On Progress Toward Information Extraction Methods for Hyperspectral Data

David Landgrebe

School of Electrical & Computer Engineering
Purdue University
West Lafayette IN 47907-1285
landgreb@ecn.purdue.edu

ABSTRACT

A focused research program has been under way for several years to discover optimally effective means for analysis of multispectral and hyperspectral data. The methods pursued are based upon fundamental principals of signal theory and signal processing. The basic approach revolves around viewing N spectral bands of data from a pixel as a single point in N dimensional space, thus, an important aspect of the work has been to discover unique aspects of higher dimensional spaces which can be exploited for their information-bearing aspects.

Substantial progress on this problem has been made in the last several years, with several key algorithms having been defined. Among these are algorithms for transforms which define optimal case-specific features, and which improve the ability of the classifier to generalize. A more fundamental finding has been to understand the characteristics of high dimensional space and the significance of design samples and their use in defining the classifier.

These results have been published in separate papers over the last several years. The purpose of this paper is to survey these results and to show how they relate to one another in achieving an effective overall analysis procedure for analyzing a hyperspectral image data set.

Keywords: hyperspectral, multispectral, data analysis, information extraction

BACKGROUND

The multispectral approach to the mapping of land surface cover has been a key approach for three decades[1]. A principal motivation for it is that it makes possible identification and mapping of cover types without the need to use very high spatial resolution, thus greatly reducing the cost of the sensor systems and the volume of data that results. However, until recent years, land remote sensing has been substantially limited by the relatively small numbers of spectral bands that could be built into spaceborne sensor systems. A principle factor causing this limitation in the early days was the state of solid state device development which necessitated that one or a small number of detectors had to be scanned across the scene for each spectral band, usually by mechanical means.

In recent years, advances in solid state detector technology has made possible two dimensional arrays of considerable size, thus allowing for many scene pixels to be viewed in parallel, greatly increasing the dwell time per pixel and thus allowing for both large increases in the number of spectral bands possible and the signal-to-noise ratio present in each. It should be possible to derive information from such data of much greater detail and to higher accuracy than was previously possible.

However, increasing the number of bands available from four or seven, as is the case with Landsat data, to several hundred as now possible, and increasing the detail of measurement from that provided by 6 or 8 bit data systems to those of 10 or 12 bits due to the improved signal-to-noise ratios means that the measurement complexity for each pixel has increased by several orders of magnitude. It seems clear, then, that previously successful data analysis processes, though they might still be useful in some cases, will not yield the full potential that such new data provides. Whole new approaches to data analysis, specifically designed for the remote sensing context and this high dimensional data, must be devised. Unfortunately, to achieve the full potential of such data calls for procedures which are less intuitive that those simpler ones of the past. They must be based upon theoretically sound principles of signal theory and yet they must be made acceptable to Earth scientists and remote sensing practitioners who are not signal processing engineers.

Presented at the SPIE 42nd Annual Meeting, San Diego CA, 27 July-1 August 1997

Anticipating this situation, a research program was begun several years ago to meet this need. The basic approach has as its origin the technology that has grown out of the communication engineering field of the last three quarters of a century. More specifically, it has been to seek a more fundamental understanding of high dimensional signal spaces in the context of the remote sensing problem, and then to use that knowledge to extend the methods of conventional multispectral analysis to the hyperspectral domain in an optimal or near optimal fashion. In what follows, we shall outline what has been learned to this point.

THE APPROACH

One must begin with the understanding that the subject matter, the Earth's surface, is quite complex. Not only is the spatial frequency of land scenes very high, much higher that of water or cloud scenes, but the pattern of reflected light is very dynamic in time, changing the wavelength distribution of energy emanating from the surface with time constants often of the order of minutes or seconds, especially over vegetated areas.

The matter of how the variations are represented mathematically and conceptually is an important first step in defining how the analysis process should proceed. There have been three principal ways in which multispectral data are represented quantitatively and visualized. See Figure 1.

- In image form, i. e., pixels displayed in geometric relationship to one another,
- As spectra, i. e., variations within pixels as a function of wavelength, and
- In feature space, i. e., pixels displayed as points in an N-dimensional space.

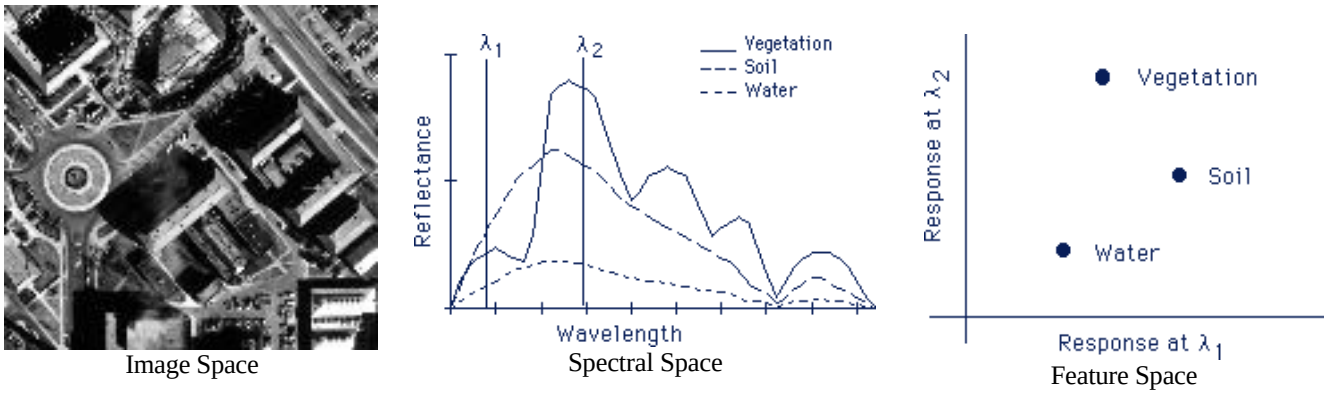

Figure 1. Three forms for representing multispectral data.

Image Space. Though the image form is perhaps the first form one thinks of when first considering remote sensing as a source of information, its principal value has been somewhat ancillary to the central question of deriving thematic information from the data. Data in image form serve as the human/data interface in that image space helps the user to make the connection between individual pixel areas and the surface cover class they represent. It also provides for supporting area mensuration activities usually associated with use of remote sensing techniques. Thus, it becomes very important as to how accurately the true geometry of the scene is portrayed in the data. However, it is the latter two of the three means for representing data that have been the point of departure for most multispectral data analysis techniques.

Spectral Space. Here, attention is focused on individual pixels, showing how reflected or emitted energy varies with wavelength. Many analysis algorithms which appear in the literature begin with a representation of a response function as a function of wavelength. Early in the work, the term "spectral matching" was often used, implying that the approach was to compare an unknown spectrum with a series of pre-labeled spectra to determine a match, and thereby to identify the unknown. This line of thinking has led, at various times, to attempts to construct a "signature bank," a dictionary of candidate spectra whose identity had been pre-established. Such an approach then places a heavy reliance upon calibration of each newly collected data set.

A second example of the use of spectral space is the "imaging spectrometer" concept, whereby identifiable features within a spectral response function, such as absorption bands due to resonances at the molecular level, can be used to identify a material associated with a given spectrum. This approach, arising from the concepts of chemical spectroscopy which has long

been used in the laboratory for molecular identification, is perhaps one of the most fundamentally cause/effect based approaches to multispectral analysis.

Feature Space. The third basis for data representation also begins with a spectral focus, i.e., that energy or reflectance vs. wavelength contains the desired information, but it is related to vector spaces rather than pictures or graphs. It began by noting that the function of the sensor system inherently samples the continuous function of emitted and reflected energy vs. wavelength and converts it to a set of measurements associated with a pixel which constitute a vector, i.e., a point in an N-dimensional vector space. This conversion of the information from a *continuous* function of wavelength to a *discrete point* in a vector space is not only inherent in the operation of a multispectral sensor, it is very convenient if the data are to be analyzed by a machine-implemented algorithm. It, too, is quite fundamentally based, being one of the most basic concepts of signal theory. Further, it is a convenient form if a more general form of feature extraction is to precede the analysis step, itself. As will be seen below, of the three data representations, the feature space provides the most powerful one from the standpoint of information extraction.

Another key characteristic which is fundamental to the engineering task of optimally designing a data analysis system is the basis for the mathematical representation of the data. A number of approaches have been considered for multispectral data over the years. The following are some examples.

- Deterministic Approaches
- Stochastic Models
- Fuzzy Set Theory
- Dempster-Shafer Theory of Evidence
- Robust Methods, Theory of Capacities, Interval Valued Probabilities
- Chaos Theory and Fractal Geometry
- AI Techniques, Neural Networks

All of these approaches have been examined to varying degrees, and each has certain facets which are attractive. Deterministic approaches, for example, tend to be the most intuitive. This is important in a multidisciplinary field such as remote sensing, where different workers have different backgrounds. However, deterministic methods tend not to be as powerful, and may have other disadvantages such as being more sensitive to noise than is necessary.

Having investigated each, we have based our work on the stochastic or random process approach[2,3]. This approach has the advantage of rigor and power, and, due to its maturity, has a large stable of tools that prove of pivotal usefulness in the work.

ON THE SIGNIFICANCE OF SECOND ORDER STATISTICS

Use of a stochastic process approach for modeling the spectral response of a ground scene requires determining the class probability distributions for each given data set. Using a parametric model for such modeling, this reduces the problem to that of accurately determining the mean vector and the covariance matrix in N-dimensional feature space for each class of ground cover to be identified.

As previously indicated, one of the advantages of the stochastic process approach is the wealth of mathematical tools available using this method. For example, it is frequently the case that one would like to calculate the degree of separability of two spectral classes in order to project the accuracy it is possible to achieve in discriminating between them. There are available in the literature a number of "statistical distance" measures for this purpose. They measure the statistical distance between two distributions of points in N-dimensional space. One with particularly good characteristics for this purpose is the Bhattacharyya Distance. In parametric form it is expressed as follows.

$$B = \frac{1}{8}[\boldsymbol{\mu}_1-\boldsymbol{\mu}_2]^T\left[\frac{\Sigma_1+\Sigma_2}{2}\right]^{-1}[\boldsymbol{\mu}_1-\boldsymbol{\mu}_2] + \frac{1}{2}\,\mathrm{Ln}\,\frac{\left|\frac{1}{2}[\Sigma_1+\Sigma_2]\right|}{\sqrt{|\Sigma_1|\,|\Sigma_2|}} \qquad (1)$$

where $\boldsymbol{\mu}_i$ is the mean vector for class i and Σ_i is the corresponding class covariance matrix. This distance measure bears a nearly linear, nearly one-to-one relationship with classification accuracy. Examining this equation, one sees that the first term on the right indicates the part of the net class separability due to the difference in mean values of the two classes, while the second term indicates the portion of the total separability due to the class covariances. This makes clear from a quantitative point of view what the relationship is between first order variations (the first term on the right) and second order variations

(the second term on the right). This illustrates, for example, that two classes can have the same mean value, and still be quite separable. Note that methods which are deterministically based only can make use of separability measured by the first term.

An example classification from a recent paper will further illuminate the matter[4]. For this experiment, a multispectral data set with a large number of spectral bands was analyzed using standard pattern recognition techniques. The data were classified using first a single spectral feature, then two, and continuing on with greater and greater numbers of features. Three different classification schemes were used, (a) a standard maximum likelihood Gaussian scheme, in which both the means and the covariance matrices, i.e., both first and second order variations, were used, (b) the same except with the mean values of all classes adjusted to be the same, so that the classes differed only in their covariances, and (c) using a minimum distance to means scheme such that mean differences are used, but covariances are ignored. It is seen from the results shown in Figure 11 below that case (a) produced clearly the best result, as would be expected.

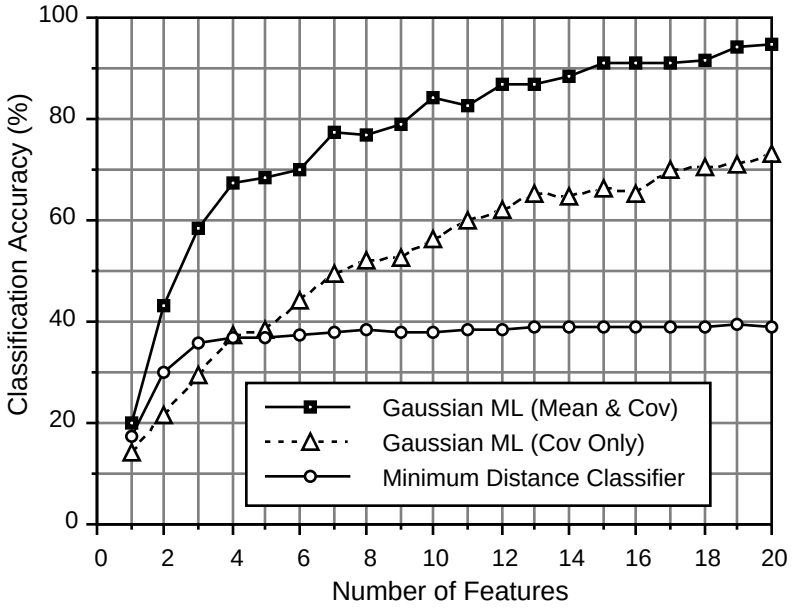

Figure 2. Performance comparison of the Gaussian ML classifier, the Gaussian ML classifier with zero mean data, and the minimum distance classifier.

In comparing the latter two, though, it is seen that, at first in low dimensional space, the classifier using mean differences performed best. However, as the number of features was increased, this performance soon saturated, and improved no further. On the other hand, while the classifier of case (b) which used only second order effects, was at first the poorest, it soon outperformed the one of case (c) and its performance continued to improve as greater and greater numbers of features were used. Thus it is seen that second order effects, in this case represented by the class covariances, are not particularly significant at low dimensionality, but they become so as the number of features grows, to the point that they become much more significant than the mean differences between classes at any dimensionality. It is, of course, also possible to show other example classifications where the mean vector dominates over the covariance[5].

However, the potential advantage of second order effects can be easily lost if increased precision in determining the class distributions is not achieved. This is what is dealt with in the following section.

ANCILLARY INFORMATION AND CLASSIFIER SUPERVISION.

From the vantage point of the above, it is clear that analysis methods which utilize both first and second order statistics can provide superior performance compared to those which utilize only first order effects. However, in many cases, this is not what is observed in practice. The explanation for this becomes apparent from the following.

With regard to the ability to discriminate between a pair of classes, an illuminating theoretical result appeared in the literature some years ago[6]. In this paper, the result shown in Figure 3 was derived. The ordinate for the curves in this figure is the mean recognition accuracy for the two class case, averaged over the ensemble of classifiers. The abscissa is measurement complexity, which in the case of multispectral data, is directly related to the number of bands and the number of gray values

per bands. The parameter for the different curves of the graph is m, the number of training samples. It is seen that each curve (except for the m → ∞ case) has a maximum, indicating that there is a best measurement complexity. It depends upon how many training samples one has, and thus how precise is the estimate of the class distributions.

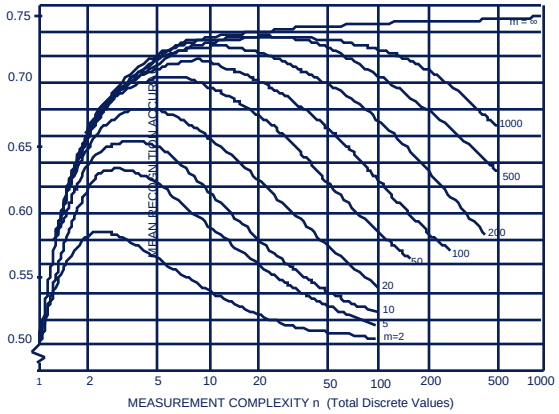

Figure 3. Mean Recognition Accuracy vs. Measurement Complexity for the finite training case.

It is important to note that the maximum of the curves moves upward and to the right as m increases, indicating that one can expect, on the average, to see improved performance as one increases the measurement complexity, but to achieve it, one will need increased precision in estimating the class distributions. The curve also shows that, given a complex enough measurement (enough features and enough values per feature) and enough training samples, one can expect to achieve arbitrarily high accuracy.

ON CLASSIFIER COMPLEXITY

In practical terms, the result of Figure 3 reduces to the following. The means for quantitatively describing a class distribution from a finite number of training samples commonly comes down to estimating the elements of the class mean vector and covariance matrix. When the number of training samples is limited, as it nearly always is in remote sensing, and the dimensionality of the data becomes large, the needed relationship between the training set size and the number of matrix elements that must be estimated quickly becomes strained even in the parametric case. This is especially true with regard to the covariance matrix, whose element population grows very rapidly with dimensionality. For example, the following table illustrates the number of elements in the various covariance matrix forms which must be estimated for the case of 5 classes and several different numbers of features, p.

No. of Features p	Class Covar. $5\{p^2 - [(p-1)^2 + (p-1)]/2\}$	Diagonal Class Common Covar. $5p$	Common Covar. $\{p^2 - [(p-1)^2 + (p-1)]/2\}$	Diagonal Common Covar. p
5	75	25	15	5
10	275	50	55	10
20	1050	100	210	20
50	6375	250	1275	50
200	100,500	1000	20,100	200

Table 1. Number of elements in various covariance matrix forms to be estimated. A case for 5 classes is assumed.

A training set of 1000 samples may sound large, and it is for a 5-dimensional problem. However, it is not so large for a 10-dimensional case, and definitely inadequate for a 20-dimensional problem. It is well known that one must have at least one more sample than the number of dimensions in order for a covariance estimate to not be singular. But just barely exceeding this amount still will not provide good results, if the separation between classes is at all dependent upon second order effects.

The relationship between training set size and dimensionality has been examined quantitatively[7], and it has been found that, as the dimensionality goes up (or the number of labeled samples available goes down), it may be advantageous to reduce the number of elements that must be estimated by reducing the algorithm complexity, i.e., by deciding between using individual class covariance matrices, a common covariance matrix, and a diagonal common covariance matrix. This allows for a more precise estimation of the parameters needed. The tradeoff of gaining precision by reducing complexity when the training sets are limited, can result in improved accuracy of classification. It has been codified into a scheme referred to as LOOC (Leave One Out Covariance) estimation which can be relatively transparent to the user. The scheme is as follows. The quantity to be estimated is $C_i(\alpha_i)$, where,

$$C_i(\alpha_i) = \begin{cases} (1-\alpha_i)\mathrm{diag}(\Sigma_i) + \alpha_i \Sigma_i & 0 \le \alpha_i \le 1 \\ (2-\alpha_i)\Sigma_i + (\alpha_i - 1)S & 1 < \alpha_i \le 2 \\ (3-\alpha_i)S + (\alpha_i - 2)\mathrm{diag}(S) & 2 < \alpha_i \le 3 \end{cases} \qquad (2)$$

Σ_i is the sample covariance matrix for class i, estimated from the training samples. The common covariance is defined by the average sample covariance matrix $S = \frac{1}{L}\sum_{i=1}^{L} \Sigma_i$ where a total of L classes are assumed. The variable α_i is a mixing parameter that determines which estimate or mixture of estimates is selected. If $\alpha_i = 0$, the diagonal sample covariance is used. If $\alpha_i = 1$, the estimator returns the sample covariance estimate. If $\alpha_i = 2$, the common covariance is selected, and if $\alpha_i = 3$ the diagonal common covariance results. Other values of α_i lead to mixtures of two estimates. Projected accuracy is estimated a priori by the well-known leave-one-out method using the available training samples.

In addition to these methods, additional aspects of classifier design have been investigated, including more complex decision logic[8,9] and ways to speed the classification computation[10,11]. With the rapid increase of computational processor speeds in recent years, processing speed has turned out not to be the pressing problem it once was, and until the more pressing problems of the analysis process are solved, complex decision logic potentials can also reasonably be postponed. Thus these aspects are being pursued at a lower priority.

One additional aspect of classifier design which appears to have significant utility has also been investigated. It has been shown[12,13] that by adding unlabeled samples to the classifier design process, better estimates for the discriminant functions can be obtained. This has resulted in an algorithm referred to as "Statistics Enhancement." The algorithm iterates between the labeled (training) samples and unlabeled samples from the data set to modify the class statistics so that a better fit to the overall data distribution is obtained. In this way, the ability of the classifier to generalize beyond its training samples is improved.

GEOMETRICAL, STATISTICAL AND ASYMPTOTICAL PROPERTIES OF HIGH DIMENSIONAL SPACES

The previous sections of this paper apply equally well to conventional multispectral data. In this section[14], we will describe some of the unique or unusual aspects of hyperspectral data, in order to illuminate some of the circumstances which must be accounted for in dealing with hyperspectral data in an optimal fashion.

For a high dimensional space, as dimensionality increases:

A. The volume of a hypercube concentrates in the corners[15]

It has been shown[16] that the volume of a hypersphere of radius r and dimension d is given by the equation:

$$V_s(r) = \text{volume of a hypersphere} = \frac{2r^d}{d}\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \qquad (3)$$

and that the volume of a hypercube in $[-r, r]^d$ is given by the equation:

$$V_c(r) = \text{volume of a hypercube} = (2r)^d \tag{4}$$

The fraction of the volume of a hypersphere inscribed in a hypercube is:

$$f_{d1} = \frac{V_s(r)}{V_C(r)} = \frac{\pi^{d/2}}{d2^{d-1}\Gamma\left(d/2\right)} \tag{5}$$

where d is the number of dimensions. It can be readily verified that f_{d1} decreases rapidly as the dimensionality increases. For example, while nearly 80% of the volume of the cube is contained in the hypersphere for d = 2, the percentage is reduced to less than 5% by d = 7. Note that $\lim_{d\to\infty} f_{d1} = 0$ which implies that the volume of the hypercube is increasingly concentrated in the corners as d increases.

B. The volume of a hypersphere concentrates in an outside shell[17,18]

The fraction of the volume in a shell defined by a sphere of radius r-ε inscribed inside a sphere of radius r is:

$$f_{d2} = \frac{V_d(r) - V_d(r-\varepsilon)}{V_d(r)} = \frac{r^d - (r-\varepsilon)^d}{r^d} = 1 - \left(1 - \frac{\varepsilon}{r}\right)^d \tag{6}$$

For the case ε = r/5, as the dimension increases the volume in the outside shell increases from about 35% for d = 2 to nearly 90% for d = 10. Note that $\lim_{d\to\infty} f_{d2} = 1, \quad \varepsilon > 0$, implying that most of the volume of a hypersphere is concentrated in an outside shell.

These characteristics have several important consequences for high dimensional data that appear immediately.

- High dimensional space is mostly empty, which implies that multivariate data in a high dimensional feature space is usually in a lower dimensional structure. As a consequence high dimensional data can be projected to a lower dimensional subspace without losing significant information in terms of separability among the different statistical classes.
- Normally distributed data will have a tendency to concentrate in the tails.

Similarly,

- Uniformly distributed data will be more likely to be collected in the corners,

making density estimation more difficult. Local neighborhoods are almost surely empty, requiring the bandwidth of estimation to be large and producing the effect of losing detailed density estimation.

C. The diagonals are nearly orthogonal to all coordinate axes[19,20]

The cosine of the angle between any diagonal vector and a Euclidean coordinate axis is:

$$\cos(\theta_d) = \pm \frac{1}{\sqrt{d}} \ , \tag{7}$$

Note that $\lim_{d\to\infty} \cos(\theta_d) = 0$, which implies that in high dimensional space the diagonals have a tendency to become orthogonal to the Euclidean coordinates.

This result is important because,

- The projection of any cluster onto any diagonal, e.g., by averaging features, could destroy information contained in multispectral data.

D. The required number of labeled samples for supervised classification increases as a function of dimensionality, and more rapidly so for more complex classification algorithms.

Fukunaga[21], for example, in a given circumstance, proves that the required number of training samples is linearly related to the dimensionality for a linear classifier and to the square of the dimensionality for a quadratic classifier. That fact is very relevant, especially since experiments have demonstrated that there are circumstances where second order statistics are more relevant than first order statistics in discriminating among classes in high dimensional data[22]. In terms of nonparametric classifiers the situation is even more severe. It has been estimated that as the number of dimensions increases, the sample size needs to increase exponentially in order to have an effective estimate of multivariate densities[23,24].

It is reasonable to expect that high dimensional data contains more information in the sense of a capability to detect more classes with more accuracy. As a matter of fact, since the curve of Figure 3 for the $m \rightarrow \infty$ case is montonically increasing, ultimately one can expect 100% accuracy, on the average. At the same time the above characteristics tell us that current techniques, which are usually based on computations at full dimensionality, may not deliver this advantage unless the available labeled data is substantial.

E. For most high dimensional data sets, low dimensional linear projections have the tendency to be normal, or a combination of normal distributions, as the dimension increases.

That is a significant characteristic of high dimensional data that is quite relevant to its analysis. It has been proved[25,26] that, as the dimensionality tends to infinity, lower dimensional linear projections will approach a normal (Gaussian) distribution with probability approaching one. Normality in this case implies a normal or a combination of normal distributions. This lends credence to using Gaussian classifiers after having reduced the dimensionality via feature extraction and indeed, to using class mean vectors and covariance matrices in evaluating the separability of classes. Properly used, parametric classifiers should provide good performance, and nonparametric schemes, with their higher demands for training data, should not be needed.

FEATURE EXTRACTION.

The findings above point to the importance of finding the lowest dimensional subspace to use for classification purposes. Thus, feature extraction becomes an important tool in the analysis process for hyperspectral data. As a result, feature extraction methods already existing in the literature were studied relative to the remote sensing context. The most suitable appeared to be Discriminate Analysis Feature Extraction (DAFE). Even so, it has several significant shortcomings for this environment, among them being that it does not perform well for cases where there is little difference in class mean vectors. It also only generates reliable features up to one less than the number of classes for the given problem. For use in problems where these shortcomings would be serious, Decision Boundary Feature Extraction (DBFE) was created[27,28,29].

However, both DAFE and DBFE calculations begin with computation in the full dimensional space in order to find the optimal transformation to a lower dimensional space, thus these calculations may, too, suffer from small training set limitations. To deal with this limitation, a Class-Conditional Pre-Processing algorithm was designed based upon a method known as projection pursuit[30,31]. This algorithm does the necessary calculations in the projected space, rather than the original, high dimensional space.

Figure 20 then shows the overall scheme. The data at point Φ might be 200 or more dimensional. Through projection pursuit, a subspace of perhaps 20 dimensions might be determined, and in this case, all calculations are done at a dimensionality of 20. This can then more optimally be followed by DAFE or DBFE to find a subspace of perhaps 10 dimensions in which to do the classification. In this way, maximal advantage can be taken of a training set of limited size.

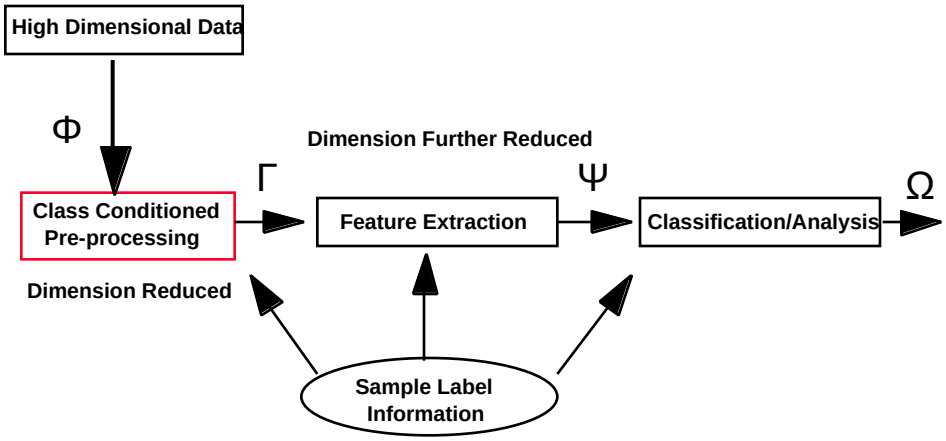

Figure 20. Classification of high dimensional data including reprocessing of high dimensional data.

SUMMARY AND CONCLUSIONS

What is sought are powerful, general analysis procedures that approach the optimum in information extraction capabilities and yet are within the reach of and practical for a broad range of Earth scientists. The techniques must be 1) powerful in terms of accuracy and detail with respect to the classes which can be discriminated, 2) objective in their performance, 3) robust with regard to the breadth of discipline problems which can be successfully approached, and yet 4) must appear sound and practical to scientists with any of a broad set of discipline backgrounds. They must be derived with appropriate mathematical rigor, but in the end, they must meet the practical conditions of the randomness of the scene, noise introduced by the atmosphere, the scene, and the sensor, and the varied skills and expectations of the users.

Summarizing, key conclusions expressed above are,

- The fact that hyperspace is mostly empty and that the data structure is in a lower dimensional subspace points to the importance of feature extraction algorithms that are able to find the optimal subspace.
- In the hyperspectral case, second order statistics, which define the shape of the class distribution in hyperspace, take on added significance.
- Training sets, by which to describe quantitatively the distributions of the classes of interest, are usually quite limited in size, in the face of the importance of second order statistics, which are quite sensitive to training set size especially as the dimensionality goes up. This means that it is especially important to use the combination of estimation procedures and the dimensionality appropriately, lest the advantages that the dimensionality should provide be lost.
- If the training set size is quite limited, it may be appropriate to reduce the classifier complexity by reducing the number of parameters that must be estimated.
- It is possible to use unlabeled samples in conjunction with the labeled ones to improve the generalization capability of a classifier.

And finally, it is recognized that a key problem is to deliver the knowledge and algorithms derived during this research to the potential users. To aid in this process, an application program for personal computers has been created with a basic multispectral data analysis capability and made available to the community without charge. Then as new algorithms emerge from the research, they are incorporated into the program and new versions of it issued. In this way, new algorithms, which may be quite complex to implement may be tried by users with a minimum of effort on their part. The program, called MultiSpec, together with substantial documentation is available for anyone interested to download from the world wide web. The URL for the web site is http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/. Some of the algorithms mentioned above

which it now contains are, Discriminate Analysis Feature Extraction (DAFE), Decision Boundary Feature Extraction (DBFE), and Statistics Enhancement. Additional ones resulting from the research but not described here are included as well.

A longer and more complete report of the work summarized in this paper, in the form of a white paper, is available for downloading from the Documentation of the MultiSpec web site referred to above. Further, several of the referenced published papers, which contain details of the individual algorithms mentioned above, may be downloaded from that site as well.

ACKNOWLEDGMENT

Work leading to the material presented here was funded in part by NASA Grants NAGW-925(1986-94), NAG5-3924 (1994-97), and ongoing Grant NAG5-3975.

REFERENCES

1 Landgrebe, David, "The Evolution of Landsat Data Analysis," (Invited), Photogrammetric Engineering and Remote Sensing, Vol. LXIII, No. 7, July 1997, Vol. LXIII, No. 7, July 1997, pp. 859-867.

2 Cooper, G. R. & C. D. McGillem, *Probabilistic Methods of Signal and System Analysis,* Second Edition, Holt, Rinehart & Winston, 1986, Chapter 7.

3 Papoulis, A., *Probability, Random Variables, and Stochastic Processes,* Second Edition, McGraw-Hill 1984.

4 Chulhee Lee and David A. Landgrebe, "Analyzing High Dimensional Multispectral Data," IEEE Transactions on Geoscience and Remote Sensing, 31, No. 4, pp. 792-800, July, 1993.

5 Jimenez, Luis, and David Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," *IEEE Transactions on System, Man, and Cybernetics*, To appear January, 1998. Downloadable from
http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html

6 G. F. Hughes, "On The Mean Accuracy Of Statistical Pattern Recognizers," IEEE Trans. Infor. Theory, Vol. IT-14, No. 1, pp. 55-63, 1968

7 Hoffbeck, Joseph P. and David A. Landgrebe, "Covariance Matrix Estimation and Classification with Limited Training Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, no. 7, pp. 763-767, July 1996.

8 B. Kim and D. Landgrebe, "Hierarchical Classifier Design in High Dimensional Numerous Class Cases," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 29, No. 4, July 1991, pp. 518-528.

9 S. Rasoul Safavian and David Landgrebe, "A Survey of Decision Tree Classifier Methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 21, No. 3, May/June 1991, pp. 660-674.

10 Chulhee Lee and David A. Landgrebe, "Fast Likelihood Classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 29, No. 4, July 1991, pp. 509-517.

11 Byeungwoo Jeon and David A. Landgrebe, "Fast Parzen Density Estimation Using Clustering-Based Branch and Bound," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 9, pp. 950-954, September 1994.

12 Behzad M. Shahshahani and David 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.

13 Behzad M. Shahshahani, "Classification of Multi-Spectral Data By Joint Supervised-Unsupervised Learning," PhD Thesis and School of Electrical Engineering Technical Report TR-EE-94-1, January, 1994.

14 Details for this section may be found in Luis O. Jimenez, "High Dimensional Feature Reduction Via Projection Pursuit," PhD Thesis and School of Electrical & Computer Engineering Technical Report TR-ECE 96-5, April 1996. See also reference [5].

15 Scott, D. W. "Multivariate Density Estimation." New York: John Wiley & Sons, 1992.

16 Kendall, M. G., A Course in the Geometry of n-dimensions, Hafner Publishing Co., 1961.

17 Kendall, M. G., A Course in the Geometry of n-dimensions, Hafner Publishing Co., 1961.

18 Wegman, E. J., "Hyperdimensional Data Analysis Using Parallel Coordinates," Journal of the American Statistical Association, Vol. 85, No. 411, pp. 664-675, 1990

19 Scott, D. W. "Multivariate Density Estimation." John Wiley & Sons, pp. 27-31, 1992.

20 Wegman, E. J., "Hyperdimensional Data Analysis Using Parallel Coordinates," Journal of the American Statistical Association, Vol. 85, No. 411, 1990

21 Fukunaga, K. "Introduction to Statistical Pattern Recognition." San Diego, California, Academic Press, Inc., 1990.

22 Chulhee Lee and David A. Landgrebe, "Analyzing High Dimensional Multispectral Data," IEEE Transactions on Geoscience and Remote Sensing, 31, No. 4, pp. 792-800, July, 1993.

23 Scott, D. W. "Multivariate Density Estimation." John Wiley & Sons, pp. 208-212, 1992.

24 Hwang, J., Lay, S., Lippman, A., "Nonparametric Multivariate Density Estimation: A Comparative Study.", IEEE Transactions on Signal Processing, Vol. 42, No. 10, 1994, pp. 2795-2810.

25 Diaconis, P., Freedman, D. "Asymptotics of Graphical Projection Pursuit." The Annals of Statistics Vol. 12, No 3 (1984): pp. 793-815.

26 Hall, P., Li, K. "On Almost Linearity Of Low Dimensional Projections From High Dimensional Data." The Annals of Statistics, Vol. 21, No. 2 (1993): pp. 867-889.

27 Chulhee Lee and David A. Landgrebe, "Feature Extraction Based On Decision Boundaries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, April 1993, pp. 388-400.

28 Chulhee Lee and David A. Landgrebe, "Decision Boundary Feature Selection for Non-Parametric Classification," *IEEE Transactions on System, Man, and Cybernetics*, Vol. 23, No. 2, March/April, 1993, pp. 433-444.

29 Chulhee Lee and David A. Landgrebe, "Decision Boundary Feature Extraction for Neural Networks," *IEEE Transactions on Neural Networks*, Vol. 8, No. 1, pp. 75-83, January 1997.

30 Luis Jimenez and David Landgrebe, "Projection Pursuit For High Dimensional Feature Reduction: Parallel And Sequential Approaches," Presented at the International Geoscience and Remote Sensing Symposium (IGARSS'95), Florence Italy, July 10-14, 1995.

31 Luis Jimenez and David Landgrebe, "Projection Pursuit in High Dimensional Data Reduction: Initial Conditions, Feature Selection and the Assumption of Normality," IEEE International Conference on Systems, Man, and Cybernetics, Vancouver, Canada, October 22-25, 1995.