12/3/97

On Information Extraction Methods for Hyperspectral Data

David Landgrebe

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

ABSTRACT

Multispectral image data methods have been under study for use in land remote sensing applications for the last three decades. Though early airborne-based studies with a sensor having 12 to 18 bands provided very promising results from the beginning, spaceborne sensors have been generally restricted to 3 to 7 bands, thus significantly limiting their information-providing potential. Over the last decade, the emergence of hyperspectral sensors, with up to several hundreds of bands have largely eliminated that limitation. The key problem that remains is how to analyze this much more complex hyperspectral data most effectively so as to realize the full potential of such high dimensional data.

A more fundamental signal processing approach to the analysis problem has been under study for the last several years. In this paper, we outline some of the concepts that have emerged by taking this perspective, and we provide some of initial illustrations of its use.

Keywords: hyperspectral, multispectral, data analysis, information extraction

INTRODUCTION

Multispectral data analysis has been under study since at least the middle 1960's. Early progress toward practical means for analyzing multispectral data of land surface areas was quite rapid during the 1960's and 1970's. The primary limiting factor during that time was the rather crude spectral characterization of the reflected and emitted energy from the land surface subject material (3 to 7 spectral bands and 6 to 8 bit precision). The coming of hyperspectral sensors over the last few years has largely removed that limitation. Instead the primary limiting factor to deriving maximal scientific and practical information from such data now becomes the deriving of suitable means for analyzing the much more complex hyperspectral data.

Over the last decade or so, progress on multispectral analysis technology in the field at large has not been as great as might be expected, given the substantial effort by a large variety of workers. This slowed progress suggested that perhaps a different approach to the problem was needed, taking an alternate, more fundamental point of view. The intent was to come to understand the first principles controlling the extraction of information from such data and to see if a more effective approach could be found to the analysis problem.

Briefly stated, in the immediate past, the general approach for the large body of effort in the field at large has been to try to correct the data of a new data set for the confounding observational and measurement factors that arise in the data collection process. This has turned out to be a quite daunting and perhaps never ending challenge. Further, the central idea has apparently been to do the correction so that the new data set can be compared to existing spectral reflectance curves for each material of interest. The limitations of this approach are that a) rarely can the observational parameters, e.g. solar and atmospheric effects, non-lambertian surface characteristics, etc. be measured to adequate precision on a pixel by pixel basis as would be required, and b) the highly dynamic nature of the reflectance of Earth surface classes is easy to underestimate, such that the existence of such standard spectral responses (spectral signatures?) which are adequately stable from time to time and place to place is called into question. Further, c) such approaches tend to be based on single spectral curves to characterize a given target material. This ignores significant aspects of spectral responses which are quite diagnostic in nature.

AN ALTERNATE APPROACH

Rather than approaching the problem by trying to adjust a new data set to previous standards, we are focusing on learning how to model the classes of interest within the data set itself to adequate precision using information which is likely to be available to an analyst at the time the analyst begins the analysis process. In the process of pursuing this alternate

method, much fundamental knowledge has been acquired about hyperspectral data in recent years[1]. The work, approached from the standpoint of signal theory as studied in signal processing engineering, revolves around viewing the data of each pixel as a point in an n-dimensional signal space, where n initially corresponds to the number of bands in the data. This signal space is referred to as feature space because as processing proceeds, linear transformations may be carried out on the data, turning the dimensions of the space into more focused spectral features which can be used in discriminating between classes of interest.

Example specific characteristics of high dimensional feature spaces which are especially relevant to the task at hand are,

1. The geometry of high dimensional feature spaces is very different from conventional three-dimensional geometry.

For example, The fraction of the volume of a hypersphere inscribed in a hypercube is given by:

$$f_{d1} = \frac{V_s(r)}{V_c(r)} = \frac{\pi^{d/2}}{d2^{d-1}\Gamma(d/2)} \qquad (1)$$

where d is the number of dimensions. It can be readily verified that f_{d1} decreases rapidly as the dimensionality increases. Figure 1 shows that, while nearly 80% of the volume of the cube is contained in the hypersphere for d = 2, the percentage is reduced to less than 5% by d = 7. Note that $\lim_{d\to\infty} f_{d1} = 0$ which implies that the volume of the hypercube is increasingly concentrated in the corners as d increases.

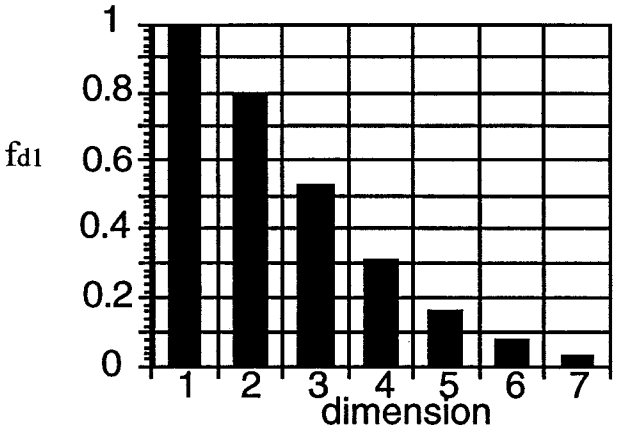

Figure 1. The ratio of the volume of a hypersphere to that of a hypercube of the same size, as a function of the dimensionality of the hyperspace. This illustrates the fact that most of the volume of the hypercube is in the corners, outside of the hypersphere.

The volume available in such higher dimensional spaces becomes enormous as dimensionality grows, and, for any given practical data set, hyperspace is thus mostly empty. As a result, it is likely that any data set to be analyzed will be sparse, not functioning as a dense cloud, but rather as a scattering of points distributed sparsely over a subregion of the feature space. As a practical matter, the positive impact of this characteristic is that, since pixels are rarely precisely coincident in such a high volume space, any subclass of pixels can be discriminated from another subclass if the subclasses can be quantitatively described to adequate precision.

2. As a direct impact of this extraordinary discrimination potential, attention must be focused upon the precision and completeness of the quantitative specification of the classes. Use of a single spectral curve or the average of a number of such curves, while it may be useful in some cases, is clearly limiting in terms of the ultimate performance possible.

To illustrate this point, there are available in the literature a number of "statistical distance" measures for the purpose of measuring the separability of two classes (distributions in feature space). Such distance measures measure the statistical distance between two distributions of points in the N-dimensional space. One with particularly good characteristics for this purpose is the Bhattacharyya Distance. This distance measure bears a nearly linear, nearly one-to-one relationship with classification accuracy. In parametric form it is expressed as follows.

$$B = \frac{1}{8}[\mu_1-\mu_2]^T\left[\frac{\Sigma_1+\Sigma_2}{2}\right]^{-1}[\mu_1-\mu_2] + \frac{1}{2}\mathrm{Ln}\frac{\left|\frac{1}{2}[\Sigma_1+\Sigma_2]\right|}{\sqrt{|\Sigma_1|\,|\Sigma_2|}} \qquad (2)$$

where μ_i is the mean vector for class i and Σ_i is the corresponding class covariance matrix. Examining this equation, one sees that the first term on the right indicates the part of the net class separability due to the difference in mean values of the two classes, while the second term indicates the portion of the total separability due to the class covariances. This makes clear in a quantitative fashion what the relationship is between first order variations (the first term on the right) and second order variations (the second term on the right). This illustrates, for example, that two classes can have the same mean value, and still be quite separable. Note that methods which are deterministically based can only make use of separability measured by the first term. Second order statistics of a class, providing information

about the shape of a class distribution takes on added significance as the feature space dimensionality increases, however, the challenge that this raises is that second order statistics are more difficult to estimate accurately than are first order ones.

3. It is characteristic of the remote sensing problem that one is extrapolating from a very limited knowledge of a scene to a detailed, usually pixel by pixel mapping of the themes of interest in the scene.

This, then, forms the basis of one of the key problems that must be solved in devising a practical and effective analysis process: what form of preliminary information will the analyst have about the scene at the outset, and how can this preliminary information be turned into an adequately precise quantitative description of the classes one desires to identify. For example, one such means for utilizing analyst preliminary knowledge, not requiring an on-site visit, to define more precisely descriptions of the classes of interest is illustrated below. In this case, the analyst preliminary knowledge was in the form of spectral absorption features of imaging spectroscopy. Use of this information to define a larger set of training samples, instead of attempting to use them directly, led to a higher performance classification and less dependence upon a very high signal-to-noise ratio.

4. Other characteristics of high dimensional feature spaces which are especially relevant have been identified. It is known, for example, that because of the high volume of such spaces and the fact that they are thus mostly empty, the significant structure of data subsets of interest lie in a lower dimensional subspace, but one which varies from data set to data set.

This led to a search for good feature extraction schemes, linear transformations which find the subspace in which the classes of interest can best be separated. Based upon past work, a number of such algorithms have been identified, with families of characteristics which reasonably well span the space of circumstances likely to be encountered. In addition to the well-known discriminant analysis algorithm, these include decision boundary feature extraction[2,3,4] and projection pursuit.[5]

5. The most logical way of quantitatively defining the classes of interest in a given analysis situation is via design or training samples drawn directly from the data set to be analyzed.

Doing so, in effect, normalizes for sensor and observation variables which exist in the scene at the time of observation. This fact substantially reduces the need to make data adjustments based upon individual sensor and scene conditions, thus significantly reducing the overall complexity of the processing, making the analysis more robust and easier for the user to carry out.

6. As the dimensionality of the feature space increases, the number of such samples needed to adequately characterize the classes rises very rapidly.

Again, because the number of such prelabeled samples is likely to be small compared to the need, the additional factor of the classification algorithm complexity comes into play. It has been found that simpler algorithms out perform more complex ones when the number of samples available is too limited. An algorithms to define the correct degree of algorithm complexity, called LOOC[6], has been defined to assist with this process, and an algorithm to mitigate the small sample size effects has been created[7].

Thus, a number of apparently practical and useful algorithms now are in hand which should significantly contribute to the analysis process. In addition to better understanding and improving the effectiveness of these algorithms, key problems which remain are the molding of these algorithms into step by step procedures for the various practical circumstances likely to arise, and the specification of these procedures to potential user scientists and other remote sensing practitioners.

In many ways, delivering these procedures to the Earth scientist and other remote sensing practitioners of the user community, procedures which are to these users perhaps much less intuitive than those which have been the focus of attention of the field for some years, may well be the most difficult part of the overall engineering problem of devising and delivering maximally effective hyperspectral data analysis schemes. It is this fact that motivated the creation of MultiSpec©.

MultiSpec is an application program for personal computers which initially was given a basic multispectral data analysis capability. As new algorithms emerged from the research, they have been incorporated into the program, and new versions of it made available to the community without charge. In this way, new algorithms, which may be quite complex to implement may be tried by users with a minimum of effort on their part. The program, together with substantial documentation, is available for download from the world wide web by anyone interested. The URL for the MultiSpec web site is

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

Some of the algorithms mentioned above which it now contains are, Discriminate Analysis Feature

Extraction (DAFE), Decision Boundary Feature Extraction (DBFE), LOOC, and Statistics Enhancement. Additional ones resulting from the research but not described here are included as well.

AN EXAMPLE

We conclude with a single early example achieved by following this line of analysis[8,9]. The data for this analysis was collected by the AVIRIS sensor in 1992 over the frequently studied Cuprite Nevada Test Site. The principle focus for the analysis task was several minerals of geologic interest, and the basis for class definition was known absorption bands for these minerals in the 2.0-2.35 μm region. A conventional approach would use the spectral shape of these absorption features directly to achieve a final labeling of the pixels in the scene. This would necessitate the accurate calibration of the data from radiance to reflectance and the removal of the effect of the atmosphere. It also precludes the use of the discrimination power of second order statistics.

However, in the present case, rather than such radiometric adjustments to the data, itself, a log residue transformation which provides a more crude form of the radiance to reflectance adjustment was used in an interactive mode. Though not as precise as an actual radiance/reflectance/atmospheric removal procedure, it was accurate enough to label a significant number of each desired class, thus allowing for the evaluation of both first and second order statistics of each. The classification which resulted is shown in Figure 2 compared with a conventional geologic reconnaissance map drawn based upon a ground survey. It is seen that a final product with more detail and including discrimination between several additional minerals was possible by this means. Further details are contained in the referenced documents.

An additional example in an agricultural context is available for downloading from the MultiSpec web site.

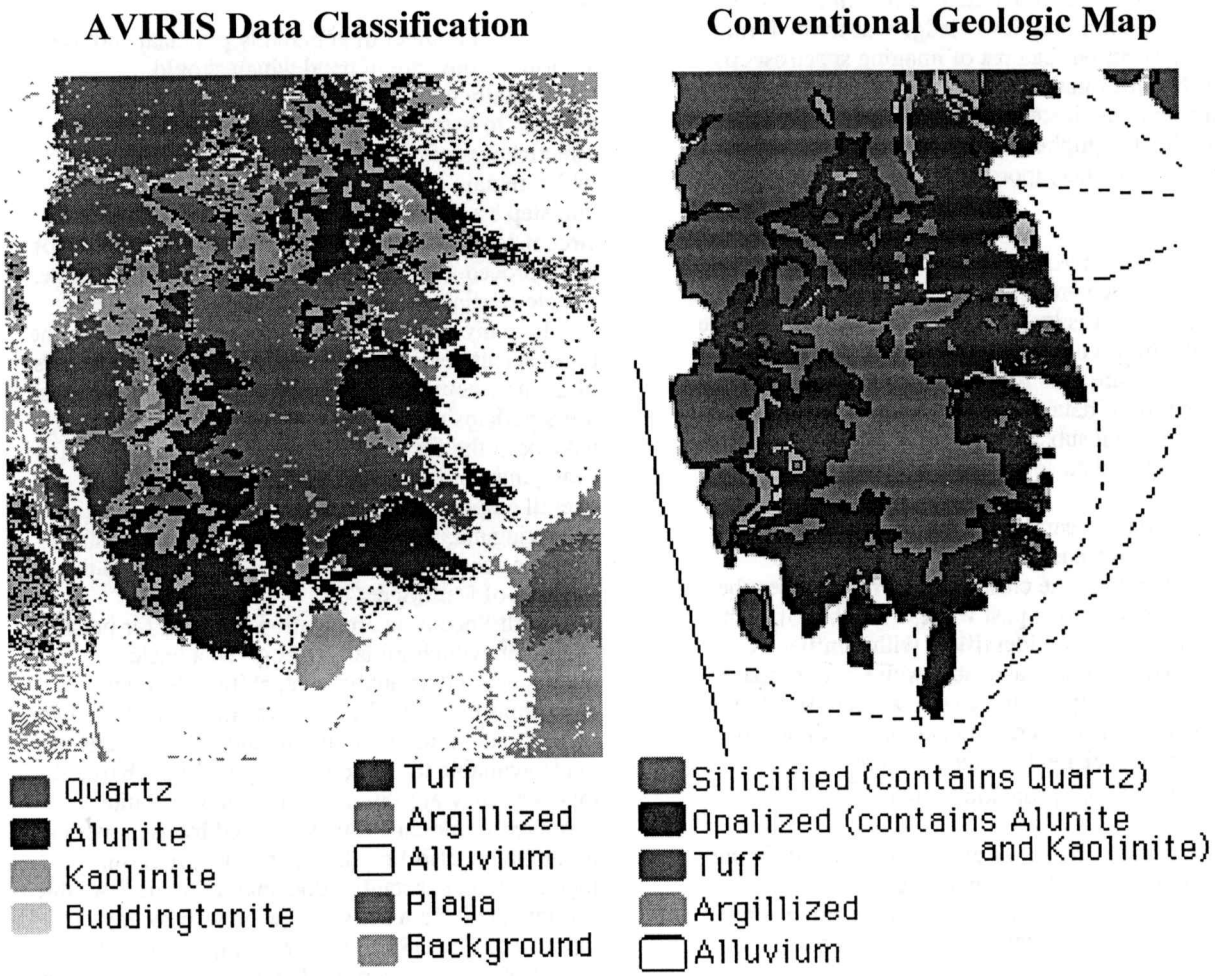

Figure 2. Comparison of a classification with a conventionally drawn geologic map of the Cuprite Nevada test site using 1992 AVIRIS data. (Original in Color).

ACKNOWLEDGMENT

Work related to this paper was supported in part by NASA under Grant NAGW5-3975.

REFERENCES

1 See, for example, Chulhee Lee and David A. Landgrebe, "Analyzing High Dimensional Multispectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 31, No. 4, pp. 792-800, July, 1993, David A. Landgrebe, "On Progress Toward Information Extraction Methods for Hyperspectral Data," SPIE 42nd Annual Meeting, San Diego, CA, 27 July-1 August 1997, and Jimenez, Luis, and David Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," *IEEE Transactions on System, Man, and Cybernetics*,To appear in Volume 28 Part C Number 1, Feb. 1998. Copies of these and other works are available for downloading from http://dynamo.ecn.purdue.edu/~landgreb/publications.html

2 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.

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

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

5 Luis O. Jimenez, "High Dimensional Feature Reduction Via Projection Pursuit," PhD Thesis, Purdue University, May 1996, also available as 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.

6 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.

7 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.

8 Hoffbeck, Joseph P. and David A. Landgrebe, "Classification of Remote Sensing Images having High Spectral Resolution," *Remote Sensing of Environment*, Vol. 57, No. 3, pp. 119-126, September 1996.

9 Joseph P. Hoffbeck, "Classification of High Dimensional Multispectral Data," PhD Thesis, Purdue University, May 1995, also available as Joseph Hoffbeck and David A. Landgrebe, "Classification of High Dimensional Multispectral data," School of Electrical Engineering Technical Report TR-EE-95-4, May, 1995.