values

$$f_Z(z) = k \int_0^{\infty} y (zy)^{c_1-1} y^{c_2-1} \exp\left[-\frac{zy}{b_1}\right] \exp\left[-\frac{y}{b_2}\right] dy$$

Performing the above integration, we obtain

$$f_Z(z) = k \frac{z^{c_1-1}}{(z b_2 + b_1)^{c_1+c_2}} (b_1 b_2)^{c_1+c_2} \Gamma(c_1+c_2) \quad z > 0 \quad (\text{A3})$$

Equation (A3) represents the density function for the ratio of two eigenvalues of \mathbf{S} . To test if (A3) fulfills the condition of normality as any density function, standard tables of integrals can be used to show that

$$\int_0^{\infty} f_Z(z) dz = 1$$

Once the density function for the ratio of eigenvalues has been defined, the next step consists in defining the density function for the summation of the ratios of eigenvalues that compose the diagonal matrix \mathbf{A} , i.e., the shape of the cluster

$$R = \frac{\lambda_1}{\lambda_p} + \frac{\lambda_2}{\lambda_p} + \dots + \frac{\lambda_{p-1}}{\lambda_p} \quad (\text{A4})$$

The derivation of the density function for a sum of random variables is a somewhat more complex problem. It can be proved [7] that it is given by the convolution of the densities (A3) of the elements in (A4)

$$f_{R(r)} = f_Z\left(\frac{\lambda_1}{\lambda_p}\right) * f_Z\left(\frac{\lambda_2}{\lambda_p}\right) * \dots * f_Z\left(\frac{\lambda_{p-1}}{\lambda_p}\right) \quad (\text{A5})$$

where the eigenvalues of \mathbf{S} are sorted in an ascending order. Finally, to obtain the density function of the trace of \mathbf{A} , i.e., of the cluster shape, the constant term 1 must be added to the sum in (A4)

$$T = \text{trace}(\mathbf{A}) = \frac{\lambda_1}{\lambda_p} + \frac{\lambda_2}{\lambda_p} + \dots + \frac{\lambda_{p-1}}{\lambda_p} + 1$$

A theorem in mathematical statistics proves that given a relation involving two random variables X and Y, such that

$$Y = a X + b$$

then

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{(y - b)}{a}\right)$$

Here, $a, b=1$. Thus, the density function for the shape of the cluster can be obtained from (A5) by a simple transformation of variables.

Acknowledgement

Dr. Haertel would like to express his appreciation to Dr. Landgrebe who made available the facilities at the Multispectral Image Processing Laboratory, School of Electrical and Computer Engineering, Purdue University, where this work was done.

List of References

- [1] K. Fukunaga, "Introduction to Statistical Pattern Recognition", second edition, Academic Press, 1990.
- [2] C. Lee and D.A. Landgrebe, "Analyzing High Dimensional Multispectral Data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 31, No. 4, pp 792-800, July 1993. See also "Feature Extraction and Classification Algorithms for High Dimensional Data", Purdue University, West Lafayette, IN, Technical Report TR-EE 93-1, January 1993.
- [3] B. M. Shahshahani and D. A. Landgrebe, " The Effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, No. 5, pp. 1087-1095, September 1994. See also B. Shashahani and D. Landgrebe, "Classification of Multi-Spectral Data by Joint Supervised-Unsupervised Learning", Purdue University, West Lafayette, IN, Technical Report TR-EE 94-1, January 1994.
- [4] L. Jimenez and D. A. Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical and Asymptotical Properties of Multivariate Data", *IEEE Transactions on System, Man and Cybernetics*, Vol. 28, No. 1, pp. 39-54, February 1998. See also L. O. Jimenez and D. Landgrebe, "High Dimensional Feature Reduction Via Projection Pursuit", Purdue University, West Lafayette, IN, Technical Report TR-ECE 96-5, April 1996.
- [5] J. H. Friedman, "Regularized Discriminant Analysis", *Journal of the American Statistical Association*, Vol. 84, No. 405, pp. 165-175, March 1989.
- [6] T. Hastie, A. Buja and R. Tibshirani, "Penalized Discriminant Analysis", *The Annals of Statistics*, Vol. 23, No. 1, pp. 73-102, 1995.

- [7] J. F. Hoffbeck and D. A. Landgrebe, "Covariance Matrix Estimation with Limited Training Data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 7, pp. 763-767, July 1996.
- [8] A. C. Rencher, "Methods of Multivariate Analysis", John Wiley & Sons, 1995.
- [9] J. D. Banfield and A.E. Raftery, "Model-Based Gaussian and Non-Gaussian Clustering", *Biometrics*, vol. 49, pp. 803-821, September 1993.
- [10] D. Morrison, "Multivariate Statistical Methods", Mc Graw-Hill, second edition, 1976.
- [11] A. M. Mood, F. A. Graybill and D. Boes, "Introduction to the Theory of Statistics", McGraw-Hill, third edition, 1974.
- [12] A. Papoulis, "Probability, Random Variables and Stochastic Processes", third edition, McGraw-Hill, 1991.
- [13] C. W. Therrien, "Decision and Estimation Classification", New York, John Wiley, 1989.
- [14] R. L. Kettig and D. Landgrebe, "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects", *IEEE Transactions on Geoscience Electronics*, vol. GE-14, No. 1, pp. 19-25, 1976.
- [15] D. A. Landgrebe, "The Development of a Spectral-Spatial Classifier for Earth Observational Data", *Pattern Recognition*, Vol. 12, No. 3, pp. 165-175, 1980.