Multispectral Data Analysis: A Signal Theory Perspective©

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

Preface

This document is intended to be a "How to..." document. How does one analyze multispectral data in a fundamentally sound, multivariate fashion? The audience in mind is an Earth science researcher who must conduct such analyses to obtain information needed for his/her research. After a brief introduction, there are two major parts, the first having to do with multispectral data of conventional dimensionality and the second on analysis of hyperspectral data. Each part begins with a brief discussion of the basic concepts to be applied. Example analyses using the MultiSpec data analysis software system are given to lead one through applying the theory covered in the concepts section to actually analyze an example data set. Copies of MultiSpec and the data are available from http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/.

The problem of analyzing airborne or spaceborne multispectral data is obviously a broad one, and clearly not all of the possible facets can be covered in such a brief document as this. The intent here is to pursue one particular facet of such analysis from beginning to end with enough depth to result in an analysis process which is well founded on sound principles of signal theory and their application. The particular facet chosen to be the vehicle here is that of analyzing the multispectral data into useful themes or classes by use of discriminant analysis. In the process, the added potential of hyperspectral data over data of conventional dimensionality is demonstrated.

Introduction

The analysis of multispectral image data is now a quite common task for Earth scientists and other users of such data. However, it is not a trivial task. Extraction of specific information from a multiband data set, especially as the number of bands becomes large, is a complex problem requiring the application of techniques based on rigorous fundamentals of both Earth science and signal processing theory if the full potential of such data is to be realized. Although many perfectly useful ad hoc, situation-specific techniques are described in the literature, in this document, we shall describe rigorously based procedures which are sufficiently general to be suitable for use in a broad range of both local and large area problems. The presentation is divided into two main parts. The first deals with analysis of multispectral data of conventional dimensionality (of the order of 10 to 20 spectral bands). The second deals with hyperspectral data (100 or more spectral bands). To clarify the motivation for the procedures, we shall begin with a very concise statement of the basic principles and concepts involved.

© Copyright 1994, 1998 by David Landgrebe

Permission is hereby granted to reproduce this document for any non-commercial purpose.

Analysis of Conventional Multispectral Data

Background and Concepts

Basic Tenant of Remote Sensing

We begin with the basic idea of remote sensing, that is, that information is available at the aperture of a sensor, based upon the electromagnetic fields emanating from the surface and arriving at the sensor, and in particular via the

- Spectral,
- Spatial, &
- Temporal Variations

of those fields. For simplicity, at this time we shall concentrate upon the optimal use of spectral and spatial variations, focusing primarily upon spectral characteristics, since multispectral data provides unique potential for information extraction in this case. We shall leave use of temporal ones to another time.

The Broad View and Signal Space Concepts

Figure 1 provides a systems overview of a passive, optical remote sensing system. It is seen that, using illumination provided by the sun, the sensor system views the portion of Earth of interest, and the resulting data are transmitted to the processing station. There, some type of preprocessing is often performed, followed by the application of an appropriate data analysis algorithm and delivery of the resulting information to the user. Use of ancillary data and human interaction with the data stream typically take place during preprocessing and information extraction. Thus, rather than being automatic, the processing in mind is more accurately described as human-assisted machine processing. This is significant because in a fully automatic system one gives up the strength that human intelligence and adaptability can add to the analysis process. We will thus proceed from the viewpoint that a well-designed analysis scheme should use the strengths of both human and machine, but do so in a manner that maintains economy and objectivity while maximally enhancing performance.

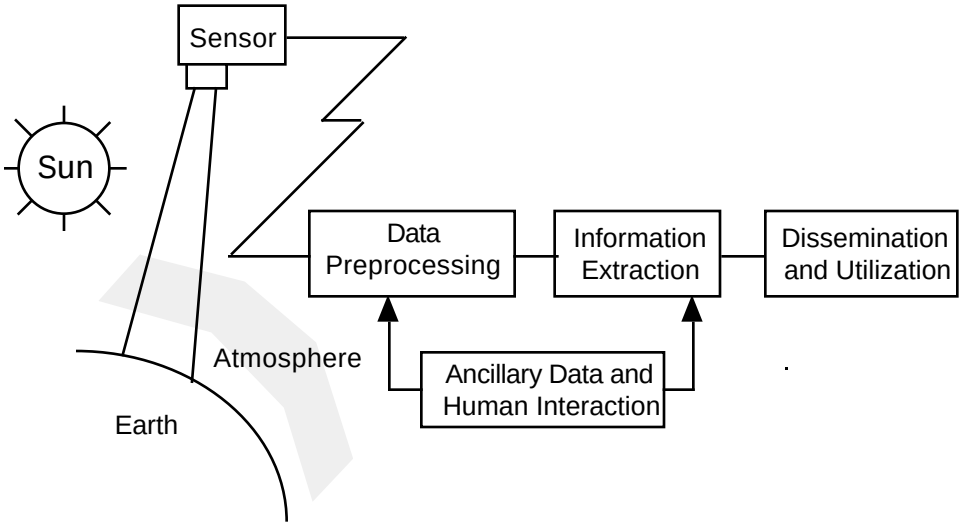


Figure 1. A Remote Sensing System Overview.

Figure 2 focuses attention more directly on the information extraction process, and casts it in the framework of a multivariate classifier. The classes to be recognized are patterns in the spectral variations that have been sensed.

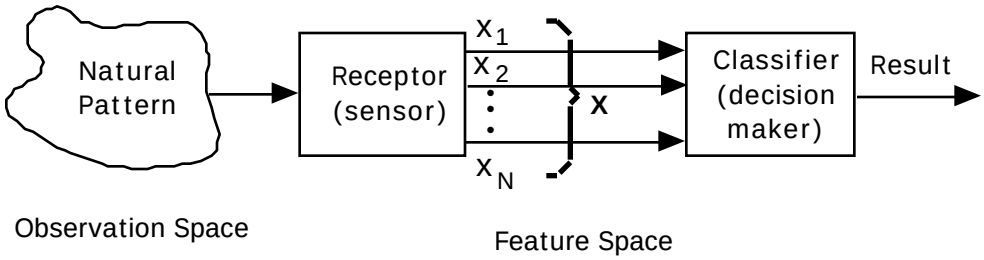


Figure 2. A Functional Diagram for a Remote Sensing System.

More specifically, Figure 2 draws attention to the fact that the job of the sensor system is to transform the information contained in the observation space into a feature space. In the context of a multispectral remote sensing system, this amounts mathematically to transforming *continuous* spectral functions into a *finite dimensional discrete* space. As an example, Figure 3 shows sample spectral patterns from three simple classes in (continuous) spectral space being sampled at two discrete wavelengths, λ_1 and λ_2, and they then appear as three points in a (two dimensional) feature space.

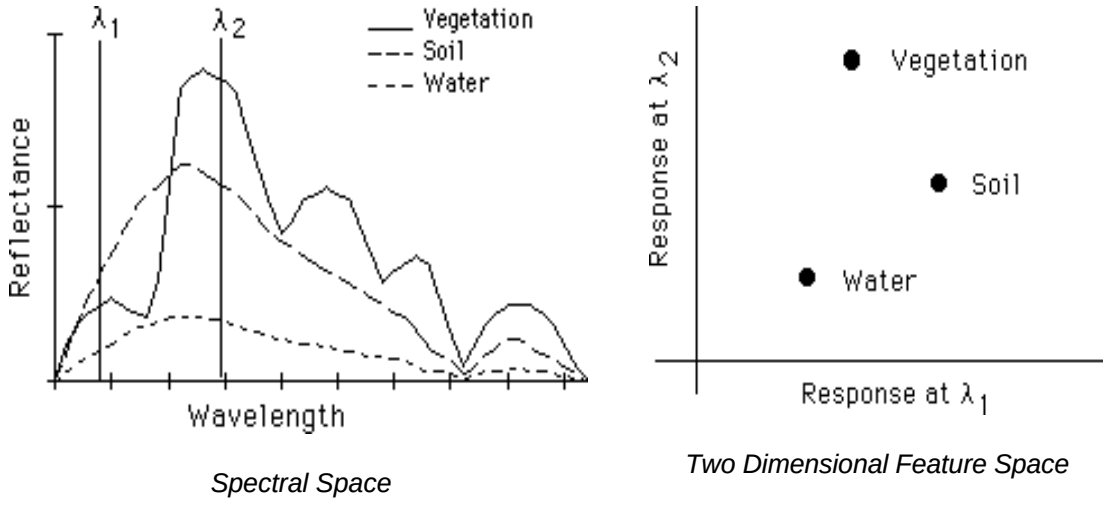
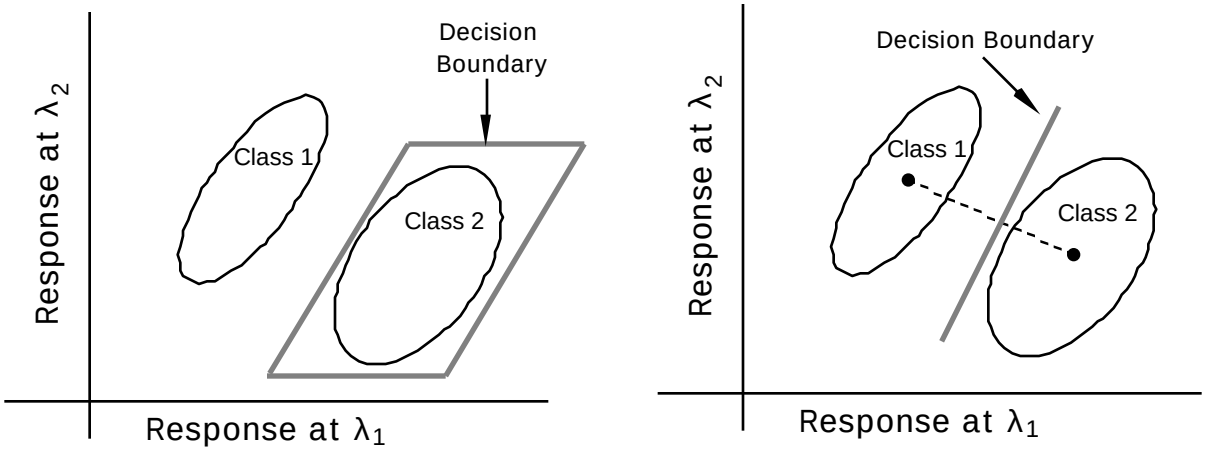


Figure 3. Spectral Space and Feature Space Representations of Spectral Responses.

Of course, sampling at additional wavelengths in Spectral Space would be possible; the result would be to increase the dimensionality of the resulting feature space. Though spaces of dimensionality greater than three cannot be diagrammed for human view, there is no fundamental reason why higher dimensional spaces could not be used and would be potentially beneficial. It is also the case that transformations more complex than simple, discrete wavelength sampling are possible, and frequently desirable. This becomes increasingly important with sensors with large numbers of bands and will be treated later.

By this transformation, from Spectral Space to N-dimensional feature space, the problem of recognition of a pattern is thus converted into one of determining into what part of the N-dimensional feature space a given unknown falls so that its association with the appropriate class can thus be determined. There are many ways that this could be accomplished. Two examples are shown in the following diagrams.



(a) Parallelepiped Classifier

(b) Minimum Distance Classifier

Figure 4. Decision Boundaries for Two Classification Algorithms

Assume that data from two classes falls in the two areas indicated in Figure 4 (a) and (b) above. The Parallelepiped Classifier is of a type known as an “absolute classifier,” i.e., it identifies pixels as being of Class 2 if they fall within the region marked “decision boundary” in Figure 4(a). Absolute classifiers are ones, which decide in favor of a given class without regard to the location of other classes. It is generally the case that, though they may be conceptually simpler, absolute classifiers do not perform as well as “relative classifiers,” i.e., classifiers that decide in favor of a class after having considered all possible classes.

An example of a relative classifier is the Minimum Distance Classifier. In this case, one associates a given feature vector with the class which has the nearest mean value. In the case of the two classes of Figure 4, the decision boundary partitioning the feature space into Class 1 and Class 2 regions would be as shown in Figure 4(b). After locating the centroid_ or conditional class means for the two classes, the decision boundary becomes the perpendicular bisector of the line connecting these mean values.

These two algorithms may provide satisfactory performance in simple cases, but most often algorithms, which result in nonlinear decision boundaries, are required. For example, consider the case of the classes shown in Figure 5.

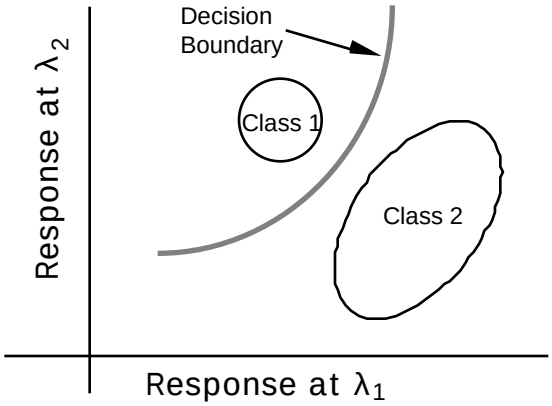


Figure 5. A nonlinear decision boundary.

In this case, the decision boundary is again a relative classifier, but with a second order decision surface. It is located not with respect to the class means alone, but taking into account the spread or distribution and orientation of the two classes. In particular, it is located halfway between the nearest points of the two classes, and its curvature accounts for the fact that with actual data, the points are not necessarily confined to a tightly defined region, but may scatter into outlying regions. This curvature accounts for the fact that Class 2 with its larger variance may be expected to scatter more than Class 1.

A classifier with the characteristics of this latter algorithm has been found in practice to be very practical, i.e., powerful enough to deal with typical data complexity, but not so complex as to require excessive calculation or training. Indeed, the concepts described here in rather simplistic terms in 2-dimensional space have been rigorously studied and derived based upon fundamentally sound principles of signal processing in noisy environments. This is certainly the situation one finds in practical remote sensing circumstances. Indeed, while it is customary when discussing classification algorithms, to show the classes as quite separable as was done above, in practice, the data usually occupy a continuum with no clumping or clustering into discrete units. Thus, one must choose classifiers and determine their parameters in some optimal way, recognizing that 100% separability is not usually possible.

Optimal System Design

It is, thus, desirable that the procedure to be used be quantitatively optimal in some known and appropriate sense. What must one have in order to carry out an optimal system design? The requirements are:

1. A Quantitative System Model. In this case this implies
   - Scene Model: (We will use Stochastic Process via Training Samples)
   - Sensor Model: (We will use a generalized sampling model)
   - Processing Model: (We will use a pattern classifier model)

2. An Optimality Criterion, objectively expressed. (We will use classification accuracy)

We have already described our means of sensor modeling, that of generalized sampling and transformation to N-dimensional space, as above. The Processing Model and Optimality Criterion then defined in terms of the partitioning of that N-dimensional space into non-overlapping regions, one each for each class present in the data. The matter of Scene Modeling bears some further comment.

A scene model describes the manner in which the continuous spectra (i.e., before the above transformation) is modeled. There are a number of theoretical approaches upon which to base such a model of the scene. Possibilities include

- A Deterministic Model
- Fuzzy Set Theory
- A number of others which are less well developed theoretically, e.g., Interval Valued Probabilities, Dempster-Shafer Theory of Evidence, Chaos and Fractal Technology.
- The Theory of Stochastic or Random Processes

Each of these has desirable features. A deterministic model has the advantage of conceptual simplicity, but does not take advantage of all information available from the spectra. Fuzzy Set Theory is focused on situations where knowledge of circumstances and parameters is not precise. The same is true for the third set, which also deal with cases when partially conflicting evidence about a given parameter may be present, characteristics which are certainly present in aspects of remote sensing.

However, these are not the key characteristics of such data. Rather, our focus is upon models that maximally convey information of interest. Thus, we choose to focus upon the high degree of the scene complexity. Even the noise-free spectral response of a given material of interest does not occur in practice as a deterministic response, but has a degree of variability about it. Added to this variability, to be sure, is noise or corrupting influences. However, since a significant part of this overall variability is diagnostic of the material itself, it is important to include it in the model. This is precisely the circumstance for which the *random process model* was designed. The theory of Random Processes[1] is very extensively and rigorously developed in the literature.[2,3,4,5] It has found its greatest use, for example, in the design of modern communication systems, due to its ability to deal with complex signals, containing complex mixtures of families of signals embedded in noise. As a result of its wide use, many theoretical tools and techniques are available for it.

---

1 Some texts use the name *Stochastic Process* in place of *Random Process.*

2 Davenport, W.B., Jr. and William L. Root, *An Introduction to the Theory of Random Signals and Noise,* McGraw-Hill, 1958.

3 Papoulis, A., *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill, 1965.

4 Helstrom, Carl W., *Probability and Stochastic Processes for Engineers,* 2nd Edition, Macmillan, 1991.

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

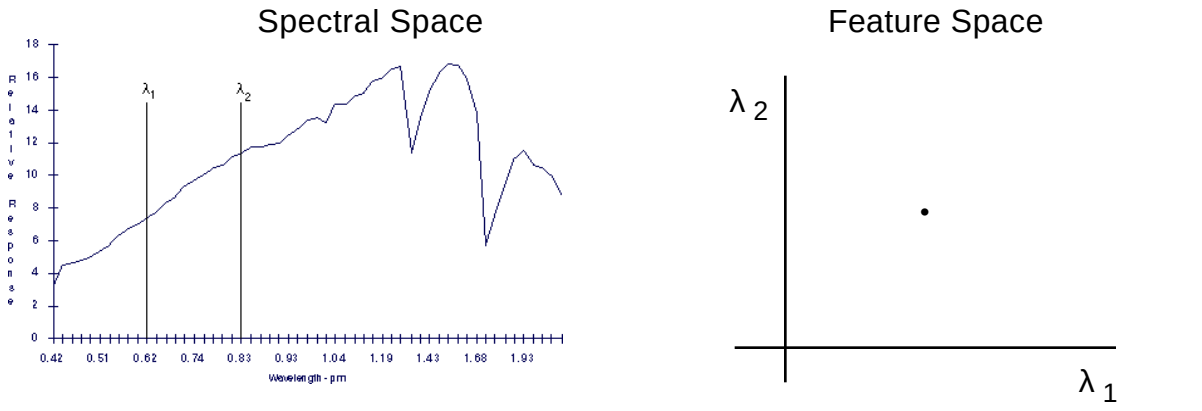
The formal definition of a Random Process is given as consisting of

- the ensemble of random functions, plus
- A probabilistic description about them.

In this case, what makes up the ensemble of the random functions is the list of all spectral responses over which the system is to be optimized, or at least a representative subset of them. What this amounts to, of course, depends upon the application context.

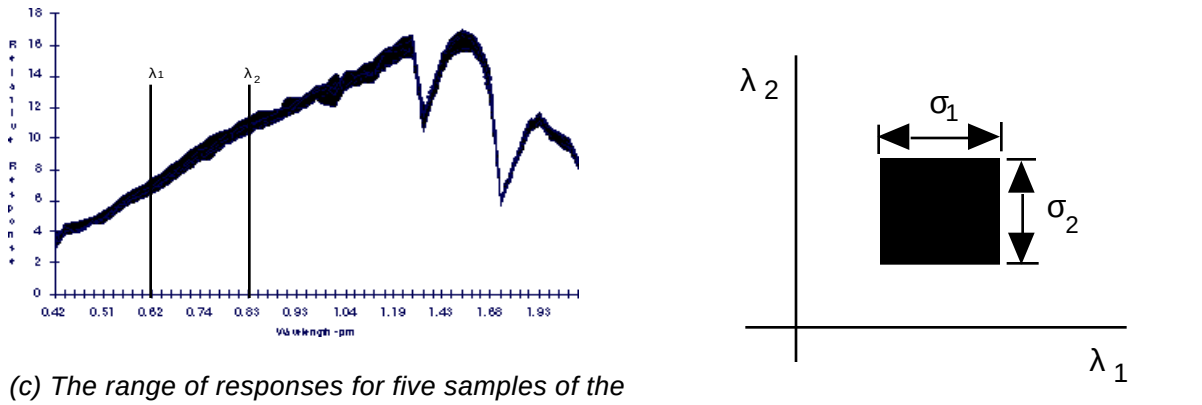
To make the implications of this theory more clear, Figure 6 shows a series of spectral response functions of a certain soil type, together with the corresponding feature space diagram for a particular two-dimensional space. Parts (a & b) of the figure show a deterministic model, as might be obtained by averaging several observations. Parts (c & d) show a “signal + noise” model of this soil. Part (c) was actually derived by displaying the maximum variation present between five samples of the same soil at each wavelength.

However, in order to describe fully this soil spectrally from the five observations, the data of parts (e & f) would be necessary. Note that the significant degree of correlation present between the two bands selected here, rather than being an indication of useless redundancy, serve to define the shape of the distribution in the feature space (f), and indeed indicate that the data in feature space is more compactly defined and thus less likely to overlap other nearby classes.



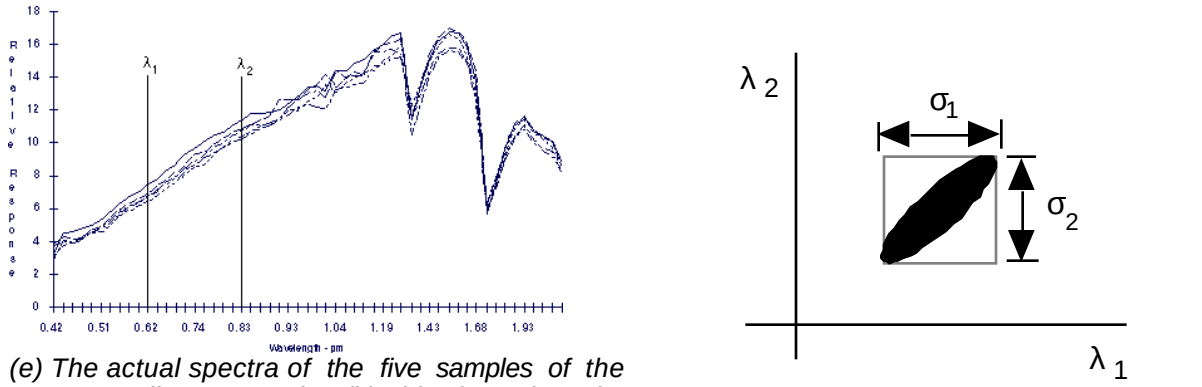
(a) The spectral response of a specific soil.

(b) The 2-D feature space equivalent for (a).



(c) The range of responses for five samples of the same soils. This amounts to a signal + noise model.

(d) The 2-D feature space equivalent for (c).



(e) The actual spectra of the five samples of the same soil. Compared to (b), this shows how the spectra vary about the mean.

(f) The 2-D feature space equivalent for (e).

Figure 6. Various models of the spectral response for a specific soil type.

This example helps to display why the random process model and the kind of transformation used is significant to the preservation of information in spectral features. Note that,

- In the deterministic case, only the location of the centroid of the spectral response is maintained. No information about how the spectra vary about their mean in spectral space or how the class distributes itself over the feature space is provided.

- In the signal + noise case, the centroid and some of the information of the shape of the distribution, that having to do with the range of the variation, are modeled, but not other characteristics of that variation.

- In the last of the three, if each of the five (and perhaps a lot more) 2-dimensional data points were to be plotted in the 2-dimensional space, they might only occupy the area indicated in black. If there were another class in the near vicinity of this class, there is less likelihood of overlap if the more complete representation of case 3 is used than that of case 2.

Thus, case 3 provides a more complete representation of the data. It shows not only how the data vary in each band, but also how they co-vary between bands, something that cannot be readily visualized in spectral space. In the feature space on the right, this characterization shows up as not only where in the space the centroid is, but also the shape of the variation about that centroid. Indeed, if the spectral space on the left were sampled at enough wavelengths and these samples expressed in all N of these dimensions, this would provide a complete characterization of how the data from this family of spectra vary, complete in the sense that the spectral space representation of each individual sample could be completely reconstructed from the feature space representation. Thus, one could be sure that no information had been lost in transforming from spectral space to feature space[6].

We note in passing that it has been found that the shape of the distribution becomes increasingly important as the dimensionality of the representation increases[7]. Indeed, for high dimensional cases such as for hyperspectral data, one must avoid using two or three-dimensional concepts, as the rules of geometry in high dimensional space are quite different than those of the two or three dimensional spaces which have formed our intuitive understandings. One must also avoid any type of preprocessing which, while it may adjust the location of the centroid in an assumed useful way, it also unknowingly distorts the shape of the distribution. Further we note that, with this model, there is neither requirement nor an advantage in the data being converted from radiance response to percent reflectance, nor to any particular set of radiometric units.

---

6 Indeed, a much more generalized means for transforming between spectral space and feature space exists. Mathematically this is expressed by the transformation pair,

$$x_n = \int_a^b X(\lambda)\ \varphi_n(\lambda)\ d\lambda \qquad X(\lambda) = \sum_{n=1}^{N} x_n \varphi_n(\lambda)$$

Where $X(\lambda)$ is the spectral function, x_n is the component in feature space and the φ_n are the (generalized) basis functions. The instantaneous sampling used above is just a special case of this, using a specific set of basis functions. This generalization becomes especially important and practical for hyperspectral data, where the sensor measures the value of $X(\lambda)$ at many λ's; It is also important for optimal feature extraction.

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

Under various circumstances, any of the above models might be used. In fact, all are found or implied by the various analysis methods found in the literature. Generally speaking, simpler classification tasks can be done with simpler data (fewer bands, fewer gray-values per band) using simpler models and correspondingly simpler classification algorithms. For example, a deterministically based decision algorithm might work satisfactorily for easily identifiable classes modeled in a deterministic fashion. One might also choose a linear classifier where somewhat more difficult classes have been modeled in signal-plus-noise fashion. However, to extract the most information (most or most detailed classes, highest accuracy), one will require more complex data and will want the most complete model. In this case, one would also want to use a decision algorithm that can make full use of the model complexity. Hereafter we will focus upon this latter possibility, using a full random process model and a classification scheme that can take advantage of both the location and shape of the class distributions in N-dimensional feature space.

Specifying the Classes

Clearly there must be some way for the analyst to specify to the analysis process what information is desired. That is, "How does one specify to the analysis algorithm the classes of materials desired to be identified?" There are a number of possible approaches to this part of the problem. For example, if one were to proceed with a deterministic model in mind, one might approach this part of the problem via previously stored "spectral signatures," i.e., previously measured spectral responses from the various classes of material of interest. While straightforward in concept, this approach has several significant drawbacks, two of which are the following.

(1) Extensive preprocessing is required. This is the case because one must reconcile the observation conditions present when the data were collected (atmosphere, sun angle, view angle, terrain effects, etc.) with the conditions under which the "spectral signatures" were measured. This requires a great deal of additional data, quite complex calculations, and is usually difficult to accomplish to precision approximating the precision of the original data.

(2) One must have already on hand the spectral responses of the materials to be identified. Further, since it is fundamentally true that relative decisions can be made more accurately than absolute ones, one must have spectral responses of other materials that may occur in the same scene as well. This makes the approach less robust against problems involving large areas and arbitrary locations over the Earth.

Instead, a simpler procedure from an operational standpoint is to label samples within the data set to be analyzed which display the characteristics of the desired classes. These samples are referred to as design samples or training samples, and in effect form the model of the classes to be identified. This allows the analysis process to be more robust in the sense that it can be applied to data collected from arbitrary sites. Further, because the design samples are from within the data set to be analyzed, many of the observation variables are accounted for, thus there is a reduced need for complex preprocessing. This is the approach in mind in the following.

The actual means by which labels for the design samples can be accumulated varies greatly with the situation. Often ancillary information or knowledge of the scene can be used to interpret the needed labels from imagery of the scene. In some large area classifications, low altitude photographic missions, which are relatively inexpensive to carry out, may provide the needed information[8]. In some cases, deterministic knowledge of specific spectral characteristics may be used[9]. Suffice it to say that it may well not be necessary to be able to collect "ground truth" from the ground in order to be able to accumulate an appropriate set of design samples.

The Gaussian Assumption

In the remote sensing situation, there is often an additional factor to the class modeling process. There is usually a paucity of these design samples by which to estimate the distribution in N-dimensional space for any given class. Thus, from a limited number of samples, one must determine as accurately as possible what the location and shape of the distribution of each class of samples is likely to be in feature space.

Experience has shown that, properly used, the assumption that each of the class data subsets may be modeled in terms of one or a combination of Gaussian distributions is a quite practical and powerful way to proceed. In feature space this is seen to amount to fitting one or a small number of Gaussian hypershapes to the actual distribution of each class. Among the advantages of using the Gaussian model is the mitigation of the need for large training sets to adequately define the desired classes, especially when the spectral dimensionality is large. However, use of this assumption does impose upon the analysis process some means for identifying the various modes of each desired information class and the fitting of densities to each of these modes. The analysis process then consists of finding the parameters for a complete set of Gaussian distributions which fit the data set and for which subsets of the distributions correspond to the classes of interest.

Basic Requirement of Feature Space Analysis

In summary, to properly train a classifier one must establish a list of classes in terms of their distributions in feature space, including training samples for each, which satisfy the following.

- The list of classes must be **exhaustive**, in the sense that there is a logical class to which to assign every pixel in the scene,
- The classes must be adequately **separable** using the available features, and
- The classes must be **of informational value**, i.e., they must be classes of interest to the user.

---

8 Bauer, M.E., T.E. Burk, A.R. Ek, P.R. Coppin, S.D. Lime, T.A. Walsh, D.K. Walters, W. Befort and D. F. Heinze, "Satellite Inventory of Minnesota Forest Resources," *Photogrammetric Engineering & Remote Sensing, Vol. LX, No. 3, pp. 287-298, March 1994.*

9 Hoffbeck, Joseph P. and David A. Landgrebe, "Classification of High Dimensional Multispectral Image Data," Fourth Annual JPL Airborne Geoscience Workshop, Arlington, Virginia, October 25-29, 1993.

An equivalent statement to this is that a well-trained classifier must have successfully modeled the distribution of the entire data set, but it must be done in such a way that the different classes of interest to the user are as distinct from one another as possible. What is desired in mathematical terms is to have the density function of the entire data set modeled as a mixture of class densities, i.e.,

$$p(\mathbf{x}|\theta) = \sum_{i=1}^{m} \alpha_i p_i(\mathbf{x}|\varphi_i) \qquad \text{and} \quad \sum_{i=1}^{m} \alpha_i = 1 \qquad (1)$$

where p is the probability density function describing the entire data set to be analyzed, $\mathbf{x}$ is the measured feature (vector) value, p_i is the density function of class i desired by the user, α_i is the probability of class i, m is the number of classes, and φ_i denotes the parameters necessary to define class i. Given the Gaussian assumption, these parameters are a set of $\overline{\mathbf{x}_i}$ and Σ_i, the mean vectors and covariance matrices for each class. Thus the task of analyzing a given data set reduces to finding the set of $\{\varphi_i: \overline{\mathbf{x}_i}, \Sigma_i\}$ which satisfy the exhaustive, separable, and informational value criteria.

Note again that even when classes are not Gaussian, the above equation is generally enough to handle this situation, in that one may model a given non-Gaussian class as a linear combination of Gaussian subclasses. In theory, by using enough such Gaussian subclasses to model a single class, one can assure that the model can fit any class to any given precision, e.g., by using a Parzen density approximation with a Gaussian kernel to fit an arbitrary nonparametric density. In practice, however, it is usually the case that only a small number (one or two) of such subclasses are necessary in actual cases.

Summary of the Concepts

The following is a brief summary of the concepts that have been presented as a sound basis for the analysis of multispectral data:

- The spectral variations measured at a pixel location play the major role in conveying extractable information via multispectral data. Spatial variations play a lesser but still significant role.

- Rather than achieving a fully automated analysis method, a more powerful goal is to achieve an analysis method that effectively combines human interaction with the quantitative strengths of machine methods.

- The role of the multispectral sensor is to transform the continuous spectral functions emanating from a pixel area to a finite dimensional discrete space. In the limit as N, the number of spectral samples sensed, increases, this N-dimensional space becomes complete in the sense that one could precisely recreate the original continuous spectral function from the coordinates of a sample in this space, thus implying that no information is lost in undergoing the transformation which the multispectral sensor accomplishes.

- The role of the analysis algorithm is to partition the resulting N-dimensional feature space into M mutually exclusive subregions, each one of which being a class of surface cover of interest to the analyst. Optimally locating the boundaries of these subregions, then, is the central task of any analysis process, and, indeed, is a perfectly general way of defining the analysis process, however implemented.

- Relative classifiers, which consider all possible classes in determining the optimal one, may generally be expected to out-perform absolute classifiers, which reach a decision based upon response characteristics defined by an absolute basis.

- A key concept is the manner in which the spectral response is viewed or modeled. Though any of a number of theoretically based approaches to this modeling could be taken, the most general and most thoroughly developed theoretically is the random process model. It has the advantage of dealing effectively with very complex classes in a noisy signal environment, but it can also be gracefully simplified to deal with less complex problem situations.

- The defining of the boundaries of the subregions of interest involves bringing together or reconciling the data collected and the conditions under which it was collected with pre-knowledge which delineates the information desired from the analysis. A useful way of accomplishing this reconciliation is to select samples from within the data set which exemplify the classes of interest, for this mitigates many of the non-information bearing variations present in the data. It thus reduces the need for preprocessing adjustments of the data, adjustments which many times can themselves unknowingly limit the information that can be extracted from the data.

- The use of the multivariate Gaussian model to model the distributions of the classes in N-dimensional feature space is an especially useful way to proceed. It is effective when there is the usual paucity of design samples by which to estimate class distributions, and by using more than one Gaussian distribution per class, it can be effective even when a class, itself, is not Gaussian.

- The goal of the analyst is thus to define a list of classes which is exhaustive, separable, and of informational value. Mathematically, this means that the density function of the entire data set must be partitioned as:

$$p(\mathbf{x}|\theta) = \sum_{i=1}^{m} \alpha_i p_i(\mathbf{x}|\varphi_i) \qquad \text{and} \quad \sum_{i=1}^{m} \alpha_i = 1$$

  Where p is the probability density function describing the entire data set to be analyzed, **x** is the measured feature (vector) value, p_i is the density function of class i desired by the user, α_i is the probability of class i, m is the

number of classes, and φ_i denotes the parameters necessary to define class i.

Turning Theory to Practice: How to Analyze a Data Set with MultiSpecã [10].

We shall next proceed to give an example of analysis of a data set that makes use of the concepts described above. In doing so, we shall use the facilities of the MultiSpecã data analysis software system. This system and the data set to be used are available at http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/ without charge to anyone desiring them; thus one is encouraged to follow along the analysis steps in parallel with their description below. It is assumed at this point that the reader has reviewed the documentation of MultiSpec and thus is familiar with the various algorithms implemented in it.

Before undertaking the example analysis, we will begin to relate the previously stated principles and concepts to the various processors contained in MultiSpec. Please note that the approach that has been described is quite general, and the applications to which it might successfully apply are quite varied. There are many specific procedures to analyzing multispectral image data, almost as many as there are analysts. The steps used in any given analysis must necessarily be based upon the scene, the information desired, and the data analyst's initial assumptions about the data characteristics. In general terms, to meet the requirements of a sound analysis as previously described, there are several helpful tools in MultiSpec:

- *Clustering* (sometimes called unsupervised classification) algorithms are focused upon finding classes of data that are **separable**. However, the resulting clusters do not necessarily have any relationship with the classes of informational value.

- *Training samples* focus the analysis upon classes of **informational value**. However, the classes initially so defined are not necessarily separable. Thus some combined or iterative use of clustering and training sample selection may be required.

- *Clustering* can also initially assist in obtaining a list of classes that are **exhaustive**. Later in the process a feature in MultiSpec called the *Probability Map* becomes valuable in determining any needed additional classes.

The problem is that data do not ordinarily occur in N-dimensional space as separable "clumps," let alone clumps of the desired informational classes. Rather, data in N-dimensional space usually exist as a continuum. Thus it is not possible for a clustering algorithm by itself to divide data into separable classes of interest.

There are two additional matters to be mentioned before turning to specific steps useful in analyzing data. The first has to do with choosing or calculating the best set of

---

10 MultiSpec © Purdue Research Foundation. A copy with its documentation may be obtained from http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/.

features to use for a given classification. There are at least two reasons why one might not want to use all of the features available in the data set to do a classification. One is that if adequate performance can be obtained with a subset of the features smaller than the full set, much computation time can be saved.

However, perhaps a more compelling reason is that one may well expect higher accuracy if a smaller set of features are used. This is due to what is known as the Hughes effect[11]. It is easy to understand that if one has very precise knowledge of the class distribution functions, the classification accuracy should be a non-decreasing function of the number of features used. However, because there are only a finite number of training samples available, and indeed, that number is often not large, the precision to which the class distributions can be estimated is often limited. As a result, the accuracy of the analysis that can be achieved will usually increase for a while, as more features are used, then peak and begin to decline. A further practical complication is that it is usually not possible to determine before the fact just what the best feature dimensionality is for a given classification.

For these reasons, feature selection, (i.e., selecting e.g., the best 4 of the available 10 features) and feature extraction (i.e., calculating, usually via a linear transformation on the existing features, a new feature space which is more suited to the specific problem) capabilities are very important in the analysis process. Several such capabilities are implemented in MultiSpec. Two, in particular, are to be noted. Discriminant Analysis Feature Extraction (DAFE) is a method that is well known in the field of pattern recognition[12]. It provides an optimal transformation to a feature space of up to M – 1 dimensions, where M is the number of classes. The second, Decision Boundary Feature Extraction (DBFE), provides an optimal linear transformation to a space of arbitrary dimensionality and gives information as to how many dimensions need be used in any given problem. It is predicted by theory and usually found in practice that DAFE, which is a very fast computation, provides the best results when the number of features to achieve satisfactory classification is less than M, while DBFE, a more lengthy calculation, provides better results when the number of features needed is greater than M and there are an adequate number of training samples. MultiSpec also implements the principal components transformation, which may be useful at low dimensionality, but which becomes undesirable at moderate or high dimensionality, since it tends to de-emphasize narrow band features which may be diagnostic of classes, often one of the very reasons for gathering high dimensional data.

The final characteristic of an analysis algorithm that needs to be mentioned at this point is that of its effectiveness at generalization. That is, given that it can accurately classify its training samples, how well can it classify other members of the same classes in the data set? This is closely related to how well the set of classes defined by the training samples result in equation (1) above being an equality. An algorithm in MultiSpec called “Statistics Enhancement” is directed at improving the fit of the classes

---

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

12 Fukunaga, K. *Introduction to Statistical Pattern Recognition,* Academic Press, 1990.

as indicated by equation (1). The algorithm is more fully described in the reference[13], however, basically it is a technique for using a combination of the training statistics and a portion of all other samples (unlabeled) in the data set in an iterative fashion to improve the degree of equality of equations (1).

Typical Analysis Steps using MultiSpec

The steps for accomplishing the analysis task according to the discussed approach using MultiSpec are:

1. Familiarization with the data set.

   • Display the data using the Display Image processor. Compare the displayed image with any ground reference information about the site that may be available. Compose a tentative list of classes that is adequately (but not excessively) exhaustive for this data set.

2. Preliminary selection of the classes and their training sets.

   • Using the Cluster processor, cluster the area from which training fields are to be selected, saving the results to disk file. Display the resulting thematic map for use in marking training areas.

   • Using either the display of the original data or that of the thematic cluster map, make a preliminary selection of training fields that adequately represent the selected classes.

   • Use the Feature Selection Processor to determine how separable the tentative classes are. It may then be necessary to iterate between the previous steps.

   • Cluster the training fields to check the modality. This will also be useful in identifying the need for subclasses.

3. Determination of the spectral features to be used.

   • Use the Feature Selection processor to choose the best subset of features for carrying out the classification for a given training set. The Feature Extraction processor is also available for this task.

4. Preliminary Classification of the data.

   • Classify the training fields only, using the spectral bands or features selected to verify their purity and separability.

5. Classification, Evaluation of the classification and Extraction of the desired information.

---

[13] Behzad M. Shahshahani and David A. Landgrebe, “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, September 1994.

- Review the results of training field classification. Make modifications to the training as required to obtain satisfactory results at this point. Mark as many fields as possible as Test Fields using the Statistics Processor. Classify these fields and examine the accuracies obtained to determine how well the classifier training generalizes beyond the training set.

- Depending upon the results of these evaluations it may be necessary to repeat previous steps after modifying the class definitions and training. After becoming satisfied with the results, classify the entire data set, perhaps setting a modest threshold value, saving the classification results to a disk file, and creating a Probability Results file.

- Use the Display Image processor to generate thematic map versions of the results and the probability result files for subjective evaluation purposes. The results file display is useful in determining that the classification results are appropriate and consistent from a spatial distribution standpoint. The portion of points thresholded in the results display, together with the probability map helps to determine if any important modes in the data have been missed in the class definition process. Depending on the outcome, it may again be necessary to iterate using some of the previous steps. The List Results processor can be used to provide a quantitative evaluation of the results based upon the accuracy figures of the training and test fields classification.

- From the Probability Results file displayed as a thematic map, determine if there are areas with very low likelihood of membership to any of the existing classes. For any that are found, create appropriate new classes. Also, determine how well the classifier generalizes to areas outside the training sets. Use the Enhance Statistics processor to improve the fit of the composite class statistics to the data set and reclassify.

Again, substantial variance from these steps might be appropriate, depending upon circumstances; however, these steps provide a reasonable overall picture of how to proceed.

An Example Analysis

The specific procedure to be used in analyzing a data set is controlled by what information one has about the scene before the fact and what specific information one desires as a result of the analysis. For example,

- Urban Land Use Example. One might desire to obtain a map of an urban area which delineates areas of the city that are in such land use classes as high density housing, low density housing, commercial, industrial, etc. One might expect to know, before the fact, where in the city, examples of each of these classes are found.

- Geology Example. One might have example spectra of specific minerals that are expected to be expressed on the surface and one might wish to

determine the general layout and shape of the expression of these minerals for purposes of estimating the manner, in a geologic sense, in which they came to be distributed as they are.

- Soil or Vegetation Example. One might have a general knowledge of the location of various soil categories and plant species in the scene and from that one might wish to determine tabular information giving the portion of the scene each occupies.

- Soil or Vegetation Example. Again from general knowledge of the location of various soil and plant classes in the scene, one might desire a thematic map showing not only how much of each is in the scene, but also where each is located.

Note that the goal in the above cases varies somewhat. If the results are to be in the form of a thematic map, then accuracy of the classification becomes of greater importance. On the other hand, if only tabular information is required, then accuracy decreases in importance in favor of obtaining unbiased results with regard to the classes, i.e. it is not as important that each pixel be correctly classified as it is that the proper number of pixels be assigned to each class.

We shall illustrate an analysis of a data set of the fourth case indicated above. In order to simplify the logistics of making this example analysis available to others, only a small portion of a data set was chosen for this example. It contains 145 lines by 145 pixels (21,025 pixels) and 9 spectral bands selected from a June 1992 AVIRIS data set[14] of a mixed agriculture/forestry landscape in the Indian Pine Test Site in Northwestern Indiana. The data set is designated 92AV3C9. Each of the nine bands are approximately 10 nm wide centered at the following wavelengths.

Band No.	AVIRIS Band No.	Wavelength center, μm
1	8	0.4795
2	16	0.5584
3	27	0.6675
4	39	0.7560
5	46	0.8235
6	70	1.0550
7	86	1.2092
8	136	1.6589
9	186	2.2186

In carrying out the analysis of the data set, we will generally follow through the numbered steps above, although, due to the relatively small size of the data set and the small number of spectral bands, all of the steps may not be needed. We will indicate specific choices of MultiSpec options and parameter values in **boldface** type and,

- Ancillary comments not in the mainstream of this processing task but which may be relevant in other circumstances will be indicated in indented, smaller type.

[14] Only 9 bands were selected from the AVIRIS data set to provide an analysis problem of conventional dimensionality. An analysis using the full 210 bands of the AVIRIS data set, constituting a true hyperspectral analysis problem, is given in the following section.

1. Familiarization with the data set and related information. Begin by choosing **Open Image** from the **File** menu and use the default options to display the data in simulated color infrared image form.[15] The result is shown in Figure 7.

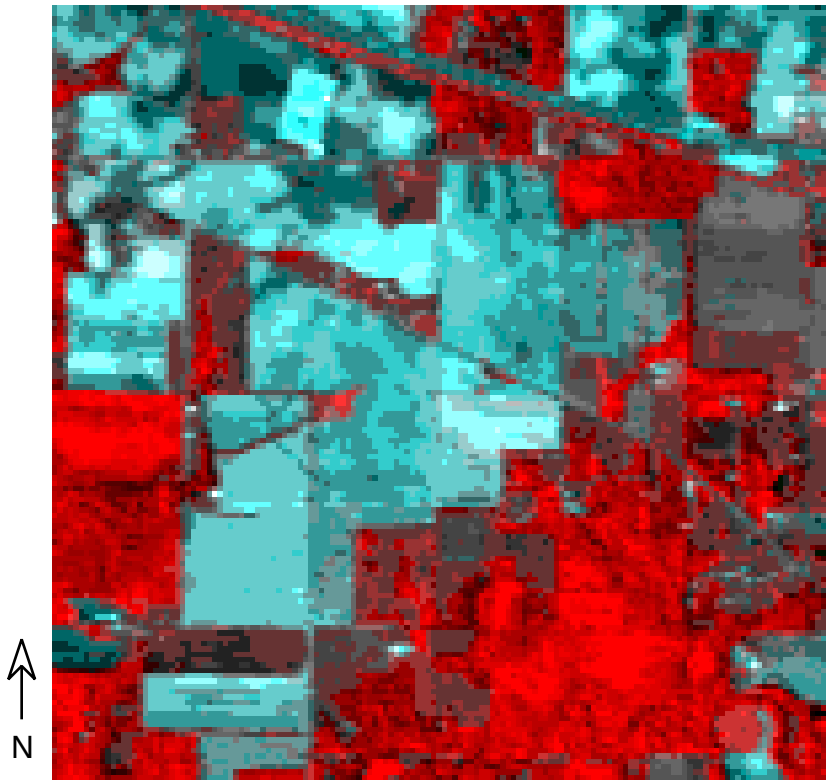


Figure 7. Data Set 92AV3C9 displayed in simulated color infrared form[16].

It is seen that the area appears to be about $^{2}/_{3}$ agriculture and $^{1}/_{3}$ forest or other natural perennial vegetation. However, due to the early season date of data collection, the cultivated land appears to have very little canopy cover as yet. There is a major dual lane highway (U.S. 52 & U.S. 231) and a rail line crossing near the top and a major secondary road (Jackson Highway) near the middle, both in a NW-SE direction. Several other county roads are also somewhat apparent, and show more clearly on a USGS quadrangle map, (Figure 8), as does the terrain relief of the area.

15 Most of the figures that follow are more effective if viewed in color.

16 The original for this and several subsequent figures are in color. Black & White hard copy versions will thus be difficult to interpret. If possible, view these figures in soft copy form on a color screen, so that the figures will be more interpretable.

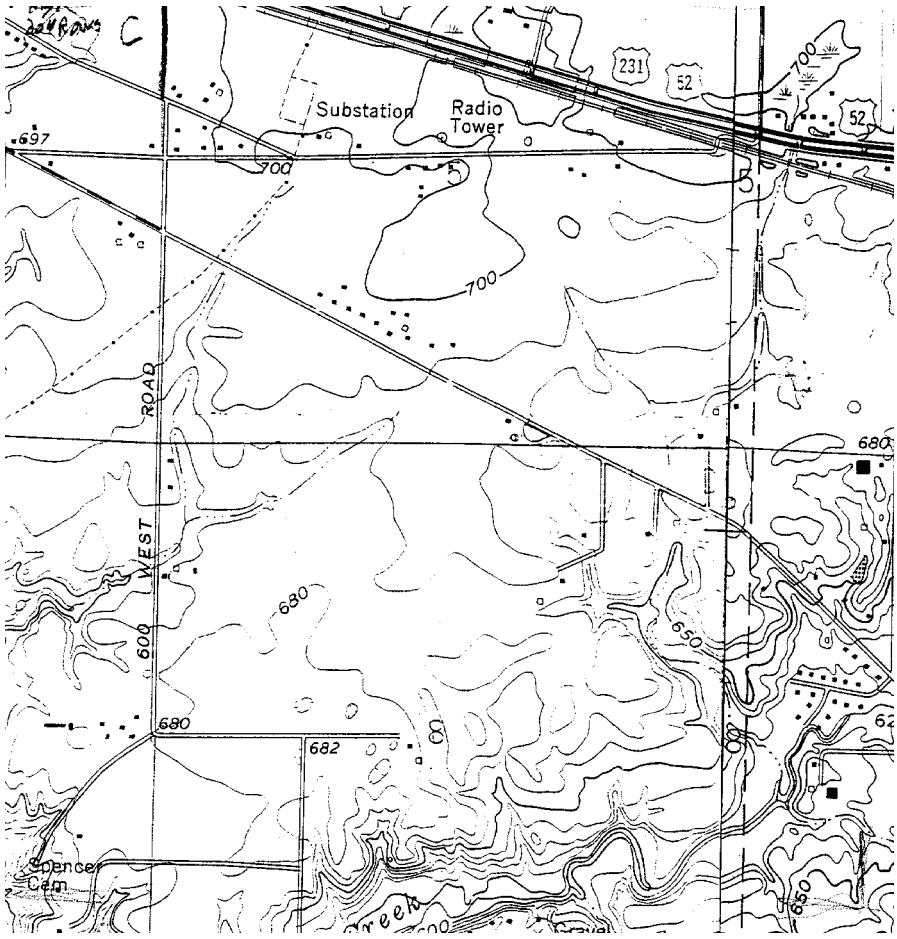


Figure 8. Portion of a USGS Quadrangle map of the test area.

The small rectangular dots on the quadrangle map indicate houses or other buildings existing at the time the quadrangle map was made. Thus a number of low-density housing areas can be found in the data. Additional information useful in deriving training samples for the ground cover of the area is contained on the following generalized reconnaissance map.

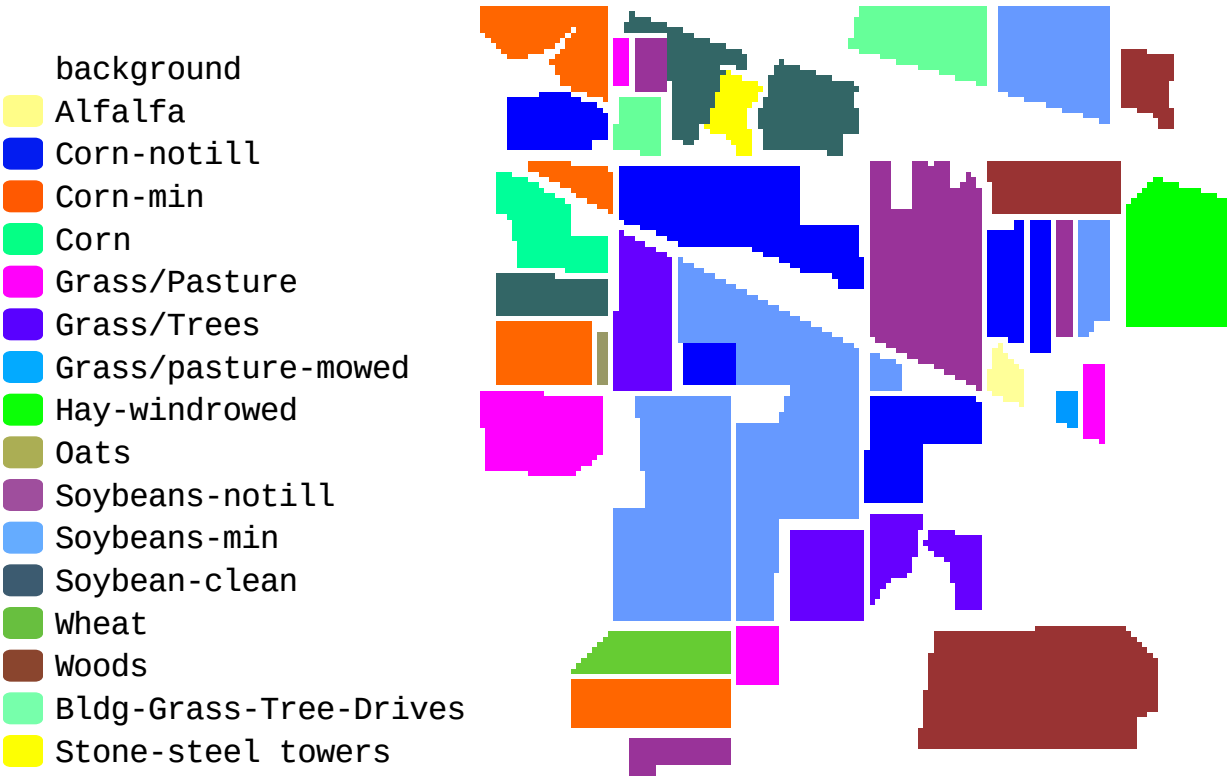


Figure 9. A generalized reconnaissance map of the test area, designated file 92AV3gt.GIS (in color).

A comparison of this reconnaissance map with the data in image form (Figure 7) shows that the reconnaissance map is highly generalized. Many small variations within fields that can be seen in the image data are ignored in the reconnaissance map. Thus, not every pixel inside of a designated area on the reconnaissance map should be expected to belong to the class associated with that area. Further, the map tends to give land use classes, rather than land cover classes. Thus, an area marked "Corn-notill" on the map may really have a land cover of bare soil and residue from previous vegetation, with only a small percent of corn vegetation as land cover.

- In the case of any given data set, various kinds of additional ancillary information might be available. Examples include soil maps, imagery from other sensors, airphotos, etc. Assembling this kind of information is an important part of a data analysis process, but obviously varies substantially for each analysis situation.

2. Preliminary selection of classes and their training sets. Next we consider how detailed a thematic map it might be both possible and desirable to develop. For this purpose, it is useful to consider possible classes in a hierarchical form. Besides the cultural features, there are several species of annual and perennial vegetation. Further subdivision possible within the annual vegetation category would be corn, oats, soybeans, and wheat. Still further subdivision within these might be to subdivide the corn and soybean classes by tillage practice, since the amount of residue from previous vegetation varies in this case. Within the category of perennial vegetation, classes such as alfalfa/hay, grass, and trees/woods would be possible categories. Due to the early season date of data collection, several species of annual vegetation have

so limited a canopy (< 5%) that these areas may not be separable in terms of plant species. Rather, the soil type or the tillage practice as manifested by the amount of surface residue from previous vegetation may be more appropriate classes.

All of the previous information will be useful in selecting classes and training sets. A very useful additional display is a map in which the scene has been subdivided into objects, i.e., areas of contiguous pixels that are spectrally similar to one another but spectrally separable from one another. Such a presentation can be obtained by selecting **Cluster** from the **Processor** menu. It is useful to adjust the parameters of the clustering so that one achieves roughly the same number of clusters as the number of classes desired, about 15 in this case. Choose **Isodata** with **one pass for initial centers**. After a little experimentation, the following parameter values would be found to lead to about 15 clusters.

Min. initial cluster size of 50
First line distance of 500
Final distance of 1000
Threshold level of 256
Use All 9 bands.
Classify the Image Area
Save to Disk File

Use the default values for all other parameters. The result of the clustering may be displayed using the **Open Image** selection of the **File** menu, as follows.

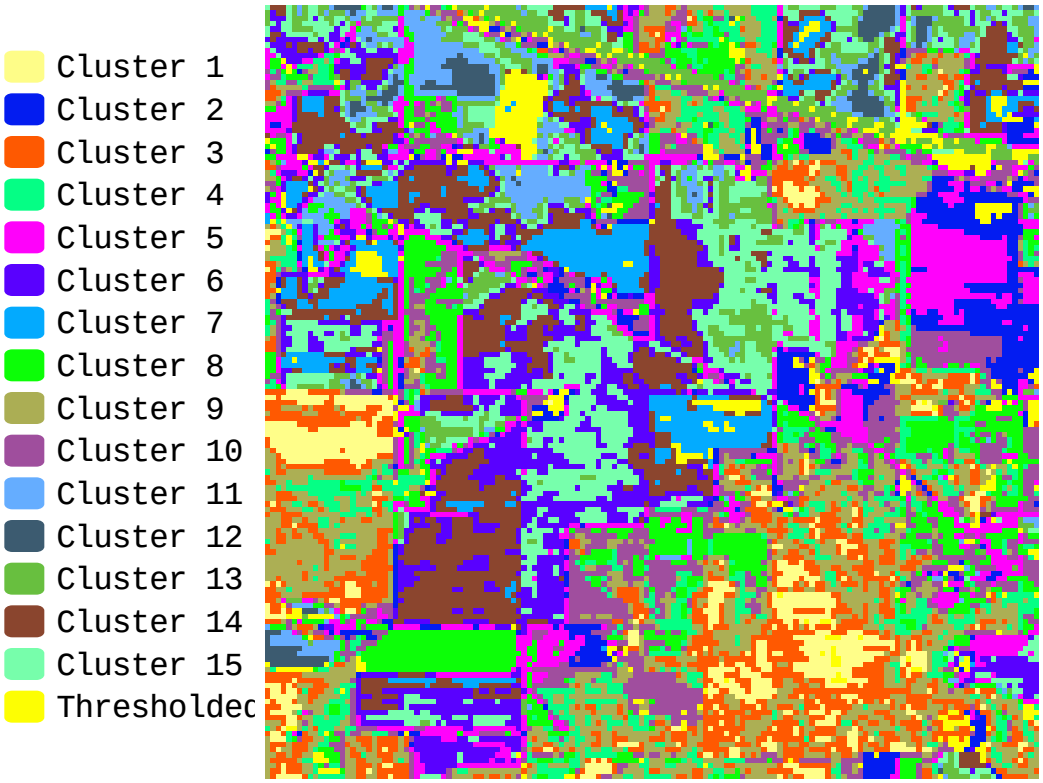


Figure 10. Cluster Classification Map (in color).

It is seen that, in most cases, even the agricultural fields are not uniform in their association with a single cluster, but are significantly mottled in appearance. This characteristic can be mitigated somewhat and a clearer association with individual clusters obtained by use of the ECHO spectral/spatial classifier using the cluster statistics. Thus select **Classify** from the **Processor** menu and choose the **ECHO** procedure, classify the **Image Area**, and save the results to **Disk File.** Continue to use all bands since their number is not large. The result will be as follows.

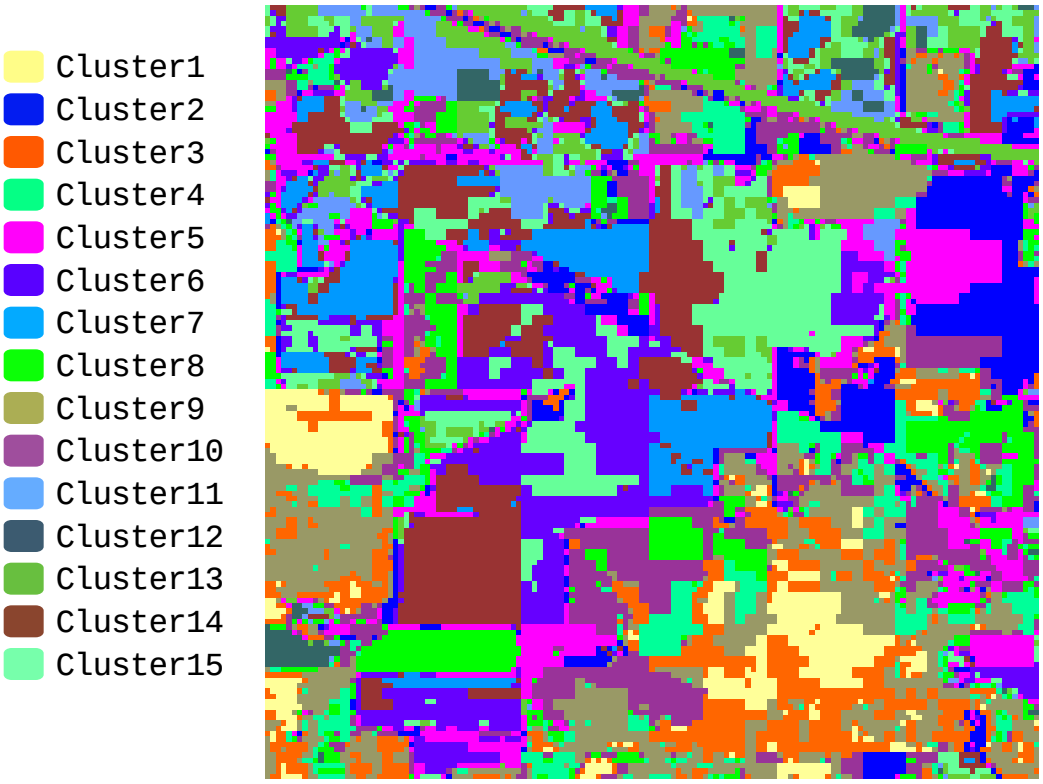


Figure 11. ECHO classification of Cluster Statistics (in color).

It is seen that this display clarifies the association of areas with spectral characteristics to some degree, although most areas are still a mixture of more than one cluster in spectral space.

- Note that this two-step procedure, clustering followed by ECHO classification, provides an effective unsupervised multivariate image partitioning capability.

  An alternate procedure might be to mark in the data the desired initial cluster centers, then have these used in the clustering process, thus "guiding" the clustering toward clusters which might turn out to be closer to the classes desired.

By now referring to the image of the data (Figure 7), the ECHO classification (Figure 11) and the generalized reconnaissance map (Figure 9), one can make a first draft at selecting classes and training fields. A possible choice will be illustrated. At this point we are assuming that all of the classes of the reconnaissance map will be separable and thus training samples have been established for each. We show in Figures 12 and 13 the training fields outlined on both the ECHO classification map and the color

infrared image, because the process is one of selecting classes and their training fields that satisfy both criteria, i.e., classes that are separable and of informational value. The ECHO map shows classes which are separable, but classes must be combined in such a way that they are also of informational value.

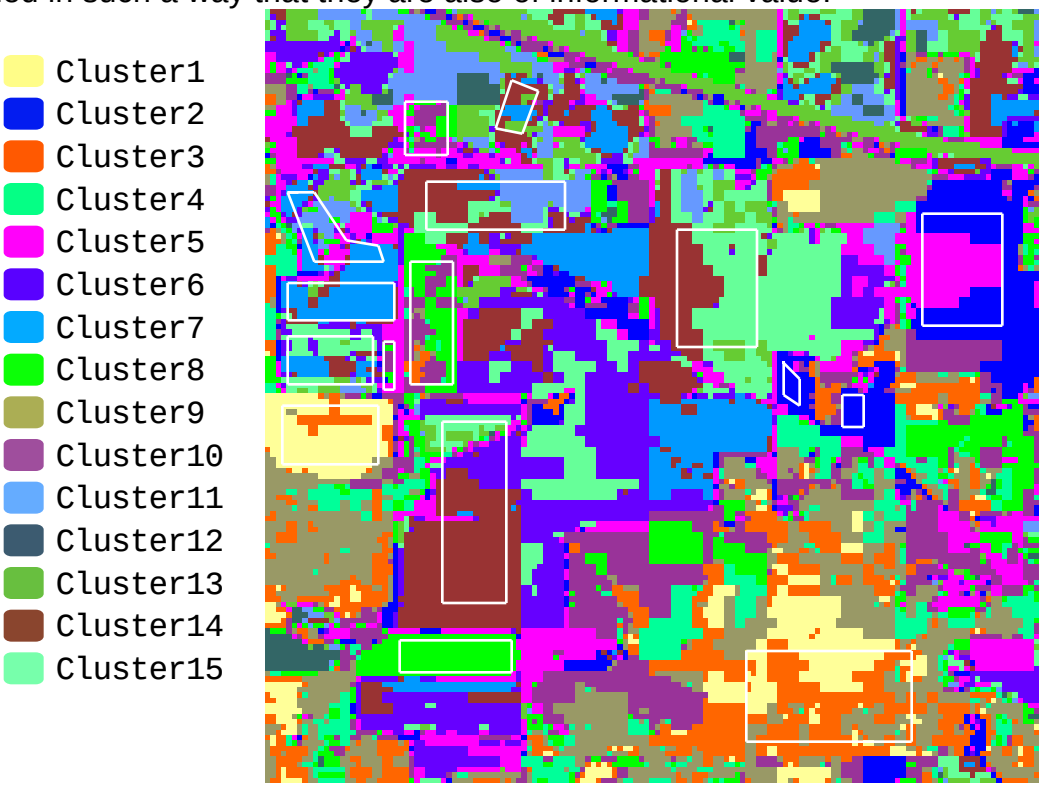


Figure 12. Unsupervised Scene Segmentation showing the training fields selected (in color).

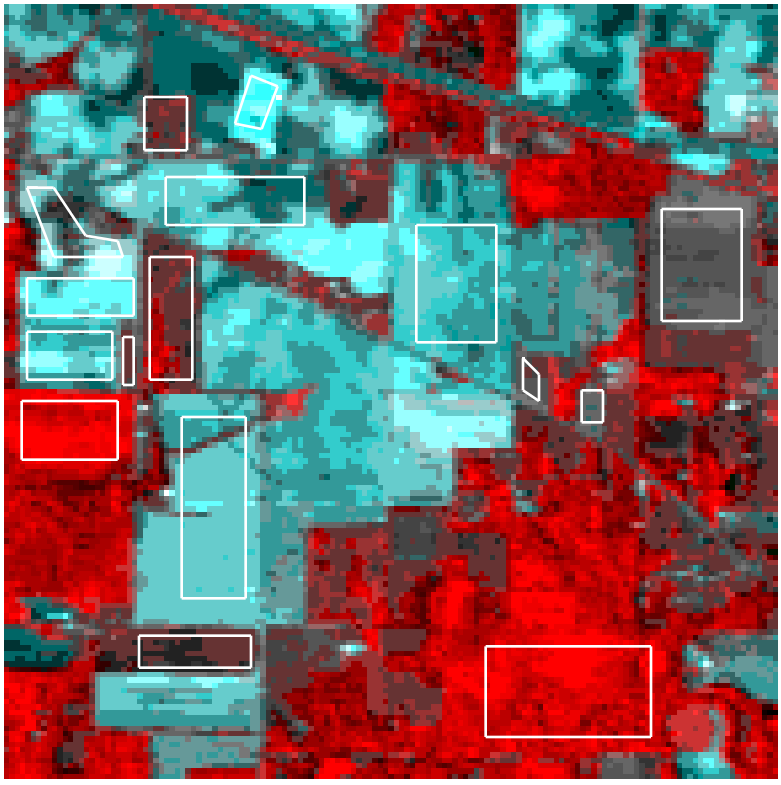


Figure 13. Data displays showing training fields (in color).

In order to allow for precise duplication of the results that follow, the addresses of the training fields selected are provided in Table 1. Note that only one field has been selected for each class. This will likely not be adequate, since one sample from a single area is usually not adequate to characterize a class in the entire data set, however, it is deemed adequate for a first trial.

Class	Line	Column	No. of Samples
alfalfa	67 73 75 70 67	98 98 101 101 98	18
Corn-notill	33 to 41	31 to 56	234
Corn-min	63 to 71	6 to 21	144
Corn	35 35 48 48 45 44 35 35	7 5 10 23 22 16 10 5	108
Grass/pasture	75 to 85	4 to 21	198
Grass/trees	48 to 70	28 to 35	184
Grass/pasture-mowed	73 to 78	109 to 112	24
Hay-windrowed	39 to 59	124 to 138	315
Oats	63 to 71	23 to 24	18
Soy-notill	42 to 63	78 to 92	330
Soy-min till	78 to 111	34 to 45	408
Soy-clean	52 to 58	5 to 24	140
Wheat	119 to 124	26 to 46	126
Woods	121 to 137	91 to 121	527
Bldg-grass-trees-drives	18 to 27	27 to 34	80
Stone-steel towers	14 23 24 16	47 44 49 52	45

Table 1. Coordinates for training fields. Those for Alfalfa, Corn, and Stone-steel towers were delineated in polygon form. All others are rectangular.

- An alternate approach to defining training statistics is to set the parameters of the initial clustering so as to provide two or three times as many clusters as the number of final classes desired, then to define classes by studying how to combine two or more clusters into a desired information class. In this case, the maximum likelihood or ECHO classification would be done using the statistics of all clusters, and the final classes achieved by grouping classes together after classification.

Next choose **Feature Selection** from the **Processor** menu. Under **Combination Groups**, choose **9**. Use the default options for all other parameters. This will list the Bhattacharyya distance between all class pairs to indicate the relative separability of each class pair. There are 120 possible such class pairs in this case. The results are given in terms of the following symbols:

Classes used:	Symbol
1: Alfalfa	1
2: Corn-notill	2
3: Corn-min	3
4: Corn	4
5: Grass/pasture	5
6: Grass/trees	6
7: Grass/pasture-mowed	7
8: Hay-windrowed	8
9: Oats	9
10: Soy-notill	A
11: Soy-mintill	B
12: Soy-clean	C
13: Wheat	D
14: Woods	E
15: Bldg-grass-trees-drives	F
16: Stone-steel towers	G

Table 2. Classes and their Symbols used in Table 3.

The class pairs and their corresponding B-distance are given in Table 3 in order of ascending B-distance size. The sorting was done by copying the row of class pair symbols and the row of their corresponding B-distances from the MultiSpec text window to a spreadsheet and sorting them there.

3C	AB	4B	6F	3B	2B	2A	AC	18	24	DF	BC	3A	78	69	34	9F	5E
1.48	2.8	2.91	2.94	2.95	3.02	3.1	3.41	3.45	3.74	4.05	4.08	4.33	4.37	4.62	4.64	5.3	5.4
23	2C	46	17	7F	EF	4A	6E	9D	4G	4C	AG	8F	BF	5F	6D	CG	49
5.64	6.56	6.63	6.79	6.82	6.98	7.63	7.95	8.15	8.17	8.74	9.31	9.58	9.61	9.73	10	10	10.4
48	1F	2G	6B	BG	4F	5D	68	7B	BD	DE	28	9B	3G	1B	67	56	6A
10.9	11	11.2	11.2	11.2	11.7	12.1	12.1	12.9	12.9	14	14.9	15	15.1	15.2	15.5	15.9	16.4
26	FG	9E	47	8B	79	59	16	BE	AF	6C	3F	4D	5B	14	2F	7D	8C
17.1	20	20.9	21.2	21.8	23.5	24.5	25.3	26.1	26.8	27.6	28.5	28.7	28.7	29.3	29.5	31.8	31.9
CF	8A	7A	89	27	36	19	29	8D	2D	7C	3D	AD	1D	38	4E	CD	9C
31.9	32	34.2	35.1	35.7	36.1	38.6	38.9	40.8	41.8	41.9	42.8	46.2	46.6	47.4	47.6	48.1	48.8
6G	39	37	12	9A	1C	8E	3E	57	58	8G	7E	1A	13	DG	7G	9G	45
55.3	57.3	58.1	61.7	64.5	66	68.6	70.7	72.8	74	75.1	76.7	77	79.6	82.4	86	93	94
5G	EG	35	2E	1E	CE	5C	1G	25	15	AE	5A						
99.9	102	112	121	135	149	156	162	162	170	171	194						

Table 3. Bhattacharyya interclass distances in order of ascending distance.

It is seen that some of the distances are quite small, implying that they may be only marginally separable. Those class pairs less than 3 indicate that it may not be possible with only the 9 bands available to separate all classes of corn and soybeans from each other, or if so, the separations may not be particularly robust in generalization to a larger area. It appears only a little more likely that it will be possible to separate tillage classes.

3. Determination of the spectral features to be used. Under some circumstances it would be appropriate to use the Feature Selection processor at this point to select a subset of features to use for initial classification. This might be called for in cases where the number of available spectral bands was large or the number of training

samples available for one or more classes was small (e.g., ≤ the number of spectral bands). However, in the present case, since neither of these conditions is present, we shall proceed directly to a preliminary classification.

4. Preliminary Classification of the data. Choose **Classify** under the **Processor** menu and use the default options to classify the training fields. This will help to verify their purity and the separability of at least the training fields. The results in terms of alphanumeric field maps and tables of accuracy appear in the MultiSpec text window. Inspection of these results indicate good accuracy (98.2% overall) with no fields showing unreasonable or unexplainable errors.

5. Classification, Evaluation of the Classification and Extraction of the desired information. Choose **Classify** from the **Processor** menu and select **Image File** for classification. Also save the results to **Disc File** and deselect **Text Window**. This will result in a maximum likelihood pixel classification of the data set that can be displayed in thematic map form, as in Figure 14.

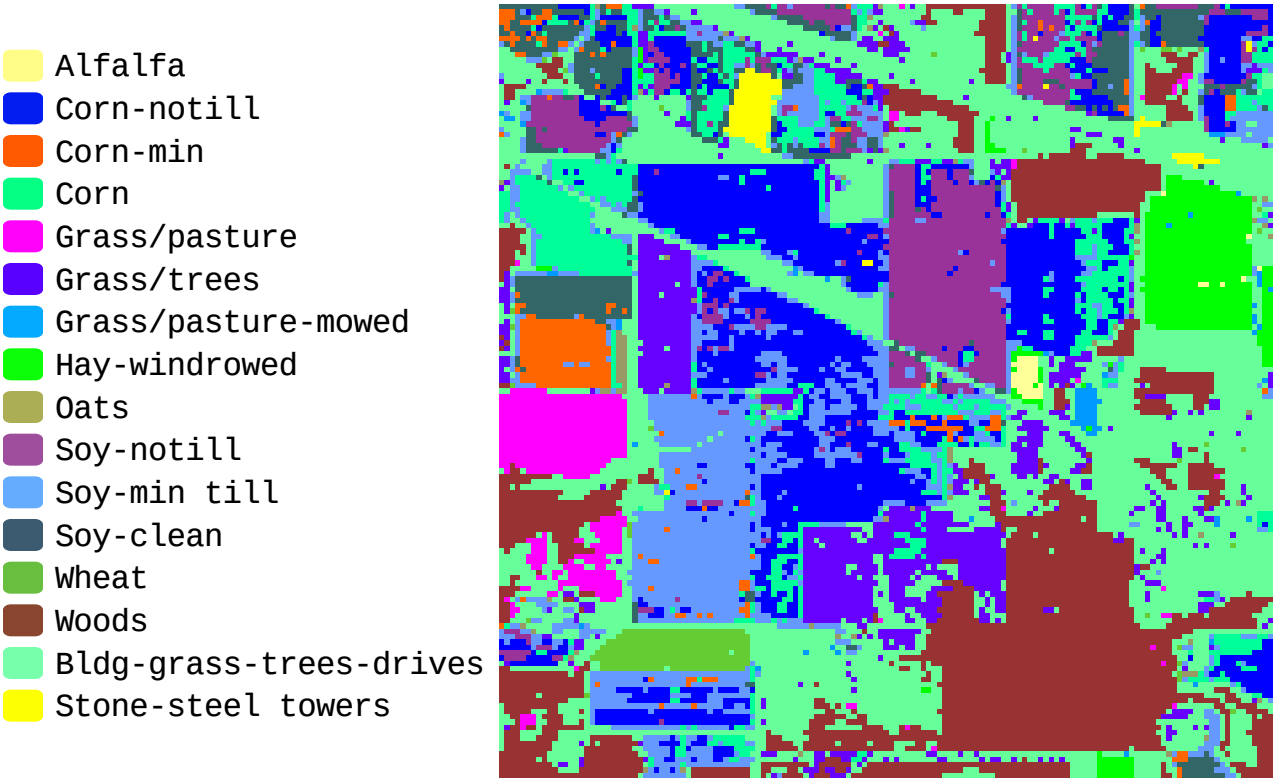


Figure 14. Maximum Likelihood Classification of original training statistics (in color).

A superficial evaluation of this result reveals that it shows some promise, but needs improvement. For example, a large portion of the large field at the center has been classified as Corn-notil when in fact it should be Soy-min till. A Corn-min till field near the lower left corner is classified partly as Soy-min till and partly as corn-notill. The improvement process can be begun by adding training fields from these three areas, as follows:

Class	Lines	Columns
Soy-min till	3 to 17	102 to 117
Soy-min till	80 to 95	51 to 71
Corn-min till	128 to 134	20 to 46

A new classification, this time using the ECHO classifier, but otherwise with default options results in significant improvement, as can be seen in Figure 15.

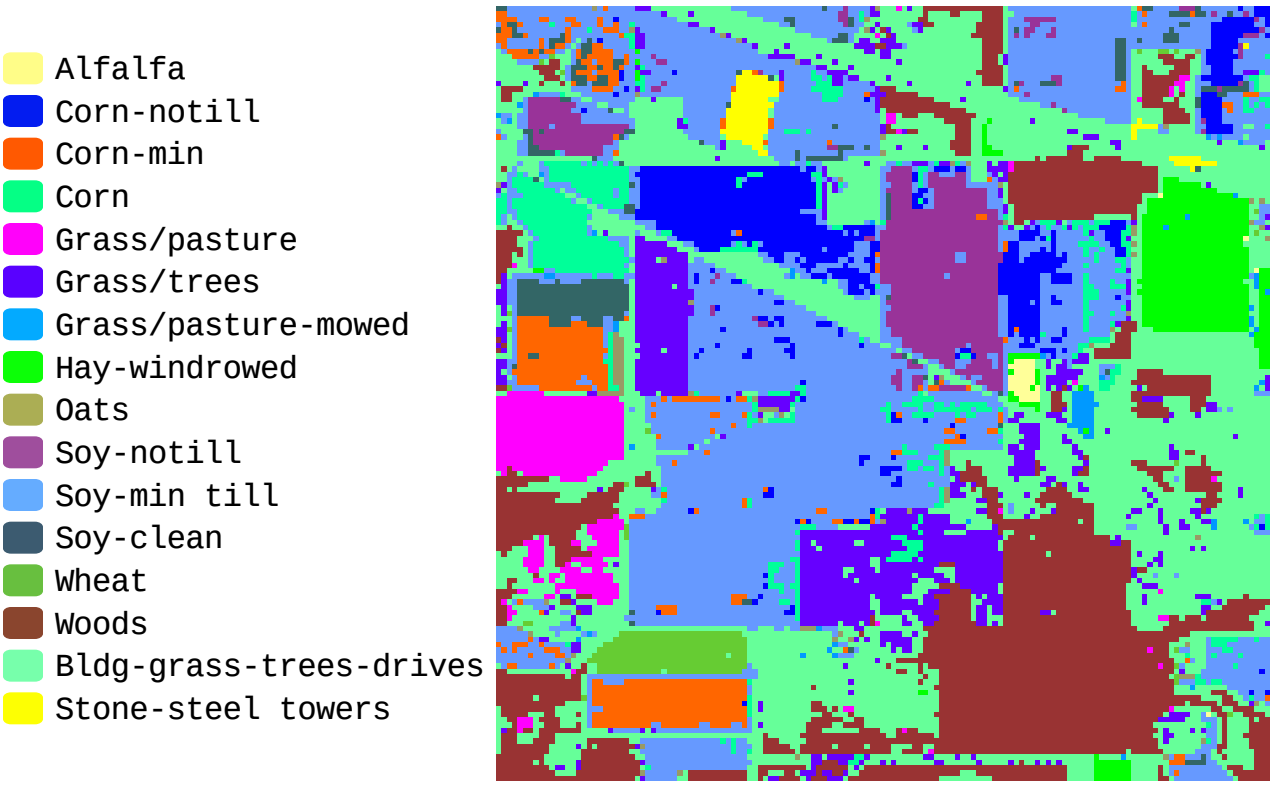


Figure 15. ECHO classification with 3 new training fields added (in color).

In order to obtain a more quantitative estimate of the accuracy of the results, the fields defined by the generalized reconnaissance map can be used as test fields. To do so, **Close** the current **Project**, being careful to save the project for future reference, and **Open** the **Project**, 92AV3.Project, containing the test field boundaries. Then, with the above ECHO classification map active, choose **List Results** from the **Processor** menu, and request a **Class Type Map**, and **Summarize by Class**. The result is as follows.

Project			Number of Samples in Thematic Image Class																
Class	Class	Percent	Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1
Name	No.	Correct	Samples	Alfalf	Corn-n	Corn-m	Corn	Grass/	Grass/	Grass/	Hay-wi	Oats	Soy-no	Soy-mi	Soy-cl	Wheat	Woods	Bldg-g	St
Alfalfa	1	66.7	54	36	0	0	0	0	0	0	11	0	0	7	0	0	0	0	
Corn-notill	2	49.6	1434	0	711	19	70	0	2	0	0	0	119	493	11	0	0	9	
Corn-min	3	61.6	834	0	6	514	80	0	0	0	0	1	0	183	49	0	0	1	
Corn	4	91	234	0	0	0	213	0	0	0	0	0	0	16	1	0	0	4	
·ass/Pasture	5	64.8	497	0	0	0	12	322	4	2	4	0	0	19	1	0	1	132	
;rass/Trees	6	90.2	747	0	0	0	25	0	674	0	0	1	0	1	0	0	1	45	
·ass/pas-mo	7	100	26	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	
ıy-windrowed	8	98.6	489	1	0	0	0	0	0	2	482	0	0	0	0	0	0	4	
Oats	9	95	20	0	0	0	0	0	0	0	0	19	0	0	0	0	0	1	
ıybeans-notill	10	68.5	968	0	10	3	32	0	2	0	0	1	663	249	6	0	0	2	
ɔybeans-min	11	90.6	2468	0	65	49	40	0	2	0	0	0	40	2236	23	0	0	13	
ıybean-clean	12	22.5	614	0	1	5	24	0	5	2	0	1	2	433	138	0	0	3	
Wheat	13	96.7	212	0	0	0	0	0	0	0	0	0	0	1	0	205	0	6	
Woods	14	94.9	1294	0	0	0	0	13	4	0	0	0	0	0	0	2	1228	47	
g-Grass-Tree-	15	70.5	380	0	0	0	0	2	21	0	0	1	0	4	1	0	83	268	
ıne-steel towe	16	93.7	95	0	0	2	0	0	0	0	0	0	0	0	0	0	0	4	8
TOTAL			10366	37	793	592	496	337	714	32	497	24	824	3642	230	207	1313	539	

OVERALL PERFORMANCE(7824/10366)=75.5

Table 4. Quantitative classification summary for the fields of Figure 9.

The overall accuracy appears to be 75.5%, however, note that many of the “errors” are not significant. For example, class 5, Grass/Trees had many of the test samples assigned to the Building-Grass-Trees-Drives, which, depending on the specific objective of the analysis, may not represent errors at all. Indeed, if the objective of the analysis was directed primarily at soils, tillage, or crop classification, the classes 5, 6, 7, and 15 could be combined into a single group, substantially raising the calculated accuracy.

Similarly, if the interest is primarily in crop classification, classes 2, 3, and 4 could be combined, as could classes 10, 11, and 12. On the other hand, if erosion related to tillage practice is the primary interest, classes 2 and 10, 3 and 11, and 4 and 12 could be combined, again increasing the calculated accuracy.

There are several additional possibilities that could be investigated as well. These include the following.

- As previously indicated, defining a class by contiguous pixels from a single area is usually not adequate, as it is unlikely that pixels from a single area show all of the characteristic variability that will exist over a whole scene. Thus additional training areas over those already added would no doubt improve the characterization of the classes.

- The extent to which the list of classes is exhaustive relative to the data. If a **Probability Map** had been requested and **Saved to Disk** at the initial classification, the following map (Figure 16) could have been viewed. This is a map showing in color-coded fashion the degree of membership of each

pixel to the class to which it has been assigned. The dark blue areas of the map indicate low probability of class membership for the class to which those pixels have been assigned, the yellow or light areas, high probability. The legend quantifies this probability in terms of a Gaussian distribution. In this case, this map points up several cases where additional classes might be established. For example, it would appear useful to define a class called Roads.

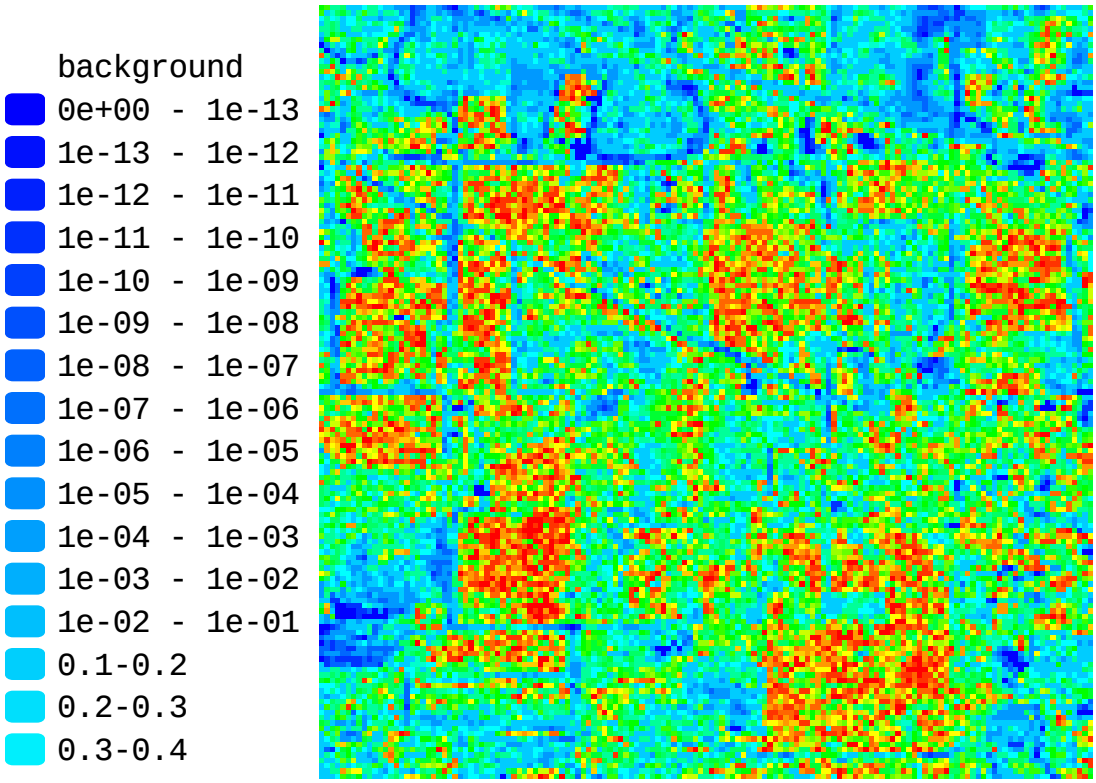


Figure 16. Probability Map from Initial Classification (in color).

- The use of feature selection or feature extraction. For example, one might choose **Feature Extraction** from the **Processor** menu and generate a new feature space via an optimal linear transformation of the spectral bands. Though a classification using all of the new features would give the same performance as the complete set in the old space, an appropriately sized subset of the new features might improve the performance.

- Once an adequately exhaustive set of classes have been appropriately defined use of the **Enhance Statistics** processor under the **Processor** menu should make the class statistics more robust in the sense of providing for better generalization to samples other than the training samples.

These are some of the additional possibilities. Obviously, there are many parameter settings, permutations and combinations of these that could be explored, too many to test here. However, in order to obtain some idea of how well this data set could be classified into the classes contained in the generalized reconnaissance map, we will

use all of the fields defined in that map as training fields for one final classification. The result is shown in Figure 17 and was obtained by using the Discriminant Analysis option of the Feature Extraction processor. The classification was done using the ECHO algorithm with the first 5 of the transformed features. Table 5 shows the accuracy summary based upon the generalized reconnaissance map fields and indicates an accuracy above 80%. As can be seen, there remains some confusion between soybeans and corn, and given the limited canopy development (≤ 5%) there appears little likelihood that this confusion can be eliminated using this data set.

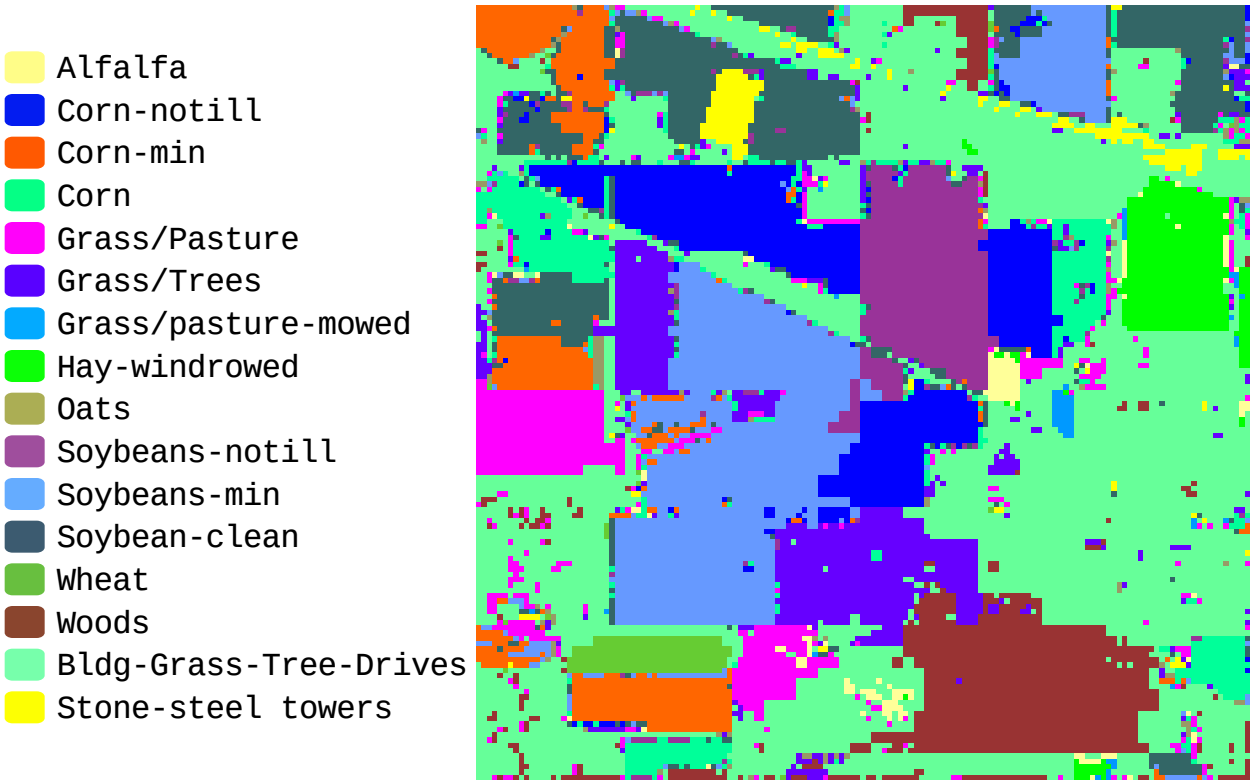


Figure 17. ECHO Classification of first 5 Decision Boundary Feature Extraction (DBFE) transformed features of recon. map training statistics (in color).

ject			Number of Samples in Thematic Image Class																
ss	Class	Percent	Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1
ne	No.	Correct	Samples	Alfalf	Corn-n	Corn-m	Corn	Grass/	Grass/	Grass/	Hay-wi	Oats	Soybea	Soybea	Soybea	Wheat	Woods	Bldg-G	St
alfa	1	83.3	54	45	0	0	0	1	0	0	3	0	5	0	0	0	0	0	
n-notill	2	78.9	1434	0	1132	56	16	2	0	0	0	1	31	82	110	0	0	1	
n-min	3	78.3	834	1	81	653	15	0	0	0	0	1	2	5	75	0	0	1	
n	4	91.5	234	1	3	1	214	1	5	0	0	1	1	1	3	0	0	3	
ss/Pasture	5	83.3	497	3	0	2	3	414	4	0	0	1	0	1	11	0	0	57	
ss/Trees	6	94	747	0	0	0	4	0	702	0	0	0	0	0	0	0	3	38	
ss/pas-mo	7	92.3	26	0	0	0	0	1	0	24	1	0	0	0	0	0	0	0	
-windrowed	8	95.7	489	10	0	0	0	0	0	6	468	0	0	0	0	0	0	5	
s	9	90	20	0	0	0	0	0	2	0	0	18	0	0	0	0	0	0	
beans-notill	10	74.9	968	0	7	1	153	0	6	0	0	1	725	1	73	0	0	1	
beans-min	11	81.2	2468	5	88	51	129	21	4	0	0	0	72	2003	91	0	0	4	
bean-clean	12	92.5	614	4	0	5	11	0	3	0	0	0	9	5	568	0	0	5	
eat	13	87.7	212	0	0	0	0	0	0	0	0	0	0	1	0	186	0	25	
ods	14	70.4	1294	0	0	0	0	2	1	0	0	0	0	0	0	0	911	380	
g-Grass-Tree-	15	73.9	380	0	0	1	3	1	3	0	0	0	0	0	1	1	88	281	
ne-steeltowe	16	98.9	95	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	9
TAL			10366	69	1311	770	548	443	730	30	472	23	845	2099	933	187	1002	801	1

ERALL PERFORMANCE(8438/10366) = 81.4

Table 5. Quantitative Classification summary based upon the fields of Figure 9.

Analysis of Hyperspectral Data

More Background and Concepts

For some years, the availability of multispectral data has been limited to data of relatively low dimensionality (< 10 bands). Recent advances in sensor technology, making possible data with several hundred bands, has provided a significant step function in the complexity of such data, and was motivated by the hypothesis that this new type of data, commonly referred to has hyperspectral data, provides the potential for gathering much more detailed information than previously possible, and doing so with greater accuracy. However, the significant increase in data complexity, quite aside from its increased volume, means that more than simply scaling up of computer speed and storage allocation will be required to realize the potential that this data provides. New analysis techniques will be required.

The techniques described in this paper and implemented in MultiSpec result from a substantial effort[17] to devise suitable means to analyze hyperspectral data. The techniques resulting (as well as those described above) are based upon principles having their origin in communication systems research of the last half dozen decades. Basically the perspective is that of modeling both the signal (variations diagnostic of the materials of current interest) and the noise (variations not currently of interest) in order to optimally discriminate among the expected list of signals. In the field of communication engineering, this approach has made possible the transmission of

17 D. A. Landgrebe and C. J. Johannsen, Principal Investigators, Final Report, NASA grant NAGW-925 "Earth observational research using multistage EOS-like data," Purdue University, May, 1994. A copy is available from http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/Final_Report5_94.html. See also 1997-98 annual report at http://dynamo.ecn.purdue.edu/~landgreb/ann.report.pdf.

information even when the signal-to-noise ratio is very much less than one, e.g., the transmission of scientific data from deep space probes using very low power transmitters.

The hypothesis followed here is that following these communication system principles in the case of multispectral remote sensing data analysis should make possible the extraction of information from data to a similarly unexpected degree. The advent of hyperspectral data makes practical a realistic test of this hypothesis. Using hyperspectral data from the site and time of the previous section but using all 220 spectral bands instead of the nine used previously, we will focus upon the six classes of soybeans and corn defined previously. These six classes represent two different plant species that have been grown under three different tillage practices, no-till, minimum-till, and clean till. Crops grown under no-till, minimum-till and clean till practices result in, respectively, a substantial amount, a moderate amount, or almost no residue from last season's crops being present on the surface. At the time of collection, the corn and soybean plants had only recently emerged and had achieved a canopy of only very limited ground cover (≈ 5%).

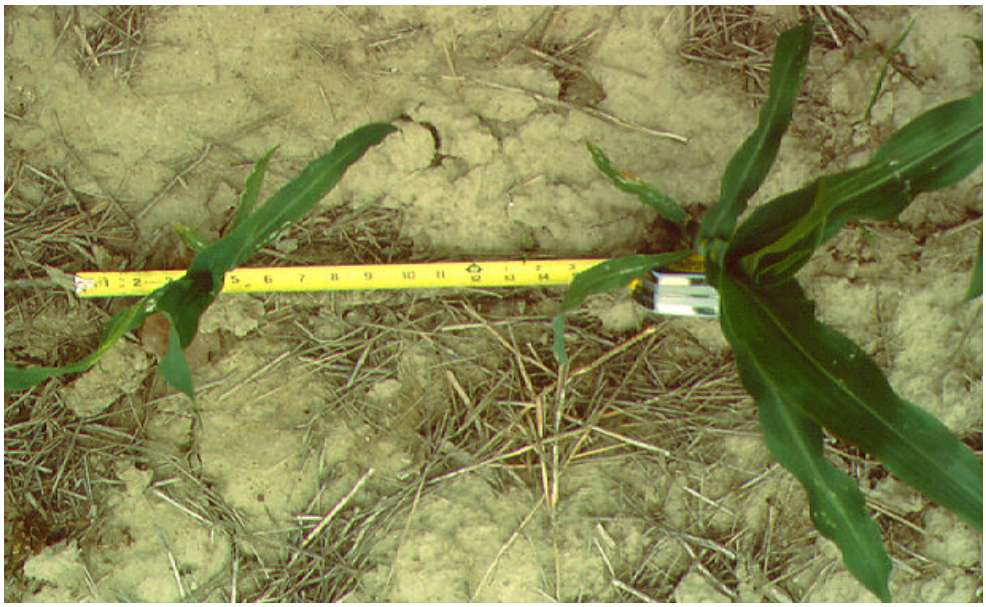


Figure 19. A close-up view of an area of the field marked 3-9 (Corn-no til) in Figure 9, showing the low percentage of canopy ground cover present at the time of data collection.

Figures 19 and 20 show examples of the degree of crop canopy cover in two example fields. Given this low canopy ground cover, the variation in spectral response due to (a) the soil type variations, and (b) the varying amount of residue from last season's crop, would provide a much greater influence upon the net pixel spectral response than would the vegetative species variation. In this sense, then, the problem of species identification represents a specific example of a generic low signal-to-noise information extraction problem. It is seen in the previous section that, even following sound practices, it was not really possible to discriminate between these two current plant species with high accuracy using the nine well-positioned spectral bands of the data of conventional dimensionality. In the following section we will lead the reader

through analysis steps that allow for discriminating between these two plant species to essentially arbitrarily high accuracy.

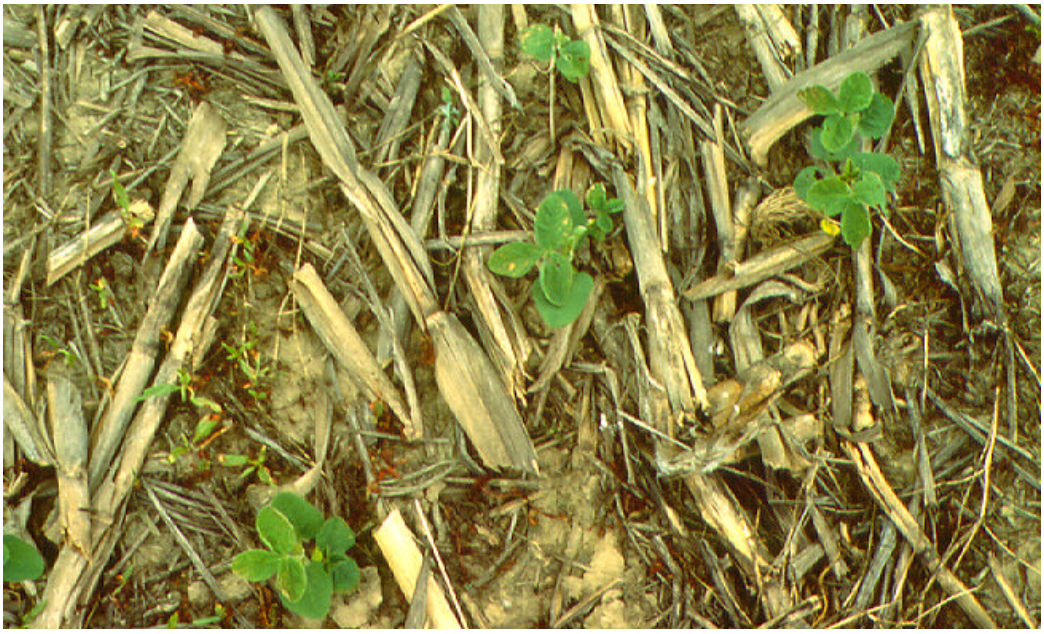


Figure 20. A close-up view of an area of the field marked 3-8 (Soybeans-min. til) in Figure 9, showing the low percentage of canopy ground cover present at the time of data collection.

In addressing the problem of high dimensional data analysis, it is important to recognize the differences that working in high dimensional space makes. One must understand as thoroughly as possible, based upon currently available fundamental knowledge about high dimensional space, what may be different about the information extraction characteristics of such spaces. This is an area in which the background fundamental knowledge is still not far advanced, but there are some characteristics that are known to be different than intuition might suggest. For example, here are two conjectures about geometrical properties whose validity would at first seem clear[18].

- Borsuk's Conjecture: If you break a stick in two, both pieces are shorter than the original.

- Keller's Conjecture: It is possible to use cubes (hypercubes) of equal size to fill an N-dimensional space, leaving no overlaps nor underlaps.

As it turns out, counter-examples to both have been found for higher dimensional spaces. Thus, one must be careful about using two or three-dimensional conceptual truths as a basis for conclusions in higher dimensional spaces.

Other aspects that are out of the ordinary are statements such as, "as the dimensionality increases, the amount of volume increases enormously fast, and most

18 Science, Vol. 259, 1 Jan 1993, pp. 26-27

of it is away from the origin[19].” Indeed, it is this enormous volume that provides the potential for discriminating between the spectral responses of many more materials, for there is volume available within which many more spectral responses can reside without overlapping others. However, even there it is the case that what seems obviously true is not quite as one would expect. As an example, consider the following graph of results[20].

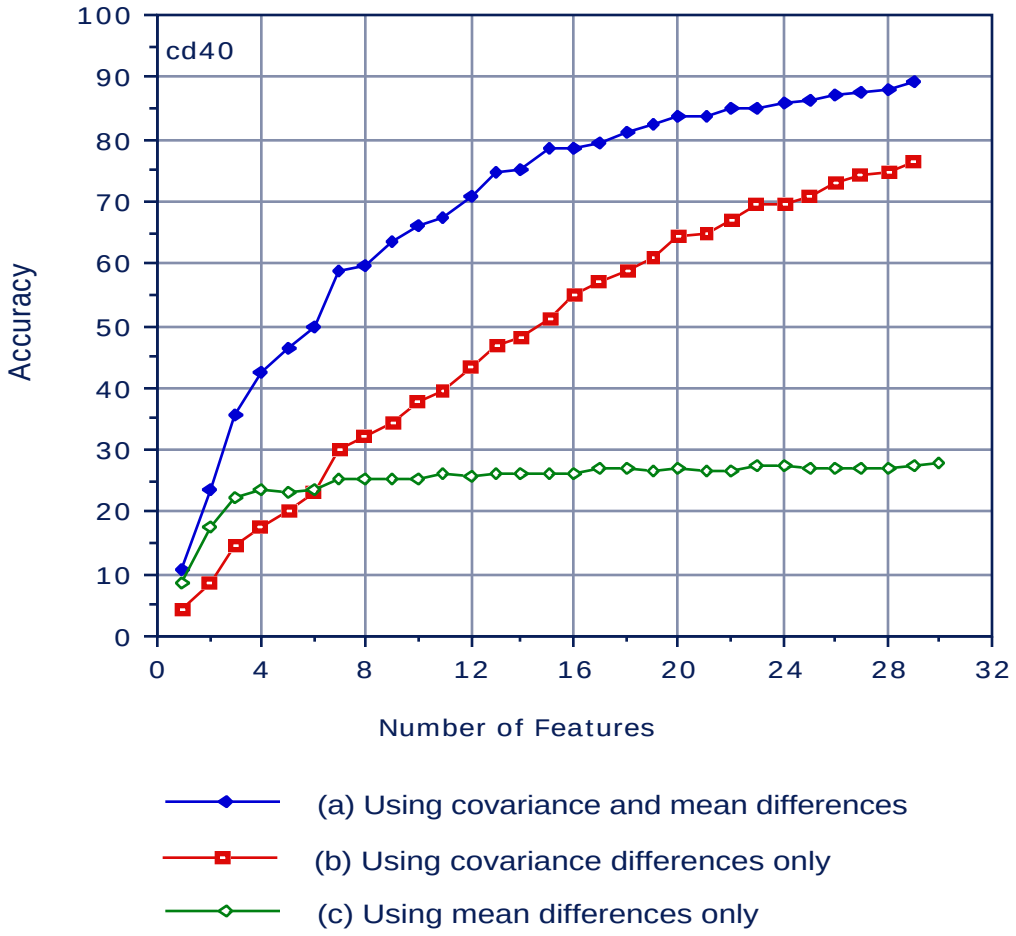


Figure 18. Classification accuracy using first order statistics only, second order statistics only, and both.

The top curve shows classification accuracy vs. dimensionality for a conventional maximum likelihood Gaussian classification. Such a classifier uses both first and second order statistics, both the mean vector and the correlation or covariance matrix. The class mean vector defines the location of the centroid of the class in N-dimensional space. The second order statistics define the shape and orientation of the class distribution. The lower two curves of Figure 18 show the performance using the mean vector information only and the covariance information only. Note that at low

19 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*, Volume 28 Part C Number 1, pp. 39-54, Feb. 1998.

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

dimensionality, N < 6, the mean-only classification provides higher accuracy than the covariance-only information, as one might expect. That is, the location of the distribution apparently contributes more to class separability than does the shape and orientation of the distribution. However, in this case, with increased dimensionality this performance soon saturates and improves no further. On the other hand, the covariance-only result continues to increase with dimensionality, indicating that for high dimensionality the shape and orientation are more significant to separation than is the location information. The implication of this is that, for hyperspectral data, one must pay close attention to the estimation of the second order statistics, in addition to the mean spectral response.

Some implications of this to the analysis of hyperspectral data are the following:

- Larger numbers of training samples will be needed to properly characterize classes than is the case with multispectral data of conventional dimensionality. In general, estimates of second order statistics converge more slowly to their final values than do first order statistics as the number of samples is increased.
- Care must be exercised that, preprocessing procedures which are intended to improve the first order characteristics do not inadvertently negatively impact the second order characteristics and thus, rather than improve, actually diminish the information content of the data. Examples of where this could occur are in attempts to “calibrate” the data or to “correct” for various observational or environmental effects.

Other differences arise for more straightforward reasons. For example, while the principal component transformation is quite commonly used with low dimensional data, the fact of now having a large number of spectral bands change the implications of its use. The principle component transformation basically focuses upon the signal variation *of the data set as a whole,* not of the separation between classes. It attempts to maximize the variation contained in the low order transformed components, relegating variations of less significant size to higher order components. Thus in 4-band data, a modest variation occurring in a single band is likely to have a significant effect, showing up in low-ordered principal components. However, the same sized variation occurring in one of 220 bands might not, and yet it could be just as diagnostic of a class of interest.

The matter of the best way to analyze hyperspectral data is not a mature technology. Indeed, it will be appropriate for this to be the subject of research for some time to come, as more hyperspectral data from different sensors over different types of scenes for different applications come under study. At present, it seems inappropriate to attempt to analyze such high dimensional data in a single step. Rather we shall follow a process as diagrammed in Figure 21. We shall introduce a preliminary or pre-processing step that is intended to re-orient the feature space toward the specific set of classes of current interest. This will allow for the selection of a lower dimensional set of features that can be effective with a conventional classification algorithm.

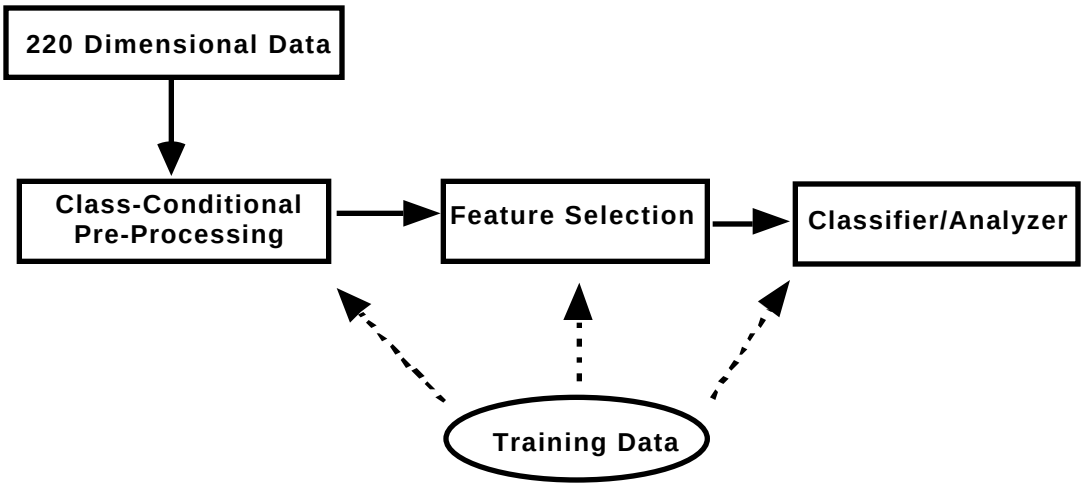


Figure 21. The concept to be followed in analyzing a hyperspectral data set.

As previously indicated, the data set to be used is a small segment of an AVIRIS data set collected in June 1992 over the Indian Pine Test site, a 100 square mile area in northwestern Indiana. The data set is the same 145 by 145 set of pixels used previously, but this time the full 220 bands will be used. The data set is designated 92AV3C, and is 9.04 Megabytes in size. The small size (21,025 pixels) of this data set, in fact, makes it too small to be of practical significance. The limited size was chosen as a compromise toward making the data set easily portable so that the reader would find it easier to download remotely. Due to the need to have more extensive sized training sets for the 220 band data, we will use the entire set of fields defined by the generalized reconnaissance map, file 92AV3gt.GIS, as the training set. Since this will mean that the training set will be the set used for accuracy evaluation, it will not be possible to test the ability of the classifier to generalize beyond its training set. This question will be left to another occasion. The file 92AV3.Project contains the fields of the generalized reconnaissance map as training fields and may be used as the MultiSpec project file.

Example Analysis Steps Using MultiSpec

Begin the analysis process by **Open**ing the project file **92AV3.Project**, and **Open** the project image using the default bands 50, 27, and 27 for red, green, and blue respectively. The fields of the generalized reconnaissance map should show outlined as training fields. The first step will be to do the class-conditional preprocessing step indicated above, to focus the needed dimensionality based upon the now available training statistics to a level that will enable a classifier to function effectively.

- At a number of points in processing hyperspectral data, exercising good judgment can reduce the amount of processing time or effect a tradeoff between processing and storage requirements. For example, the project file 92AV3.Project, has been provided without the class statistics stored with it. This significantly reduces the size of the file and speeds the process of opening the project substantially. However, it means that the project statistics must be recalculated before any processor requiring them can be used. (To do so, open the **Statistics** processor, click on the **Project** button and then the **Update P. Statistics**

button.) Saving and loading a project file with the project statistics included may be faster than recalculating them each time the file is opened, but substantially more storage space is required.

We will use the **Feature Extraction** processor under the **Processor** menu, and we will choose the **Decision Boundary** option of this processor. It will be necessary to use a subset of classes which excludes all classes that have less than 221 training samples, to avoid singularity problems with matrix inversion. Classes that must thus be excluded are Alfalfa, Grass/pasture-mowed, Oats, Wheat, and Stone-steel towers. This is a lengthy calculation. It determines a linear transformation that is optimal for the set of classes involved, and orders the features determined in order of their value in discriminating between the classes. One may then use the first N of these features to classify the data to the level of accuracy desired.

Figure 22 shows the resulting maximum likelihood pixel classification for N = 19, the largest value of N that can be used and still include all of the classes, since Oats has only 20 training pixels. Table 6 gives the quantitative evaluation of the classification. It is seen that the overall accuracy is 86.3% at this dimensionality.

- If the classification is done directly after the Decision Boundary Feature Extraction using the **Use Transformation** option of the Classifier processor, note that a 19 feature classification will still take as long as a 220 feature classification would, because the classifier must use all 220 bands to calculate the linear combination defining each feature for each pixel. A possibly faster method would be to use the **Reformat** processor of the **Processor** menu to create a transformed data set with the transform just calculated. Then classifications can be done without the **Use Transformation** option of the Classifier processor set, and only the first 19 of the features will actually be used for the 19-feature classification. That is what was done for the following examples.

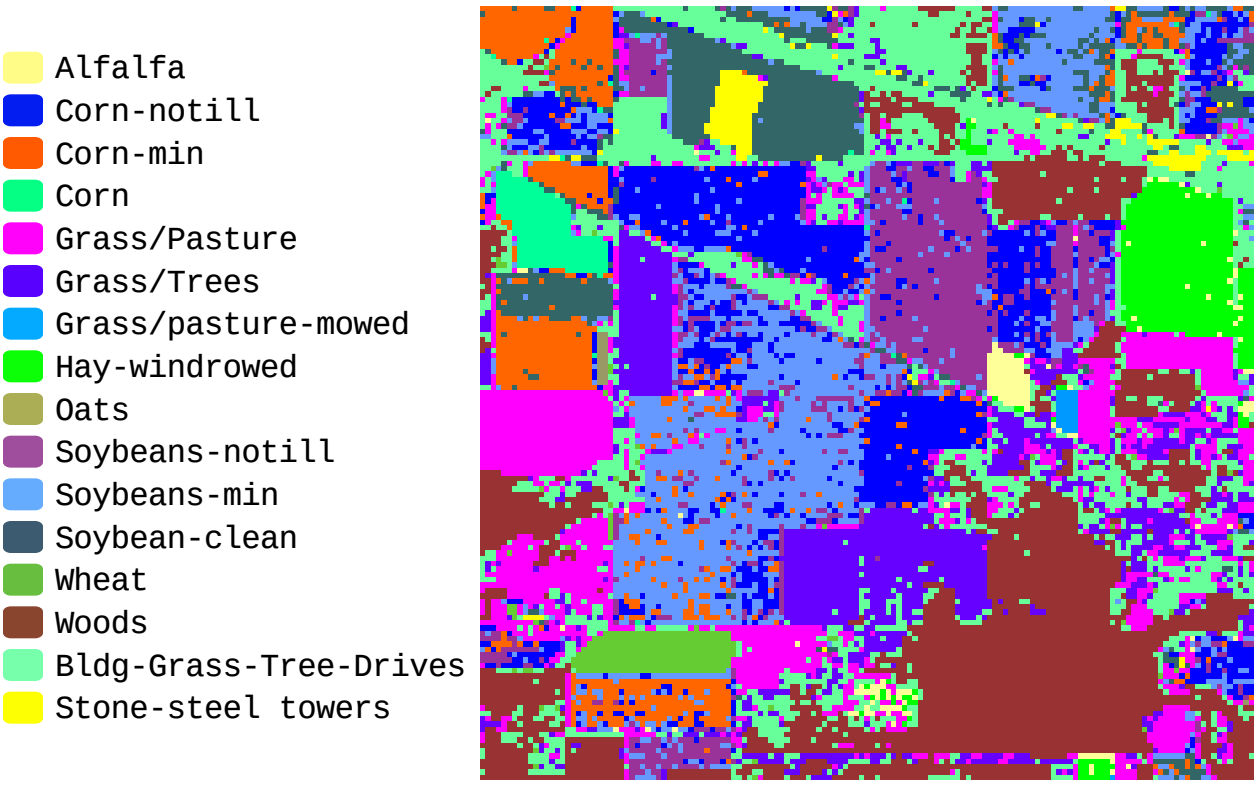


Figure 22. 19 DBFE Feature Maximum Likelihood Pixel Classification (in color).

ject			Number of Samples in Thematic Image Class																
ss	Class	Percent	Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1
ne	No.	Correct	Samples	Alfalf	Corn-n	Corn-m	Corn	Grass/	Grass/	Grass/	Hay-wi	Oats	Soybea	Soybea	Soybea	Wheat	Woods	Bldg-G	St
alfa	1	100.0	54	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
n-notill	2	83.1	1434	0	1191	46	2	2	2	0	0	0	75	110	5	0	0	1	
n-min	3	89.0	834	0	31	742	2	0	0	0	0	0	4	37	18	0	0	0	
n	4	99.1	234	0	0	0	232	2	0	0	0	0	0	0	0	0	0	0	
ss/Pasture	5	96.6	497	0	1	0	0	480	0	0	0	0	1	3	8	0	0	4	
ss/Trees	6	97.7	747	0	0	0	0	0	730	0	0	0	1	0	0	0	0	16	
ss/pas-mo	7	100.0	26	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	
-windrowed	8	98.2	489	9	0	0	0	0	0	0	480	0	0	0	0	0	0	0	
s	9	100.0	20	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	
-notill	10	87.7	968	0	47	2	0	2	2	0	0	0	849	66	0	0	0	0	
beans-min	11	68.3	2468	2	290	146	1	17	3	0	0	0	234	1685	87	0	0	2	
bean-clean	12	96.7	614	0	0	10	0	0	0	0	0	0	0	8	594	0	0	0	
at	13	99.1	212	0	0	0	0	0	0	0	0	0	0	1	0	210	0	1	
ds	14	95.7	1294	0	0	0	0	5	2	0	0	0	0	0	0	0	1238	49	
g-Grass-e	15	84.5	380	0	0	0	0	1	6	0	0	0	0	0	0	0	51	321	
ne-eltowe	16	100.0	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
AL			10366	65	1560	946	237	509	745	26	480	20	1164	1910	712	210	1289	394	

RALL PERFORMANCE (8947/10366) = 86.3

Table 6. 19 DBFE Feature Maximum Likelihood Classification Evaluation

It is seen that the classification map appears to have a good many “salt and pepper” errors to it. A more accurate classification results if the ECHO spectral/spatial classification algorithm is used on the same statistics. The results of doing so are shown in Figure 23 and Table 7.

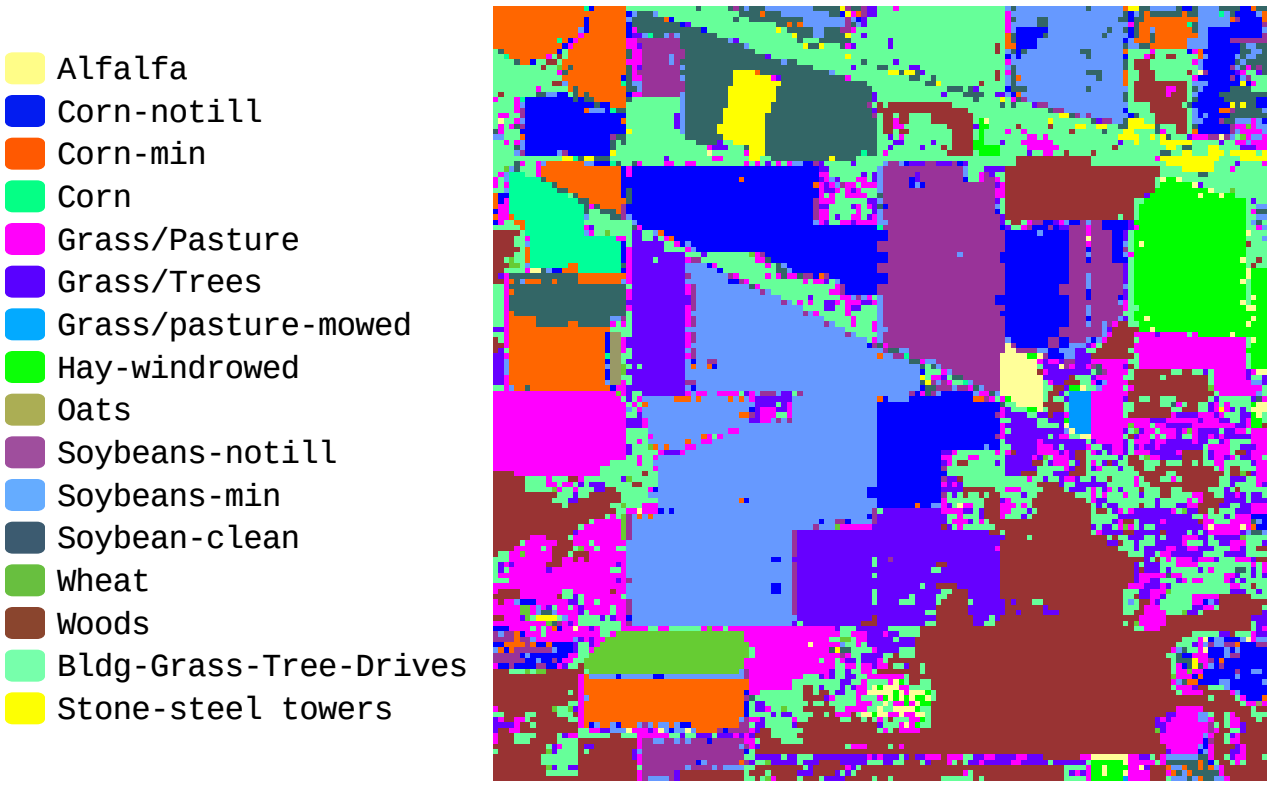


Figure 23. 19 DBFE ECHO classification (in color).

ject			Number of Samples in Thematic Image Class																
ss	Class	Percent	Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1
ne	No.	Correct	Samples	Alfalf	Corn-n	Corn-m	Corn	Grass/	Grass/	Grass/	Hay-wi	Oats	Soybea	Soybea	Soybea	Wheat	Woods	Bldg-G	St
alfa	1	100.0	54	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n-notill	2	90.2	1434	0	1294	7	2	2	2	0	0	0	36	85	5	0	0	1	0
n-min	3	96.2	834	0	6	802	2	0	0	0	0	0	2	10	12	0	0	0	0
n	4	98.3	234	0	0	2	230	2	0	0	0	0	0	0	0	0	0	0	0
ss/Pasture	5	97.0	497	0	1	0	0	482	0	0	0	0	1	3	8	0	0	2	0
ss/Trees	6	97.9	747	0	0	0	0	0	731	0	0	0	1	0	0	0	0	15	0
ss/pas-mo	7	100.0	26	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0
-windrowed	8	99.0	489	5	0	0	0	0	0	0	484	0	0	0	0	0	0	0	0
s	9	100.0	20	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0
-notill	10	96.8	968	0	12	1	0	2	2	0	0	0	937	14	0	0	0	0	0
beans-min	11	90.2	2468	1	56	16	0	17	3	0	0	0	116	2227	30	0	0	2	0
bean-clean	12	96.4	614	0	0	15	0	0	0	0	0	0	2	3	592	0	0	0	2
at	13	99.1	212	0	0	0	0	0	0	0	0	0	0	1	0	210	0	1	0
ds	14	95.2	1294	0	0	0	0	5	1	0	0	0	0	0	0	0	1232	56	0
g-Grass-Tre	15	97.4	380	0	0	0	0	1	4	0	0	0	0	0	0	0	4	370	1
ne-steeltow	16	100.0	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95
AL			10366	60	1369	843	234	511	743	26	484	20	1095	2343	647	210	1236	447	9

RALL PERFORMANCE(9786/10366) = 94.4

Table 7. 19 DBFE Feature ECHO Classification Evaluation

Indeed the accuracy can be further improved by using more of the DBFE features. Figure 24 and Table 8 give the results for the 50-feature case, again using the ECHO algorithm. Note that the accuracy has now risen to 97.8 %.

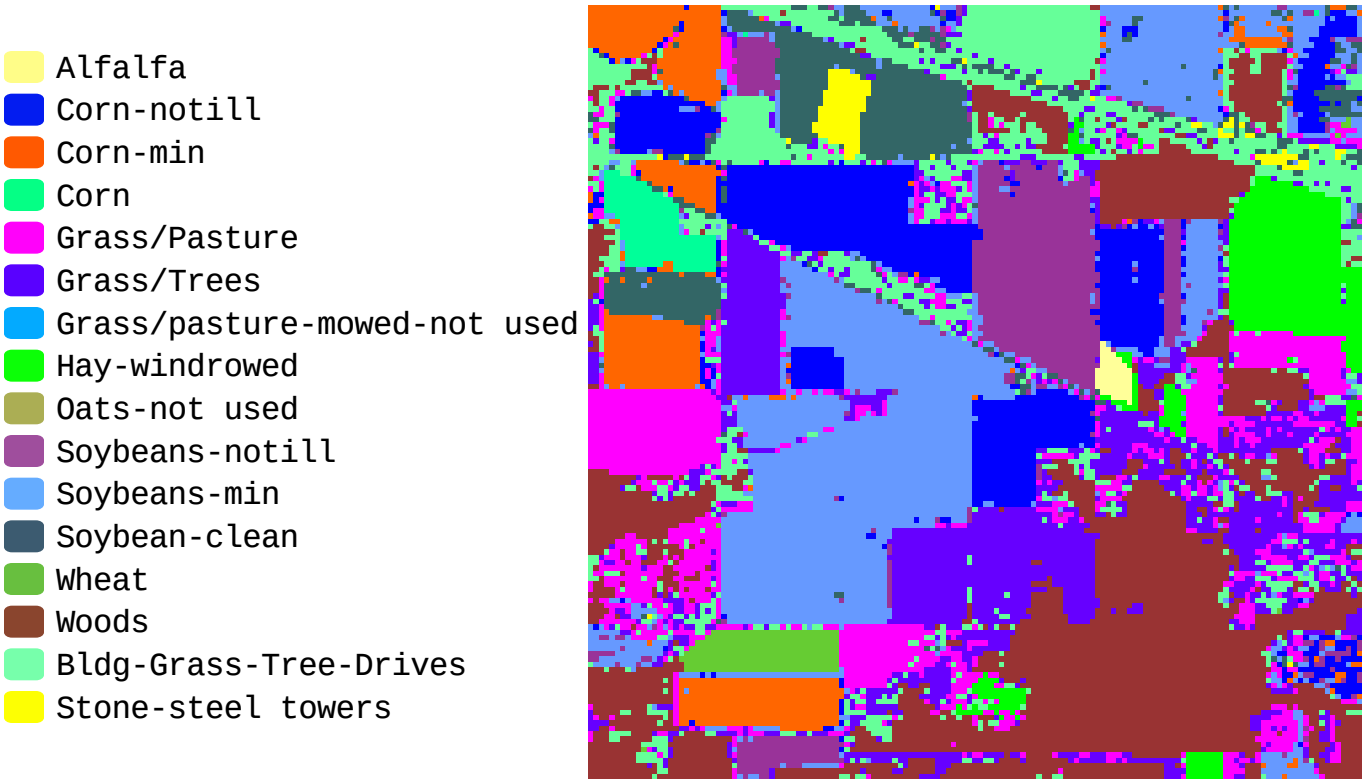


Figure 24. 50 DBFE Feature ECHO Classification (in color).

ject			Number of Samples in Thematic Image Class																
ss	Class	Percent	Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1
ne	No.	Correct	Samples	Alfalf	Corn-n	Corn-m	Corn	Grass/	Grass/	Grass/	Hay-wi	Oats	Soybea	Soybea	Soybea	Wheat	Woods	Bldg-G	St
alfa	1	100.0	54	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
n-notill	2	96.1	1434	0	1378	3	0	2	2	0	0	0	18	28	2	0	0	1	
n-min	3	99.2	834	0	2	827	0	0	0	0	0	0	0	1	4	0	0	0	
n	4	99.1	234	0	0	2	232	0	0	0	0	0	0	0	0	0	0	0	
ss/Pasture	5	99.0	497	0	0	0	0	492	0	0	0	0	0	4	1	0	0	0	
ss/Trees	6	99.9	747	0	0	0	0	0	746	0	0	0	0	0	0	0	0	1	
ss/past.-mo	7	0.0	26	0	0	0	0	1	1	0	23	0	0	1	0	0	0	0	
-windrowed	8	100.0	489	0	0	0	0	0	0	0	489	0	0	0	0	0	0	0	
s	9	0.0	20	0	4	0	0	11	2	0	0	0	0	1	0	0	0	2	
beans-notil	10	98.6	968	0	3	0	0	2	2	0	0	0	954	7	0	0	0	0	
beans-min	11	97.6	2468	0	9	5	0	12	6	0	0	0	12	2408	14	0	0	2	
bean-clean	12	98.5	614	0	0	4	0	0	0	0	0	0	2	2	605	0	0	0	
at	13	99.5	212	0	0	0	0	0	0	0	0	0	0	0	0	211	1	0	
ds	14	98.5	1294	0	0	0	0	1	2	0	0	0	0	0	0	0	1274	17	
g-Grass-e-	15	98.9	380	0	0	0	0	0	0	0	0	0	0	0	0	0	4	376	
ne-eltowe	16	100.0	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
AL			10366	54	1396	841	232	521	761	0	512	0	986	2452	626	211	1279	399	

RALL PERFORMANCE(10141/10366)=97.8

Table 8. 50 DBFE Feature ECHO Classification Evaluation.

Figure 25 shows a graph of accuracy vs. number of features used for the DBFE features and shows that at 50 features, the accuracy is still increasing for the Maximum Likelihood pixel classifier. For ECHO, the accuracy is slightly higher than for the Maximum Likelihood Pixel classifier. In spite of the greater complexity of this algorithm, it is also usually as fast or faster than the pixel classifier.

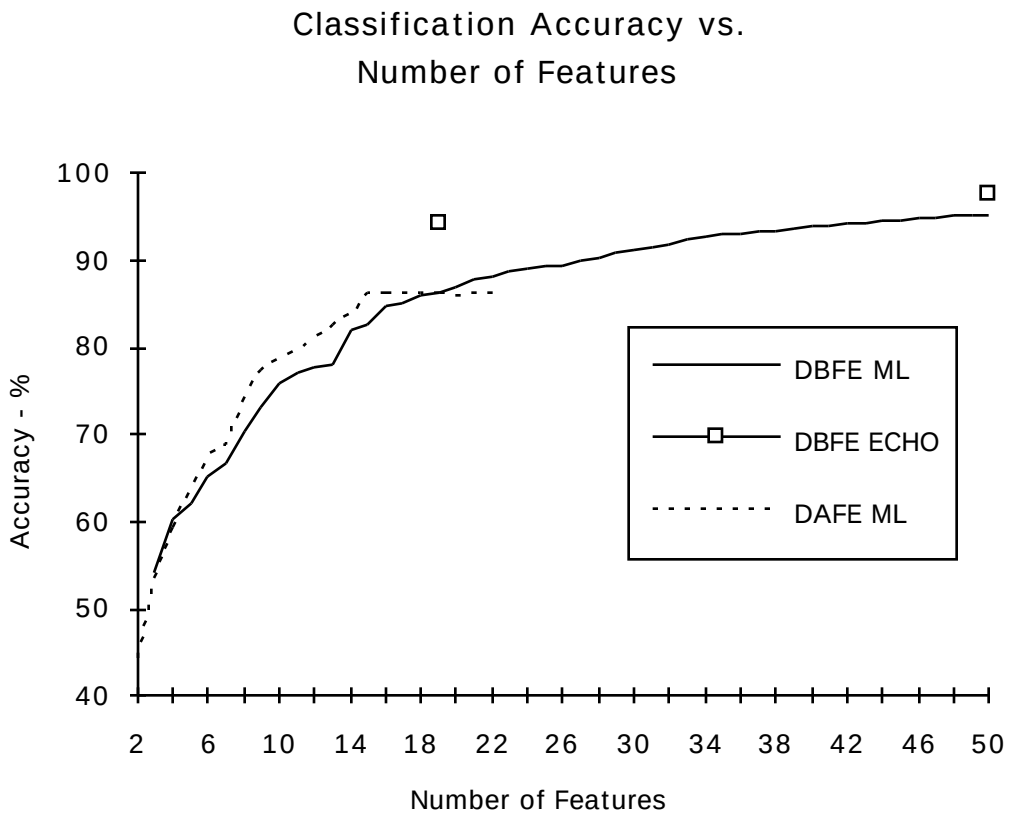


Figure 25. Accuracy vs. number of features for the Decision Boundary and Discriminant Analysis Feature Extraction algorithms with Maximum Likelihood Pixel classification. For the DBFE algorithm, the ECHO classifier accuracy is also indicated for 19 and 50 features.

- Another possible method for carrying out the Class-Conditional Preprocessing calculation is to use the Discriminant Analysis Feature Extraction (DAFE) option of the Feature Extraction Processor. This option has several desirable features. It is a much shorter calculation, and classes may be included even though they have a number of training samples less than the number of features. It also tends to give higher accuracy so long as the number of features used is less than M – 1 where M is the number of classes. However, it calculates optimal features only up to M – 1, and the features determined beyond that point are unreliable. The performance for the Maximum Likelihood pixel classifier using DAFE features is shown in Figure 25. It is seen that the rate of increase of performance became small as the number of dimensions, N, approached the number of classes, M=15, in the above graph. Thus theory and this example suggest that a good practice would be to use DAFE when the number of classes to be used is greater than the number of features needed, otherwise to use DBFE.

Closing Comments

In this example we have shown one set of steps which can be used to analyze hyperspectral data. The method produced arbitrarily high accuracy even for classes that have a signal-to-background ratio that is very small, thus demonstrating the increased value of such high dimensional data. We note, however, that the

generalization characteristics of the particular example were not demonstrated here, due to the constraint upon data set size imposed arbitrarily on the example to hold the data set to a size that can be easily transferred.

Development of these techniques continues. A number of newer features resulting from continuing research but not used in the hyperspectral analysis case above have already been incorporated into MultiSpec or are pending. Among these are the following.

- Statistics Enhancement Processor. This processor improves the generalization characteristics of a given set of training statistics relative to a data set to be analyzed[21].
- LOOC (Leave One Out Covariance estimation) This algorithm allows for a more effective estimation of class covariance matrices when the number of training samples is small[22].
- Projection Pursuit Feature Extraction. We note the following. DAFE may limit performance when there are only a small number of classes or where the classes have similar mean responses. DBFE may not perform well when the ratio of the number of training samples to dimensionality is not large. Projection Pursuit is an additional feature transformation algorithm for which all calculation is done in the final (reduced) dimensionality, rather than the initial dimensionality. It has been incorporated into MultiSpec[23], as a scheme for preliminary feature reduction preceding use of DAFE or DBFE.
- Several additional classifier algorithms have been added to MultiSpec recently, among them the Correlation Classifier sometimes called the Spectral Angle Mapper (SAM), a matched filter classifier called Constrained Energy Minimization (CEM)

A number of other aids to the training analysis process have been incorporated into MultiSpec. For example, one can pre-specify the number and location of initial cluster centers for a clustering process.

In considering the acquisition of information by multispectral remote sensing means, one must ask the following questions.

1. Is the desired information actually present in the electromagnetic variations traversing the space between the target and the sensor? If so,

2. Does the sensor successfully capture these variations with enough precision and detail to insure that the desired information is now contained in the data stream coming from the sensor to the analyst's location?

---

21 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.

22 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.

23 Luis O. Jimenez and David Landgrebe, "High Dimensional Feature Reduction Via Projection Pursuit," School of Electrical & Computer Engineering Technical Report TR-ECE 96-5, April 1996.

If the answer to both of these questions is yes, then it is up to the analyst to find a way to discriminate between the desired diagnostic variations and all others, which are invariably present. According to the framework presented above, the task of the analyst is to find a way to partition the N-dimensional feature space in such a way that the feature variations associated with materials of interest are contained within subregions of this space as uniquely and mutually exclusively as possible.

The factor that changes substantially with hyperspectral data over that of conventional dimensionality is the magnitude of the volume available in this space. This is at once both the opportunity that was sought and a challenge. The magnitude of this volume is the very thing that provides the opportunity for many more materials to be discriminatable. However, it is this very characteristic that adds substantially to the complexity of doing so. As has been seen, as the dimensionality goes up, the ability of the analyst to use intuitive concepts which otherwise seem self-evident, is precluded, and in fact many of these concepts become misleading. Instead, one must rely upon quantitative tools that objectively provide information to the analyst about the separability of spectral responses in feature space, so that the right path through the analysis process can be quickly found. Indeed, at this stage of development, it is not so much, “Is it possible to discriminate between this material and others in the scene,” as it is, “Can I find a way to discriminate between this material and others in the scene.”