LAND SATELLITE INFORMATION EXTRACTION: PAST, PRESENT, AND FUTURE

092595

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

Land satellite technology has been centered upon (multivariant) multispectral analysis since the beginning of the Landsat series. A key limiting factor in the past has been the precision and detail with which the spectral response of Earth cover materials could be measured from orbit. The first generation system, the MSS defined in 1968, had only four broad bands spaced closely together with a six bit data system; the second generation system, the TM defined in 1975, has seven bands which are more evenly spaced and an eight bit data system. Sensor technology and processing capabilities have now advanced to the point of making possible bands into the hundreds and doing so while allowing a signal-to-noise ratio justifying data of 10 or more bit precision. Along with these advances, higher spatial resolution has also become possible.

Following a brief historical perspective, in this paper, we examine the rationale for going to large numbers of spectral bands, and discuss what is different about such data that affects how one approaches the information extraction process. We also briefly address the obstacles which must be overcome to make such data sources practical for the user community and the effect of higher spatial resolution.

INTRODUCTION

Soon after the launch of the world's first weather satellite on April 1, 1960, attention began to be given to the matter of monitoring the Earth's land surface via satellite observation. The concern was that the Earth's resources are finite, and thus with continued pressure upon them from a growing population, more careful management of them was required. Considerations focused not only upon environmental but economic considerations as well. By 1964 a cohesive laboratory/field/airborne research program had been formulated and began to be pursued with an interdisciplinary team approach.

From the beginning the objective was to take advantage of the synoptic view from space to be able to gain information about broad areas very economically. Thus the techniques that began to be developed were required to be functional over large areas, but not labor intensive. Electronic computation at the time was still rather limited in its capabilities, however it had more promising characteristics relative to these requirements than did highly human-intensive methods. Within the electronic computational family of approaches, it appeared that image processing techniques would not be a good approach, because such techniques require higher spatial resolution, something that is very expensive. The volume of data goes up with the square of the spatial resolution. thus by the point at which one would have high enough spatial resolution to use these techniques to identify Earth cover types, the volume of data would be economically prohibitive. This largely continues to be the case today. Thus what was needed was a technique that would provide for identification of materials with relatively gross resolution and computational procedures which were effective without being too complex, and which would be robust enough to function over large land areas at any given observation time. It was at this point that the multispectral concept coupled with pattern recognition schemes was introduced.

Presented at the American Society of Photogrammetry and Remote Sensing Symposium, "Landsat Satellite Information in the Next Decade," Vienna, VA, September 25-28, 1995.

Multispectral methods require measurements of the up-welling radiation in enough wavelength intervals to allow discrimination between surface cover types that are present in a given scene. i.e., to insure that the information desired is contained in the data. Then, to extract the information from the data, pattern discriminators could be used to divide the data set into categories of interest. Pattern recognition methods have their origin in the same basic principles that led to advances in communication systems, where the problem is to detect the presence of a signal in the face of noise, even when the signal-to-noise ratio is well less than one. The problem then came down to (a) learning enough about the spectral reflectance of various materials to be able to predict the circumstances under which various materials of interest would be discriminable from such other materials as may be expected to be in the same scene at the same time, (b) devising suitable adaptations of the emerging pattern recognition technology for the remote sensing circumstances. This work was pursued throughout the middle 1960's, and by 1967, enough had been learned and demonstrated that specifications for an land-oriented satellite could be defined.

By this time, it was well recognized that there could be a wide variety of uses for a land observing satellite. Some of them could be best serviced by providing broad scale images which could be directly interpreted. Others would be best served by the new multispectral technology, which would enable the labeling of each pixel in the scene with a class label of interest. Thus the recommendation was for two types of sensors on board this first spacecraft, one which was image-oriented, and a second which was more spectrally oriented. This plan was put forth in 1968[1] and the satellite was ultimately launched as ERTS-1 (later renamed Landsat 1) in July 1972.

A quasi-operational demonstration that the multispectral approach was viable came about as a result of external events in 1971, the year before Landsat 1 was ready for launch[2]. A significant pathogen called the Southern Corn Leaf Blight began to emerge late in the 1970 growing season. Though it didn't progress very far before the end of the season, it's nature was such that the next season it might be able to devastate the entire U.S. corn crop. Thus there was grave concern about the 1971 growing season. As a result, plans were made to attempt to monitor the progress of this pathogen during the 1971 season by monitoring 220 selected areas located throughout the seven states of the U.S. Corn Belt every two weeks during the 1971 season. An RB-57F high altitude reconnaissance aircraft was made available to the program to fly the photographic missions over 200 of the segments, however, the only aircraft available with a multispectral scanner was a DC-3 containing what came to be known as the M-7 scanner. This aircraft was operated by a special unit of the University of Michigan (later ERIM) and flew bi-weekly missions over 20 flightlines selected over the western half of Indiana.

From a remote sensing standpoint, the task was not a simple one. One must not only be able to identify a specific plant species, corn, but one must be able to subdivide the class into categories depending on the plant condition, i.e., the degree of blight infestation, and do so throughout the growing season. The successful execution of this project was very reassuring about the basic viability of the multispectral approach. A number of other broad scale demonstrations of the multispectral concept were successfully conducted in the years following the launch of Landsat 1.[3]

LIMITATIONS ON THE SPACEBORNE TECHNOLOGY

As previously indicated, a key factor with multispectral sensing is the detail with which the spectra of pixels is measured. The M-7 scanner, which was the early airborne workhorse for remote sensing research was such that its configuration of bands could be varied, usually from 12 to 18 or so over the full optical range. A typical configuration is illustrated in Figure 1 relative to

1 See for example, the report of Panel 6, Sensors and Data Systems, National Research Council report on Useful Applications of Earth-Oriented Satellites, National Academy of Science, Washington DC, 1969.

2 MacDonald, R.B., M.E. Bauer, R.D. Allen, J.W. Clifton, J.D. Ericson and D.A. Landgrebe, "Results of the 1971 Corn Blight Watch Experiment," Proceeding of the Eighth International Symposium on Remote Sensing of Environment, Vol. I, Environmental Research Institute of Michigan, Ann Arbor, Michigan, pp 157-190, 1972.

3 See for example, descriptions of such projects in Swain, P.H. and S. M. Davis, *Remote Sensing: The Quantitative Approach*, Chapter 6, McGraw-Hill, 1978.

the typical spectral characteristic of green vegetation. Notice that the coverage is well was distributed over the spectrum and the sampling is reasonably detailed, especially in the visible portion of the spectrum.

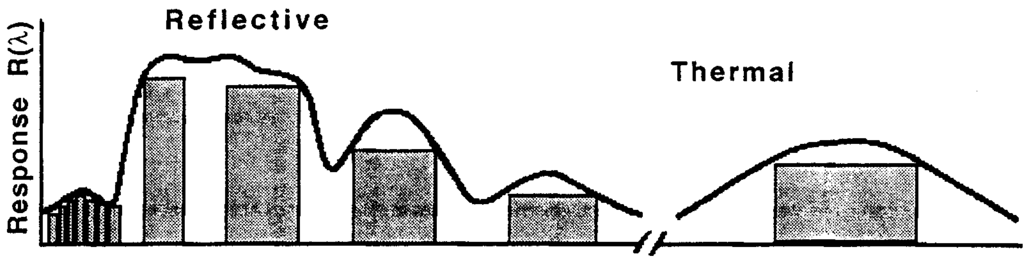


Figure 1. A typical band configuration of the M-7 airborne scanner relative to the response of green vegetation. Notice the detail of the measurement in the visible portion of the spectrum.

Unfortunately, at the time of specification of the first generation Landsat 1 scanner, known as MSS, this level of detail was beyond the state of the art for space sensors. Thus a substantially less detailed measure of spectral content was available when MSS flew. See Figure 2.

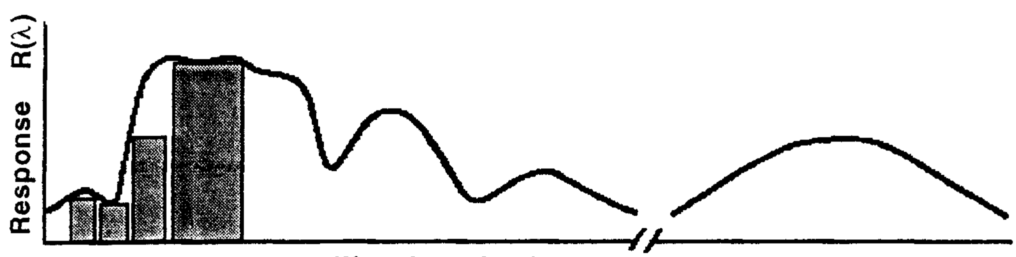


Figure 2. The band configuration for MSS, the first generation multispectral space sensor system.

The design effort toward a second generation system was begun soon after the launch of Landsat 1. It culminated in commitment to a design in 1975 containing seven spectral bands. Its larger number of better spaced bands (See Figure 3) together with its improved signal-to-noise ratio[4] represented a significant improvement over that of the MSS. Nevertheless, Thematic Mapper still was limited by the state of the space sensor art at the time of its design to something well less than that available in an airborne sensor in the 1960's.

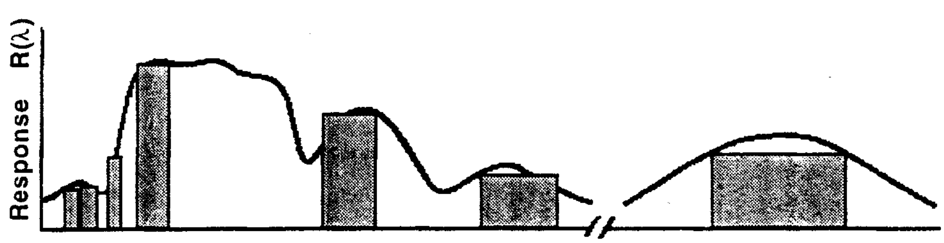


Figure 3. The band configuration for TM, the second generation multispectral space sensor system.

Thus in the past, a limiting factor in the implementation of the multispectral approach to remote sensing has been the sensor state of the art and the degree of detail with which the spectral response of land surface cover materials can be measured by a space sensor. There have been two other key factors which have limited both the research and the applications of this field. These are the timeliness with which data are delivered and the cost of the data. The long gap between the time when data are collected and the time when it is delivered to the user has had a profound effect on both research and operational use of this technology. The long gap has caused users to become effectively divorced from the data collection process. For the researcher, this means that it has

4 The second generation Thematic Mapper system of Landsat 4 and 5 has an eight bit data system as compared to the six bit system of the MSS of Landsats 1, 2, and 3.

become much more difficult to understand the cause-effect relationships between data and implications to it in scene factors, since by the time the researcher receives the data, the scene has changed significantly. From an applications standpoint, many possibly valuable uses of this technology simply have never been addressed, because their nature is very dynamic and long data delivery intervals make such uses unfeasible. However, neither of these latter two limitations are technologically based, and thus we shall remain focused on the matter of spectral detail in the sequel.

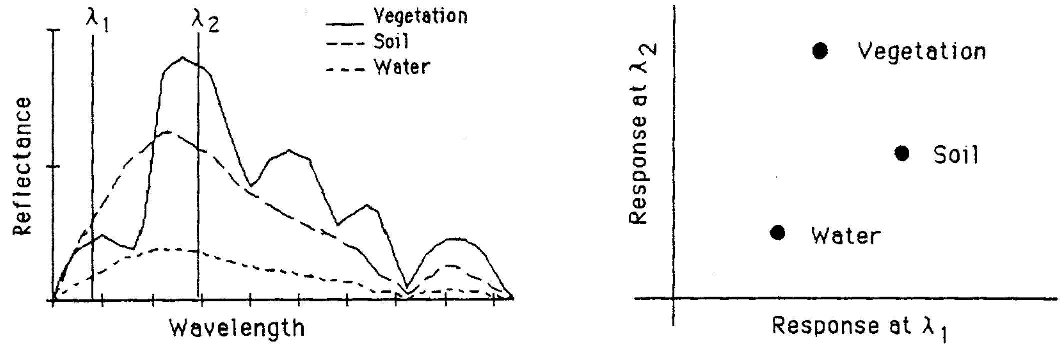
THE CHARACTER OF HYPERSPECTRAL DATA

The substantial advances which have taken place in sensor technology and particularly in solid state technology since the time that the design of the Thematic Mapper was defined now make possible substantial advancements in the level of spectral detail that can be obtained from a spaceborne sensor. Airborne systems are in operation or are under construction with as many as several hundred spectral bands. At the same time signal-to-noise ratios from such sensors are high enough to justify 10 to 12 bit data systems. Such possibilities raise the questions, is such detail useful, and if so, how must data analysis techniques be changed so as to take maximum advantage of it. We shall address the latter question first.

Such high dimensional data is often now referred to as imaging spectrometer data, due to the similarity between such data and that which arises from a chemical spectrometer in the field of chemistry. Or equivalently, it is simply called hyperspectral data. An increase in the number of spectral bands by as much as two orders of magnitude means that a cursory look at analysis technology for the new type of data would not be adequate. A much more fundamental study is required. Thus a research effort was begun a few years ago to come to understand what is different about high dimensional data and to devise data analysis algorithms that would be most effective with it. We shall briefly outline example pieces of information that have resulted from that work. First we note the usual method for representing such multispectral data mathematically.

The standard way of representing a continuous variable such as $R(\lambda)$ vs. λ is to use a generalized transform to represent it in an N-dimensional feature space. For example, Figure 4 shows how sampling can be used as the transform so that each response function sampled at two wavelengths becomes a point in 2-dimensional space. If the a response function is sampled at more wavelengths, the corresponding point in feature space remain a point, but the dimensionality of the space goes up. Thus, for a two hundred band sensor, one is working with points in 200-dimensional space.

It has long been known that land surface cover type cannot be adequately represented even in low dimensional space by a single spectral response function, i.e. a single point in feature space. Real ground cover classes have a variability about them, and thus in spectral space they must be characterized by a family of response functions. In feature space this means that they are described by a "cloud" of points or a distribution.



Spectral Space N-Dimensional Feature Space

Figure 4. The mapping from Spectral Space to N-Dimensional Feature Space.

A part of the research, then, has been to learn what higher dimensional space is like and what kind of rules apply there in characterizing distributions. It has been found that high dimensional space has very different characteristics, and the usual rules of geometry that we are so familiar with from living in a three dimensional space do not necessarily extrapolate to higher dimensional space.

For example[5], it has been shown that as the dimensionality increases,

(1) *The volume of a hypercube concentrates in the corners.* The fraction of the volume of a hypersphere inscribed in a hypercube is:

$$f_{d1} = \frac{\text{volume} - \text{sphere}}{\text{volume} - \text{cube}} = \frac{\pi^{d/2}}{d2^{d-1}\Gamma\left(d/2\right)}$$

where d is the number of dimensions. Note that as d increases, this fraction approaches zero.

(2) *The volume of a hypersphere concentrates in the outside shell.* The fraction of the volume of a sphere of radius r-ε inscribed in another sphere of radius r is:

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

Note that as d increases, this fraction approaches unity.

Both of these characteristics show that high dimensional space is mostly empty, which implies that multivariate data in an N-dimensional space is almost always in a lower dimensional structure. A concrete consequence of this is that normally distributed data will have a tendency to concentrate on the tails; meanwhile uniformly distributed data will be more likely to be collected in the corners, making density estimation more difficult.

(3) *The diagonals are nearly orthogonal to all coordinate axis.* The cosine of the angle between any diagonal vector and a Euclidean coordinate axis is:

$$\cos(\theta_d) = \pm\frac{1}{\sqrt{d}}\text{, where } \lim_{d\to\infty}\cos(\theta_d) = 0$$

5 Luis Jimenez and David Landgrebe, "High Dimensional Feature Reduction Via Projection Pursuit," Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'94),CD-ROM pp 1473-1479, Pasadena, Calif, August 8-12,1994.

This piece of information is important, because the projection of any cluster onto any diagonal, e.g., by averaging features, could destroy information of multispectral data.

Perhaps more directly to the point, it has been found that for discriminating between two classes, second order statistics, i.e., the correlation matrix, which provides information about the shape of the distribution of a class, is more significant in high dimensional space than are first order statistics, i.e., the mean vector. Indeed, two classes of data can have the same mean value, meaning that they lie on top of one another, but if they have different correlation matrices, i.e., their distributions have different shapes, they may still be quite separable. While this may make little sense in three dimensions, it "makes good sense" in high dimensional space.

This fact has very significant implication for remote sensing systems. For example, most "data correction" algorithms are directed at adjusting the mean value of the data, rather than the manner in which the data varies around the mean. Indeed it is possible that "corrections" that are made to the data may change the manner in which the data varies about its mean, but in a manner unrelated to the separability of classes of interest. Thus, it is possible that a well-intended calibration procedure could make separability between classes of interest worse instead of better.

AN APPROACH TO HYPERSPECTRAL DATA ANALYSIS

Note that in discussing what is different about hyperspectral data, we have not mentioned the high data volumes that are involved. The data volumes are indeed large, however, this is a manageable aspect of the problem given the rate at which computational technology is advancing. The more significant problems arise from the characteristics described above, together with those of practical remote sensing circumstances.

To elaborate, in order to discriminate between any two classes of land surface cover, one must be able to determine an adequate mathematical description of the N-dimensional data distributions for each class. Usually this is done with the use of training samples, i.e. samples from the data set to be analyzed which by some means one is able to label by class as representative of that class. It is characteristic of the remote sensing situation that there are only a limited number of these samples for each class, as the labeling process is often the most difficult part of the analysis process. However, as the dimensionality goes up, the number of such samples needed to adequately define the shape of the class distribution in N-dimensional space goes up quite rapidly. The inadequately accurate estimation of class distributions is the major limitation to achieving the full potential of hyperspectral data in an analysis task.

To minimize the impact of this problem, we have approached the design of analysis methods for hyperspectral data as shown in Figure 5. The classical approach to multispectral analysis is to use a feature selection scheme to determine the optimal subset of bands for use with the classifier. To this we have added a class-conditional preprocessing step. The rationale for this step is seen from the above noted characteristic of hyperspectral data, that of high dimensional space being mostly empty, and thus that the data characteristics of interest usually lie primarily in a subspace. The problem is that one cannot predict which subspace it will be until one has an adequately accurate estimate of the class distribution. That is, the particular subspace depends upon the specific data set and the classes desired. Thus what is needed is an effective means by which to determine from the training data which subspace is optimal.

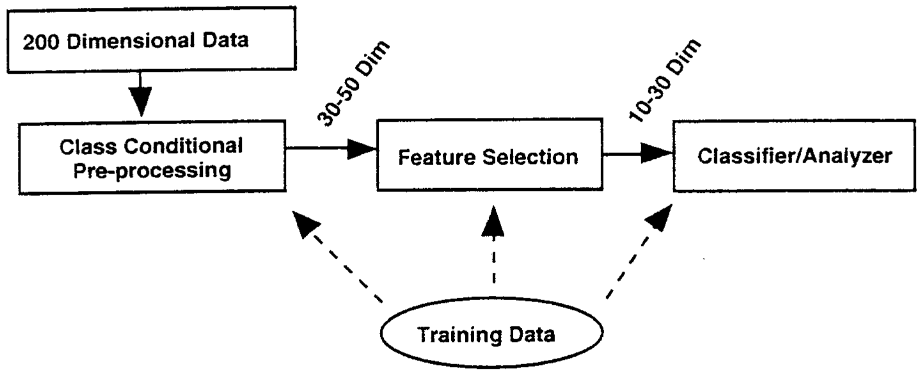


Figure 5. The steps in one concept to the analysis of hyperspectral data.

In order to deal with the varied circumstances which arise, we have devised several means for accomplishing this, as follows.

- Discriminate Analysis Feature Extraction (DAFE). This basic approach has been known for some time[6] and we have simply adapted it to the remote sensing circumstances. It is primarily useful when the number of classes small.

- Decision Boundary Feature Extraction (DBFE). This is a wholly new approach designed specifically for the remote sensing situation[7]. It is useful for both small and large numbers of classes.

- Projection Pursuit. This is a recently devised scheme that is still being explored[8,9,10]. It has the advantage that all calculations are done in a lower dimensional space.

Each of these use a somewhat different basis for finding the optimal subspace within which to proceed, given the limited training set available. In addition, we have devised a scheme for making optimal use of the limited training set size in determining the best complexity to use in the classifier[11]. Further, we have also devised an algorithm to improve the capabilities of the classifier to generalize beyond its training areas to the entire data set to be analyzed[12,13]. All of these algorithms have been devised specifically with hyperspectral data in mind.

---

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

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

8 Luis Jimenez and David Landgrebe, "High Dimensional Feature Reduction Via Projection Pursuit," Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'94),CD-ROM pp 1473-1479, Pasadena, Calif. August 8-12,1994.

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

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

11 Joseph Hoffbeck and David Landgrebe, "Covariance Estimation For Classifying High Dimensional Data," Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'95), Florence Italy, July 10-14, 1995.

12 B.M. Shahshahani, D.A. Landgrebe, "On the Asymptotic Improvement of Supervised Learning by Utilizing Additional Unlabeled Samples; Normal Mixture Density Case," SPIE Int. Conf. Neural and Stochastic Methods in Image and Signal Processing, San Diego, CA, July 19-24, 1992.

13 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.

THE TECHNOLOGY TRANSFER PROBLEM

Another factor associated with the use of hyperspectral data is that of complexity. It stands to reason that the substantially increased complexity of hyperspectral data calls for increased complexity in the analysis algorithms if the full value of such data is to be achieved. However, how is the transfer of understanding of these more complex algorithms to be made from the signal processing engineers who must devise them to the Earth science user who will take advantage of them?

To deal with this problem, early in our research program we formed a specific technology transfer element as an integral part of the overall effort. Specifically, we constructed a software system which contained all necessary elements of an analysis system for conventional multispectral data. The intention was to then add to this system the new algorithms resulting from the research program as they emerged and to give a copy of this system to anyone asking for it. The specific design goals for the systems are,

- The implementation should be on a readily available platform which has adequate processing power, but financially within the reach of any Earth science user (i.e., ≤ $5000).

- The system should be easy to learn and easy to use, even for the infrequent user, using the most modern of software environments.

- The system should provide for easy import of data in a variety of formats, and easy export of results, both in thematic map and in tabular form.

This approach to making the new technology has now been under way for the last half dozen years. The system, now called MultiSpec (© Purdue Research Foundation), has more than 500 registered requesters in nearly all states and more than 25 foreign countries. It and its documentation is now distributed via the World Wide Web from the following URL:

fttp://dynamo.ecn.purdue.edu/~biehl/MultiSpec

The current capabilities of the system may be summarized as follows.

- **Import Data** (Binary or ASCII)
- **Histogram & Display Images**
- **Cluster** (Single Pass or Iterative)
- **Define Classes** (Rectangular or Polyonal areas, or by Clustering)
- **Feature Definition** (Subsets, or Optimal DBFE or DAFE Subspaces)
- **Statistics Enhancement** (Via Labeled & Unlabled pixels)
- **Classification** (Spectral or Spectral/Spatial)
- **Results Display** (Thematic and Tabular) Bands, Mosaicing, Changing Geometry, Visualization aids, etc.)

THE POTENTIAL OF HYPERSPECTRAL DATA

But what can be said for the value of hyperspectral data to this point compared to more conventional multispectral data? In fairness, this remains a question not yet fully answerable. There have been no broad scale tests of this technology such as were done in the 1970's with conventional data, and indeed, the full complement of algorithms needed for achieving it full potential is not yet in place. However, there have been enough results appearing to show a clear incremental value to such data. A number of results have been shown regarding the mapping of minerals using more straightforward algorithms, results which would not have been achievable with conventional data. Some results have also been shown for problems in other fields as well. We will conclude by showing one such result for a problem chosen specifically for its difficulty.

The problem picked is from the field of agronomy, which lends itself especially to the quantitative evaluation of results. The data set collected was gathered at a time of the season when corn and soybeans had only recently emerged and had about 5% ground cover. In addition, the fields in the area selected had undergone a variety of tillage practices from no-till, to minimum till, to clean till. This would imply that there were various amounts of residue from the previous year's crop present on the surface. Further, there would be the natural soil type variability present as well.

Thus the problem is to distinguish between the spectral response of corn and that of soybeans when the canopy of each averages about 5% and there was a variety of residues and soil type background contained in the other 95% of ground cover. The data used was collected by the AVIRIS sensor which collects data in 10 nm intervals from 0.4 to 2.4 µm, about 210 bands all together. The ground spatial resolution is 20 m. First a subset of 9 bands were selected, the first 6 of which were the AVIRIS bands which are centered on the 6 TM bands in that spectral region. Three addition bands were selected from wavelengths not covered by TM, and conventional maximum likelihood techniques were used to achieve a classification. The result is shown in Figure 6. The overall accuracy was about 65%, a level presumable below one that would be acceptable. Using even all of the newly available schemes in MultiSpec, it was not possible to improve this above about 80%.

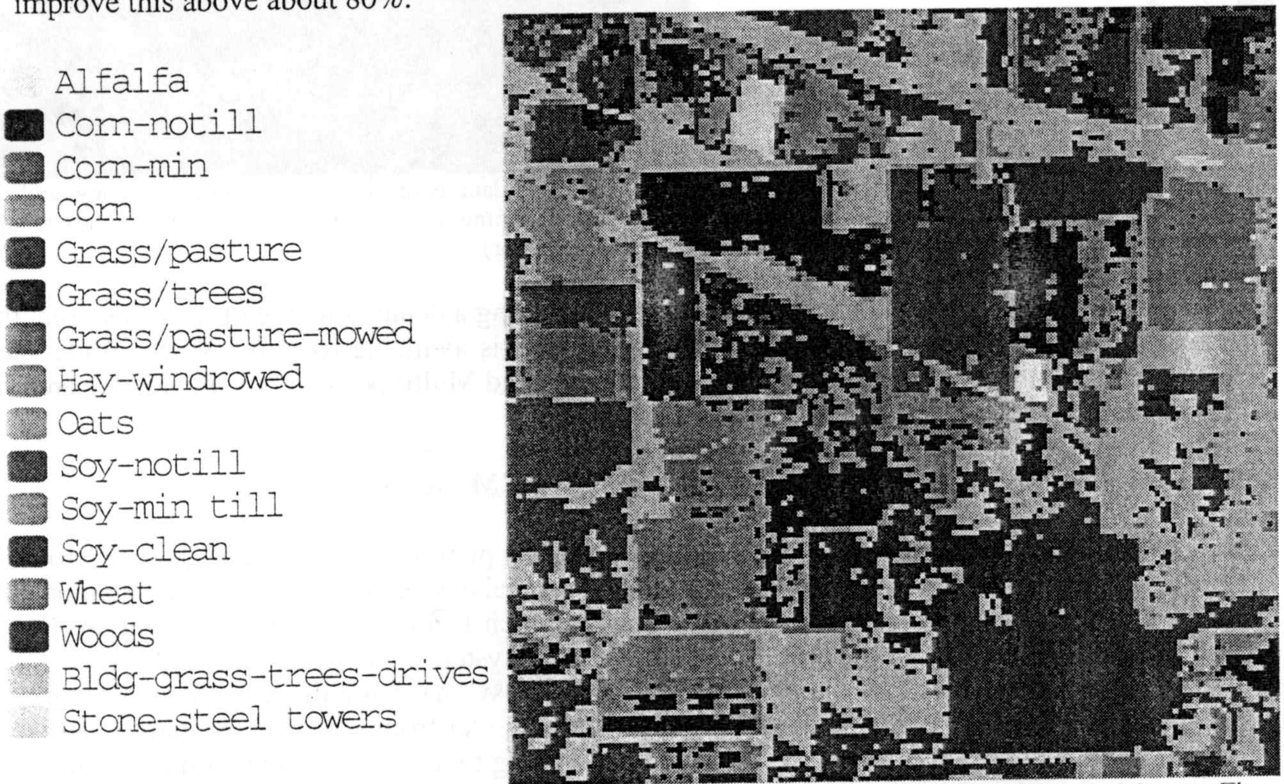


Figure 6. A conventional maximum likelihood classification using bands similar to the Thematic Mapper. The classification accuracy relative to the ground truth map of Figure 7 is 65% (original in color).

Next, the full set of AVIRIS bands was used to analyze the same data set for the same classes. The result is shown in Figure 7. The accuracy level was now near 100% measured by the same set of samples as used in the results of Figure 6. We do note, however, that the entire set of labeled samples had to be used for training because of the larger number of spectral bands. Thus the two are not entirely comparable. Nevertheless, the ability of hyperspectral data to spectrally separate classes with only very subtle differences in spectral response is demonstrated by this classification.

Figure 7. A 50 DBFE feature ECHO classification of the same data as in Fig. 6. The overall accuracy achieved is 97.8%, although in this case all test samples where use for training, due to the larger number of bands involved. (original in color)

A complete description of these experiments, including a detailed listing of the steps used, the training and test sample areas used, and the data itself is available from the World Wide Web MultiSpec page indicated above, so that both the data and MultiSpec can be downloaded and the results repeated.

CONCLUDING REMARKS

In this paper we have endeavored to give a historical picture of the Landsat program from its beginning, from a signal information content and data analysis perspective. We have attempted to show that the full potential of the multispectral approach from space systems has up until now been limited by the state of the art of spaceborne sensor systems, and that sensor system advances in recent years have advanced that potential substantially. We have attempted to show that working in high dimensional feature spaces is quite different than that from available techniques, and that some new techniques now available are already beginning to demonstrate the promise of this new technology. At the same time we caution that this technology is not yet mature. And finally, we have indicated a convenient way to access some of the new algorithms that have been created for analyzing hyperspectral data, so that familiarity and skill in the new environment can begin to be accumulated.

Acknowledgement: Work reported in this paper was sponsored in part by NASA Grants NAGW-925 and NAGW-3924.