Toward a maximally effective means for analysis of hyperspectral data

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

Toward a maximally effective means for analysis of hyperspectral data[1]

David Landgrebe[2]
Purdue University School of Electrical & Computer Engineering
West Lafayette IN USA 47907-1285

ABSTRACT

In this paper we describe efforts toward a hyperspectral land remote sensing data analysis procedure that would be maximally effective for use by a broad community of future users. Though there would be dependence of performance achieved on the spectral subspace and the classification algorithm used, the major dependence is on how well the user quantitatively defines the classes desired. Thus the attempt is to measure this dependence for typical users and to introduce means that mitigate the problem of class and training definition.

Key words: Hyperspectral Analysis, Training methods, Class definition

1. INTRODUCTION

Much of the effort on land remote sensing in past decades has been largely science research oriented. However, the ultimate value of this technology is thought to be in the economic impact it could have and thus the potential is very widespread, reaching well beyond the science community. We begin by exploring this reach, by analogy with other space-based activities, before centering on the task of specifying and devising a suitable interface for the technology with this much larger, beyond-science potential community.

Efforts to learn to use operational space-based capabilities for remote sensing began in earnest in the 1960's. This was stimulated in large part by the launch of the Soviet Union's Sputnik satellite in 1957, which had this effect on a range of potential uses of spacecraft capabilities. A 1967 study by the U.S. National Research Council "on the probable future usefulness of satellites in practical Earth-oriented applications" identified a number of potential applications for space technology. Among these were applications in forestry, agriculture, geography, hydrology and geology in the land areas; meteorology and oceanography; several types of communication such as broadcast and point-to-point communication; and navigation and traffic control. Today there is widespread routine daily reliance on such technology in all of these applications except those related to the land.

There are two primary elements that are missing that made land remote sensing applications slower to develop than the others. The first is the availability of data under user-defined conditions. These required conditions include data being collected upon a user's request, with suitable specifications, at no or low cost similar to weather data, and a belief that such data will be available on these terms in the future. Today if a potential user senses the need for land remote sensing data, the process is usually one of looking through various archival sources to see if reasonably acceptable data already exists. Very probably one must substantially compromise on ones need, as data is unlikely to have been gathered over the needed site at a time not too different than when required. Then there is the matter of cost. It also seems likely for potential users to doubt in the present circumstances that when the next need arises, user data will again be available. None of these limitations exist for weather, communication or navigation data.

One is inclined to assume that such limitations will eventually be removed, and a community of users will build large enough to cause the cost and availability factors to be mitigated. However, these limitations also impact the slow development of the second missing element.

---

[1] An invited paper prepared for SPIE Conference 4541, Image and Signal Processing for Remote Sensing VII, Toulouse France, 21 Sept 2001

[2] landgreb@ecn.purdue.edu

The second missing element for land remote sensing application technology is a suitable means for the user (as compared to the remote sensing specialist) to analyze the data. What is needed is an analysis method which does not require highly specialized signal processing engineering knowledge but which produces optimal results in terms of accuracy on a computer that is not prohibitively expensive and done so in a reasonably short analyst time. It is to this objective that the current work is directed.

2. SENSOR SYSTEM DESIGNER CHOICES

First, some comments about the sensor system and specifications for the needed data that is to be analyzed. The fundamental premise of such information systems is that the electromagnetic energy field arising from the Earth's surface contains potentially useful information in its spectral, spatial, and temporal variations. The question that the sensor system designer must resolve is how shall these variations be measured? They must be divided into spatial pixels, but what size, and they must be divided into wavelength bands but how many, located were, and what width. And given these choices and the finite magnitude of the energy field, what signal-to-noise ratio will result, thus defining the signal-to-noise ratio and therefore the measurement precision. Five hundred meter pixels are too coarse and one meter pixels too fine for an agricultural crop species survey, for example, and seven spectral bands are only marginally enough for many urban and natural land surveys. At least 50 bands covering the solar reflective and thermal regions would be adequate to solve a large number of land resource problems. In the following, we will assume something in that range and with signal-to-noise ratios justifying a 10 or more bit dynamic range.

3. ANALYST CHOICES

The analyst ordinarily has a great many choices to make in analyzing the data. Some of the key ones are,

- How subtle are the discriminations to be made?
- What is the optimum spectral dimensionality to be used in the classifier, and how is the optimum feature subspace to be determined?
- What level of complexity of classification algorithm should be used?
- How will the quantitative description of the desired classes be determined?

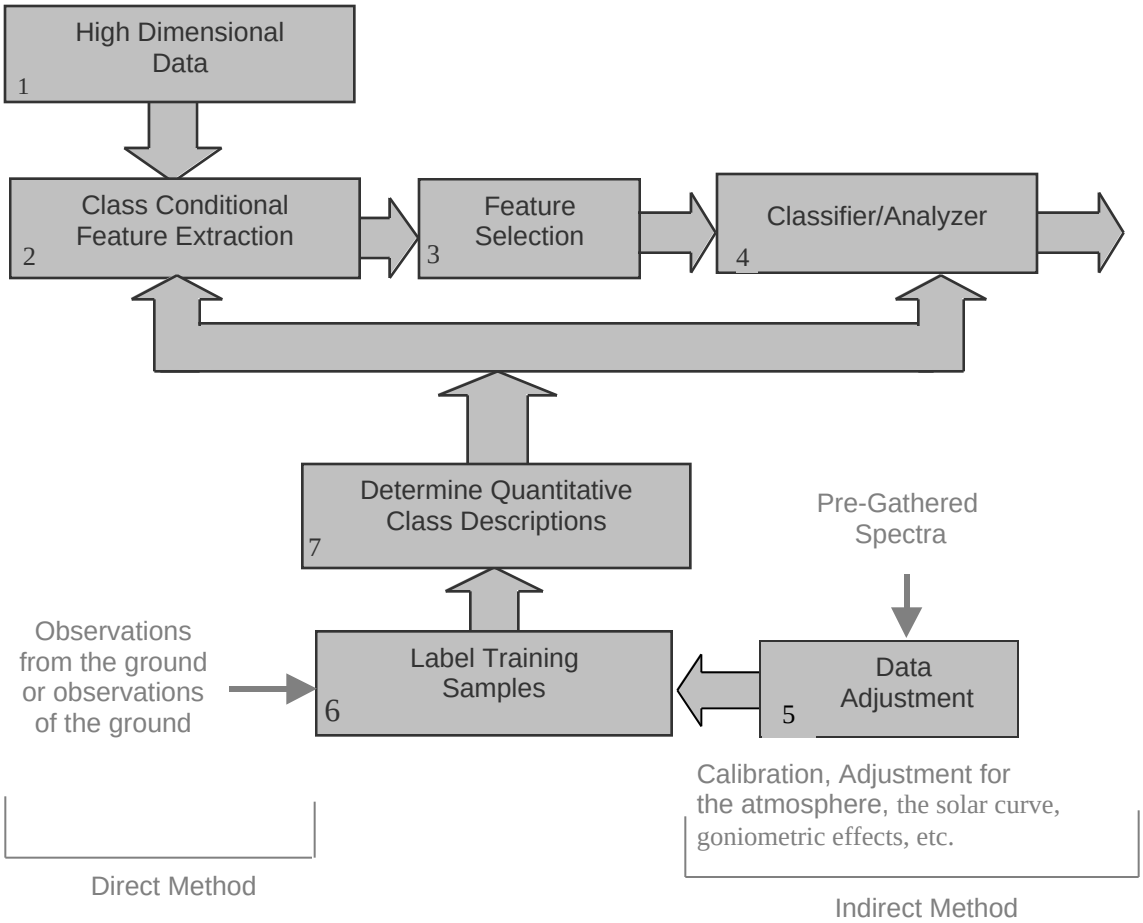

Figure 1: Schematic layout of a data analysis procedure.

Figure 1 shows a schematic layout of one possible, reasonably robust data analysis sequence. It shows the data (1)[3], which might be 50 to 200 dimensional at the outset, is to proceed through a feature extraction process (2) to define candidate optimal subspaces based upon adequately precise statistical descriptions of the user-specified classes desired. One or more feature extraction algorithms might be used to generate subspaces of perhaps 5 to 30 dimensions that are optimized for the particular set of classes desired. Then the analyst might (3) select the specific optimized feature subset based on the degree of subtly of the classes and the precision with which the analyst has been able to quantify the class descriptions. Next and again based upon these same two factors, the analyst must choose the classifier algorithm to use (4).

The most significant and sensitive among all the analyst activities usually is the matter of how to specify in a quantitative fashion and with adequate precision just what information is desired from the data. Obviously, one cannot expect accurate output if one does not provide complete and accurate input. Some would have this process start with pre-gathered spectra. In this case, the next step would necessarily be (5) making all needed data adjustments in order to reconcile measurement circumstances for the data at hand with those present at the time the pre-gathered data were taken. These adjusted data could then be used to label training samples (6) in the data set to be analyzed. Doing this, rather than attempting to use the adjusted data directly for the analysis, would help to normalize for any remaining conditions not accounted for in the data adjustment process or inaccuracies that might have resulted from that process.

This procedure using data adjustment is referred to as the indirect method. It is problematic because adequately precise data needed to make the data adjustments to a useful level of precision are not usually available. It is also a complex and lengthy process. When possible, a more powerful method is the direct method indicated above. In this case one uses observations from the ground or at least of the ground, perhaps from imagery of the scene, to label training samples. No data adjustment processing would be needed in this case.

By whatever the means, defining a suitably exhaustive list of classes by labeling an adequate number of training samples is the key step to a successful analysis, and it is one of the most time-consuming and onerous parts of the process. Further, usually it is difficult to label a large enough number of training samples, especially for high dimensional cases, which are more powerful but more sensitive to estimation error due to inadequate numbers of samples. There are a number of algorithms (7) that can be used to mitigate this limitation, so that more adequately precise and representative class statistics can be obtained. The class models thus obtained would be used to do the feature extraction and classification processes.

This process has been found to be very effective for a large number of circumstances of dimensionality and a wide variety of data and classes. In preparing a data analysis process for wide use, an additional parameter that must be considered is the anticipated skill level of the analyst and process vulnerability to variations in this skill level. Can a variety of analysts obtain reasonably consistent results? To test this issue the following experiment was run[4]. A set of 12 two-person teams of analysts was given the same problem and data set. The data consisted of 208,000 pixels of 12 band data from an agricultural scene. This was the first data set any had analyzed. They were each required to produce classifications with at least the following classes: *corn, oats, soybeans, wheat, and forage/hay.* There was an extensive amount of ground truth available for the site, but to simulate the effect of limited knowledge of the ground scene, as would ordinarily be the case, each team was limited to 1000 total training samples distributed over the classes and subclasses the team chose to use. All teams were to measure performance against a fixed 70,000-pixel test set

In order to test the effect of various training sets on classification algorithms of varying complexity, the classification algorithms to be used were the

- *Minimum Distance Classifier,* which uses only the individual class mean vectors, the
- *Fisher Linear Discriminate,* which uses the individual class mean vectors plus a covariance matrix common to all classes, the
- *Quadratic Classifier,* which uses individual class mean vectors and individual class covariance matrices, and
- *ECHO*[5,6], which, in addition to class mean and covariance matrices, uses spatial information.

[3] Numbers in parentheses refer to the number in the blocks of the diagram

[4] David Landgrebe, "On the Relationship Between Class Definition Precision and Classification Accuracy in Hyperspectral Analysis," International Geoscience and Remote Sensing Symposium, Honolulu, Hawaii, 24-28 July 2000.

[5] R. L. Kettig and D. A. Landgrebe, "Computer Classification of Remotely Sensed Multispectral Image Data by Extraction and Classification of Homogeneous Objects," *IEEE Transactions on Geoscience Electronics*, Volume GE-14, No. 1, pp. 19-26, January 1976.

[6] David Landgrebe, "The Development of a Spectral-Spatial Classifier for Earth Observational Data," *Pattern Recognition*, Vol. 12, No. 3, pp. 165-175,1980.

All of these were implemented as maximum likelihood classifiers.

Further, once having fixed their class and subclass list and the training sets for them, the teams were asked to test several schemes for improving the estimation of class statistics. In addition to standard sample mean and covariance estimation, two algorithms that tend to mitigate the limited design set size problem were used. These were the LOOC covariance estimator[7] and a Statistic Enhancement[8] scheme. The LOOC estimator examines the sample covariance estimates, the common covariance estimate, as well as their diagonal forms, and their mixtures to determine which would be most effective. Though a covariance matrix estimate would ordinarily be singular if fewer than n+1 samples are used, where n is the number of spectral bands, LOOC provides a usable covariance estimate with as few as 3 samples per class, regardless of the number of bands. The Statistics Enhancement scheme uses a sampling of unlabeled samples in addition to the labeled design samples to mitigate the limited design set problem as well as improving the classifier's ability to generalize over the entire data set. All the teams used the algorithms as implemented in MultiSpec, a personal computer based application available to anyone from http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/. The data used in the experiment is also available from that site.

The results of the tests are shown in Table 1 as the average classification accuracy obtained and the standard deviation of the classification accuracy over the 12 teams. It is seen that for the baseline 1000 sample tests, the accuracy steadily improved with classifier complexity. It is significant to note that the consistency of results, as indicated by the standard deviation of the 12 improved as well. For the last two classifiers, which are both quadratic, the classification results were remarkably consistent.

Classifier→ *Training*	Minimum Distance	Fisher Lin. Discrim	Quadratic	ECHO
Baseline	100-200 pixels/class			
Std. Ave	75.1%	86.9%	91.4%	92.8%
St. Dev.	9.1%	4.0%	1.9%	1.8%
LOOC Ave	75.1%	87.0%	92.0%	93.9%
St. Dev.	9.1%	4.0%	2.3%	2.3%
LOOC-Enh.Ave	74.4%	86.9%	91.5%	94.6%
St. Dev.	7.6%	2.9%	2.5%	2.3%

Table 1: Average accuracy and standard deviation results from the 12 teams.

These tests are intended as a step towards defining an analysis procedure that can provide consistent results over a class of perhaps somewhat inexperienced but otherwise knowledgeable users of remote sensing technology. However, there still remains the rather onerous nature of defining the classes, subclasses and selecting training sets for each. Further, the process as describe above is much too complex for the general user community. As a first step towards mitigating these limitations, an additional scheme is under investigation with initial results recently obtained[9]. The concept is to add box 8 as shown in Figure 2. An iterative process would be used in which, after a given classification, the samples classified with higher likelihoods, referred to as semi-labeled samples, would be added to the training data for the class to which each was assigned and the classification process repeated with the enlarged training set.

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

[8] 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.

[9] Qiong Jackson and David Landgrebe, "An Adaptive Classifier Design for High-Dimensional Data Analysis with a Limited Training Data Set," IEEE Transactions on Geoscience and Remote Sensing To appear. See also item B[45] at http://dynamo.ecn.purdue.edu/~landgreb/publications.html

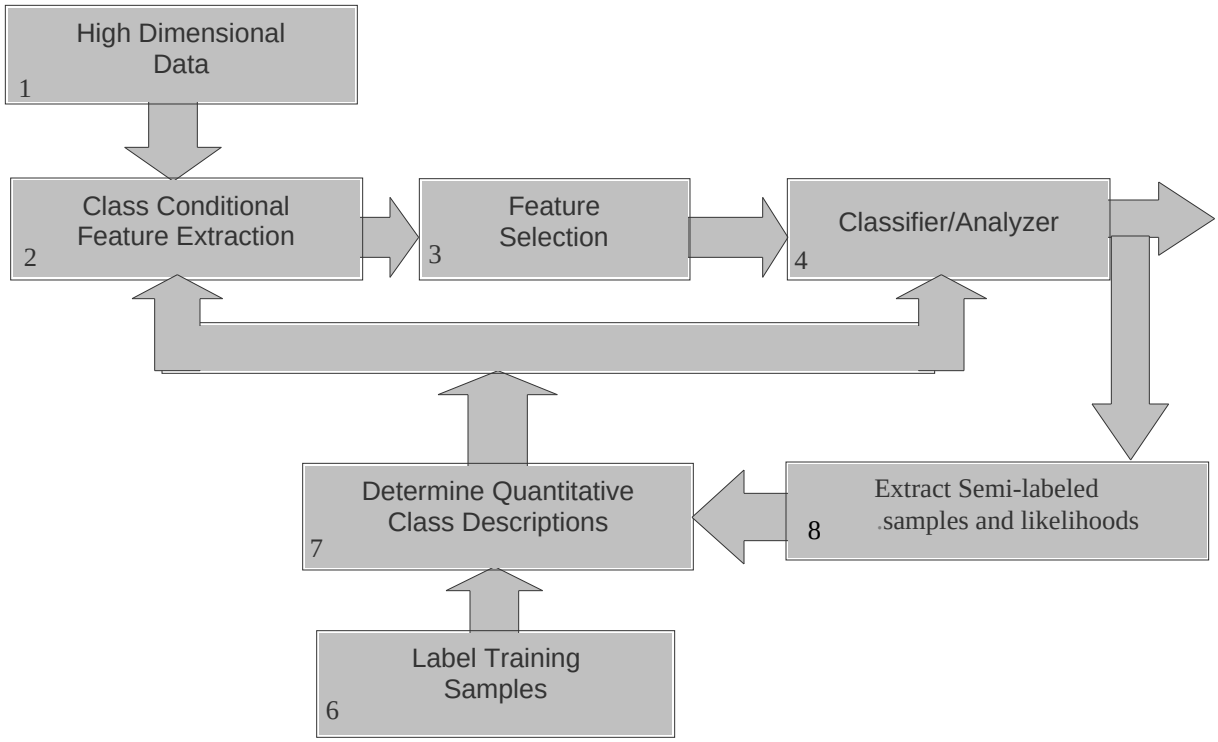

Figure 2: An augmented analysis that may prove to ease the user's problem.

The increase in the size of the training set improves the estimation accuracy of the class statistics and thus the classification accuracy. Figure 3 shows the result of this process in an early trial. This particular analysis was begun with a rather modestly drawn set of training samples, leading to rather modest initial results. However, the iterative process was able to improve the accuracy significantly. The effect from the analyst's point of view is that one can begin with a smaller set of labeled samples that would perhaps be less laborious to identify and as might be drawn by a less experienced analyst, then allow the system to move the classification accuracy toward an optimum value.

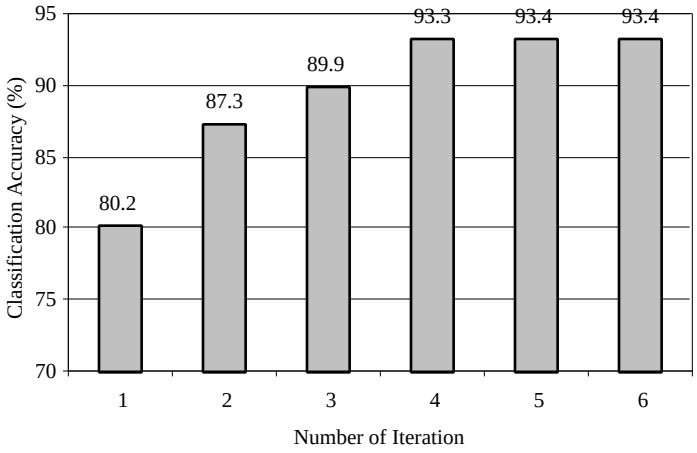

Figure 3: Classification Accuracy vs. iteration number for an example data set

4. CONCLUDING COMMENTS

To be viable to the broader community, there must be a data collection capability that is responsive to the user needs rather than in effect, requiring the user to respond to whatever the data collection process provides, and the system must be available in a manner that instills confidence that it will be available over a longer term as needed. There must also be an analysis scheme that is workable in the various situations that the user requires. It must be easy to use, not requiring a high level of training or understanding of the fundamental concepts of signal processing engineering. It cannot be "automatic" and still be responsive to the nuances of the user requirements. It simply must work as the user needs, based on the user inputs. Though the algorithms applied here may well be useful in such a user system, the process of their application as described here is much too complex to be adopted by a broad spectrum of users who could benefit from the technology. It would be necessary to evolve an implementation utilizing most if not all of the steps used here but implemented in a manner relatively transparent to the user, while at the same time providing acceptable results to the user's analysis problem.

ACKNOWLEDGEMENT.

Work reported in this paper was funded in part by the US Army Research Office under grant DAAH04-96-10444. This sponsorship is gratefully acknowledged.