Spectral Analysis Techniques for Classification of AVIRIS Data

Sylvia S. Shen, Bonnie Y. Trang and Mark T. Noga

Lockheed Palo Alto Research Labs
3251 Hanover St., Palo Alto, CA 94304

ABSTRACT

An advanced software system is described that allows scientists to perform effective and efficient analysis of high spectral resolution data. The principal focus is on new hyper-spectral information handling and extraction algorithms. Among these is an image classification technique which uses a set of diagnostic features derived from the absorption and emission bands of the spectra as discriminants. Another approach uses a new signature matching technique that employs geometric and structural information contained in the spectrum. The strategy here is to represent the spectrum by a tree, so that spectrum correlation can be performed by matching the nodes and branches of the signature and reference trees. The emphasis has been on producing fast spectral analysis techniques which allow all pixels in the scene having a signature similar to that of a specific class to be identified more effectively. Results of applying these techniques to AVIRIS data are also presented, along with a brief description of the current capabilities of our new hyper-spectral image analysis software system.

1. INTRODUCTION

The new generation of airborne and spaceborne sensor technologies offers the potential of producing broadly consistent image data with high spatial, spectral, and temporal resolutions. These types of remotely sensed image data are becoming the primary tool for studying many land and ocean processes. In fact, multispectral instruments such as the Landsat Multi-Spectral Scanner (MSS) and Thematic Mapper (TM), and NOAA's Advanced Very High Resolution Radiometer (AVHRR), have been used in land cover, land use, and meteorological studies for many years. These instruments typically provide data in 3 to 8 spectral bands. A new generation of imaging spectrometers such as JPL's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) has now been designed to disperse reflected solar electromagnetic energy from the earth's surface to form as many as 200 spectral images of the scene simultaneously. Unlike the traditional multispectral instruments, imaging spectrometers provide sufficient spectral resolution to allow unique spectral signatures to be extracted on a per-pixel basis, thus enabling direct surface material identification and more detailed analysis of the biophysical processes. However, the unprecedented volume and complexity of the new data presents a challenge to the traditional image analysis methods and requires that new approaches be developed to efficiently analyze the data.

This paper describes an advanced software system that we have developed to allow scientists to perform effective and efficient analysis of high spectral resolution data, e.g. AVIRIS, using moderate computing resources. The principal focus is on new hyper-spectral information handling and extraction techniques that we have developed. One of these is a hyper-spectral image classification technique which uses a set of diagnostic features derived from the absorption and emission bands of the spectra as discriminants. Since these bands are unique to the composition and structure of the material being analyzed, the derived features contain the key discriminating information for class identification. Another approach represents the spectrum as a tree, so that spectrum correlation can be performed by matching the nodes and branches of the trees, one representing the sample waveform signature and the other representing the signature of a known material (class). These fast spectral analysis techniques allow all pixels in the scene having a signature similar to that of a specific class to be identified more effectively. Results of applying these techniques to AVIRIS data are also presented, along with a brief description of the current capabilities of our hyper-spectral image analysis software system.

2. SPECTRAL ANALYSIS TECHNIQUES

Since the potential value of imaging spectrometers will not be realized without the development of new approaches to hyper-spectral information handling and extraction, the principal focus here is on special analysis techniques for data of high spectral dimensionality such as AVIRIS. AVIRIS is designed to acquire images with 20-m pixels simultaneously in 224 spectral bands covering the 0.4 - 2.5 μm wavelength region. Each pixel has associated with it 224 radiance values, or essentially a continuous radiance spectrum (see Figure 1). The high spectral sampling

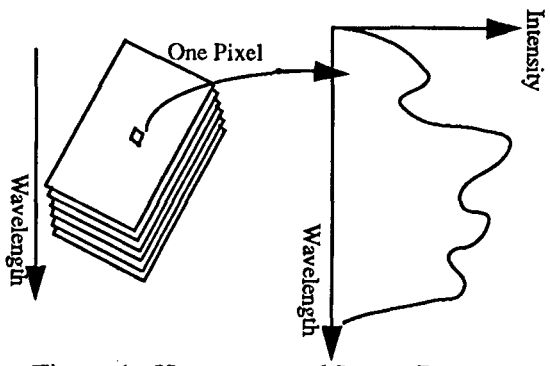

Figure 1. Hyper-spectral Image Data.

rate gives sufficient information to uniquely classify most terrestrial materials. To make effective use of this new class of data, the key step is to find a strategy to reduce dimensionality and at the same time retain sufficient information for class separation and identification. Described below are some of the spectral analysis techniques that we have developed for analyzing high spectral resolution image data.

2.1 Image Classification Based on Emission Band Characteristics

An analysis method that has been used and reported in the literature is based on the principal component transformation [1,2]. This transformation transforms the 224 spectral values into

224 uncorrelated coordinates. The first few coordinates, namely the principal components, are used as attributes for discriminating between classes. These principal components which are linear combinations of the original 224 spectral values, may be effective in separating certain classes. But they might not be optimal for separating certain other classes. Construction of this transformation, requiring on the order of 10^{13} arithmetic operations for a typical 11-km-square AVIRIS scene, can be computationally expensive [3]. Therefore, more efficient methods of extracting discriminant features are ultimately needed.

We have developed a scheme which uses attributes that are more characteristic of the radiance spectrum. Our method [4] uses a set of diagnostic features derived from the absorption and emission bands of the spectra. Because a material is specified by its composition and structure, which determine the specific energies at which the electronic excitations and molecular vibrations occur, each material has a characteristic set of excitation energies. When radiation at the appropriate energies penetrates into the material, it is absorbed by the excitation process, resulting in a loss of radiation at those specific excitation energies. These absorption bands are therefore diagnostic signatures of the material. Emission peaks are similarly diagnostic of the composition and structure of materials, although the signature phenomenology is somewhat different than that for the absorption bands. Emission peaks are typically the by-product of excitations. Excitation is followed by relaxation back to the ground state, this relaxation corresponds to a loss of energy by the excited species as it returns to its lower ground state configuration. This loss of energy takes the form of radiation emission at the specific energy of relaxation. Just as with the absorption bands, the emission peaks are also diagnostic signatures of the material.

Our method first extracts the positions, widths and strengths of the emission peaks from the spectra, and then uses them as attributes for class separation. Since these bands are unique to the composition and structure of the material being analyzed, the derived features contain the key discriminating information for class identification. A nonparametric classifier has also been developed based on these attributes for identifying the various classes in the scene. Training samples or prototype samples can be specified from the image data themselves or from a spectral reference library. Similar diagnostic features can be extracted from the absorption bands and used as discriminants for class separation and identification [4]. Furthermore, the combined set of features derived from both the emission and absorption bands can also be used as discriminants [4].

2.2. Spectrum Correlation Via Tree Matching

An alternative approach to pixel classification is to correlate the radiance spectrum of a given pixel with the library spectra of known classes. Matching by cross-correlation can often be very time-consuming. Therefore a fast spectral signature matching technique is needed. A binary spectral encoding algorithm was developed by Mazer et al [5] which uses a binary code vector constructed from the spectral amplitude and the local slope to represent the spectrum. It then uses Hamming distance as the similarity measure for signature matches. This type of signature matching techniques can be used to find all the pixels in the scene having a spectral signature similar to that of a specific class. Our strategy is to represent the spectrum by a more efficient data structure from a processing standpoint, namely a tree (see Figure 2), so that spectrum correlation can be performed by matching the nodes and branches of the trees. A tree, if properly de-

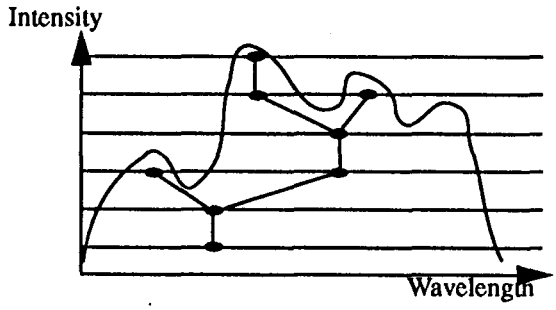

Figure 2. Tree Representation.

signed, can describe not only the succession of the peaks and the valleys of the spectrum, but also the relative height of these entities, and the self-embedded structure of the spectrum. We have developed a scheme to rapidly construct such trees from the pixel spectra. We have also developed an efficient recursive tree matching algorithm [6] based on a dissimilarity or distance measure as described below.

The tree matching approach can be best described as a recursive method that traverses two trees simultaneously looking for differences in the number and nodes and branches between the two trees. Signature and reference trees are first converted to general lists (lists of lists) because this data structure minimizes the number of pointers required to store any tree with an arbitrary number of branches per level (see Figure 3). The algorithm is complicated by the fact that, in

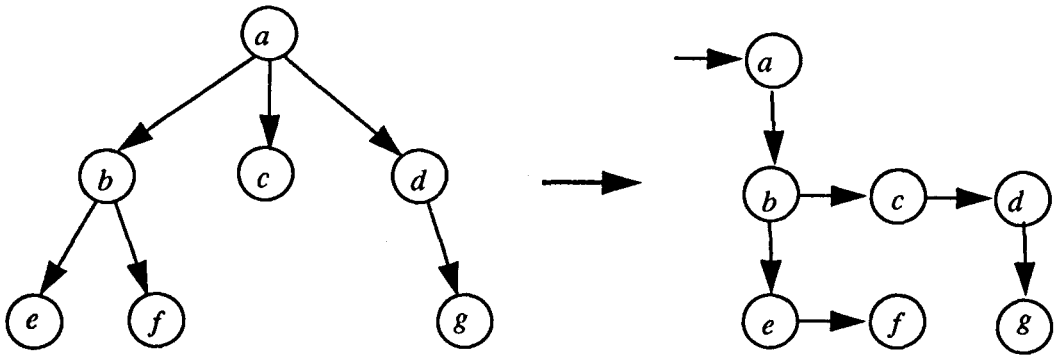

Figure 3. Tree to list mapping.

general, it is not possible to determine a priori how sublists should be paired (see Figure 4). Our approach, which is reasonable with respect to the intended application, is to compare all sublist combinations maintaining a left-to-right ordering. Fortunately, the recursion keeps track of all possible combinations of branch/node comparisons regardless of level. Before the traversal begins, the difference D is set to zero. When two trees differ at any level, a weighted value is added to D. At the end of the procedure, the minimal D is the value reported by the algorithm (see Figure 5). Definitions of the various functions used in the algorithm described in Figure 5 are

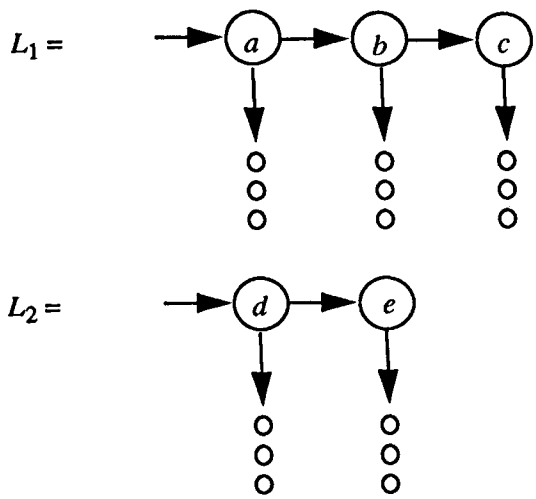

Figure 4. Should the sublists associated with the combinations (a,d) and (b,e) be paired, or should (b,d) and (c,e) be paired, etc.?

```
DIFF(L1,L2):
    if (IS_EMPTY(L1) and IS_EMPTY(L2)) return (0);
    else if (IS_EMPTY(L1)) return SIZE(L2);
    else if (IS_EMPTY(L2)) return SIZE(L1);
    else return (MIN3(DIFF(SUBS(L1),SUBS(L2)) + DIFF(TAIL(L1),TAIL(L2)),
                      1 + SIZE(SUBS(L1)) + DIFF(TAIL(L1),L2),
                      1 + SIZE(SUBS(L2)) + DIFF(L1,TAIL(L2)));
```

Figure 5. Tree difference algorithm, weighting factor is one per missing node.

as follows: The TAIL of a list is the list to the right of the head node, the SUBS of a list is the list below the head node, and the SIZE of a list is the total number of nodes in the list.

Again, the prototype spectra of the classes of interest can be specified from the image data themselves or from a spectral reference library. This technique is very efficient and can, much like the previous approaches, be used to search the entire scene for occurrences of similar spectra to those in the on-line spectral library of candidate materials.

3. RESULTS

We have applied both techniques to the 1987 AVIRIS data over Moffett Field area. Both algorithms have successfully identified 12 surface materials including asphalt, concrete, water, salt pond, grass, tree and shrubbery, soil, marsh, etc. The left image in Figure 6 shows the true RGB composite of the Moffett scene and the right image is the color coded classification results using the automated algorithm based on emission band characteristics as described in Section 2.1.

Highly comparable classification results have also been obtained using the tree matching technique. These algorithms can also be used for vegetation identification, mineral classification, and change detection. As an example, we have begun to scan the 1989 AVIRIS data over Jasper Ridge Biological Reserve for vegetation species using these techniques.

4. CURRENT SOFTWARE SYSTEM CAPABILITY

Effective use of new classes of data with high spectral dimensionality has motivated the development of a highly interactive software system that allows users to visually interact with the hyper-spectral data and their statistics. We have developed and implemented a variety of display and processing routines using low cost computer workstations. These routines perform functions such as image display, spectral curve plotting, histogramming, contrast adjustment, statistics calculation, zoom, thresholding, color coding, etc. Display and processing routines have been integrated with the special algorithms described in Section 2 to provide a wide range of capabilities to users for extracting information contained in the data.

The current system capabilities include rapid display of projection of hyper-spectral images in the spectral direction, cursor-designated spectral plots of single pixels or averages over a spatial window, the ability to generate, delete and update library (reference) radiance curves from specified regions in the scene, and display of spectra from a spectral library for visual comparison. There are also functions which perform linear, logarithmic and inverse logarithmic contrast enhancement, subwindow zoom, extraction of key discriminant features from radiance curves as described in Section 2.1, rapid unsupervised organization of pixels having similar spectral signatures (this enables a better understanding of the natural clustering and structure of the data), search of all the pixels in the scene having a spectral signature similar to a specified library spectrum for efficient scene classification using the signature matching techniques (e.g. tree matching, binary encoding, cross correlation, emission and absorption band characteristics, etc.) as described in Section 2 and portrayal of those pixels in a unique color. Color-map adjustment functions include designation of color codes for classes of interest, and display of color coded classification result, false color image and true color composite of the red, green and blue channels of the image data. Figure 7 shows a snapshot of a current version of our software system in operation.

5. CONCLUSIONS

Our advanced software system incorporates a wide variety of special algorithms and a large collection of display and processing tools to allow scientists to perform detailed analysis of data with high spectral dimensionality. The special hyper-spectral techniques can differentiate object substances that look identical to an analyst using conventional tools. More specifically, our techniques identify an object by detecting its absorption and emission band characteristics. Since every substance has a unique set of such bands, our techniques are more robust than the conventional methods. The display and processing tools allow the users to visually interact with the data and their statistics. This software system is highly interactive, maintainable, and extensible. All software is written in the C language under UNIX. The interactive portion of

the software is written for Silicon Graphics Iris 4D Workstations and their windowing environment. An X port is planned that would make the software usable from a number of different workstation platforms including those from Sun Microsystems. Implementation has been carried out in such a way as to make additions and modifications easy, while at the same time maintaining a user friendly application environment. So far, we have demonstrated that our software system is capable of extracting and identifying surface materials from AVIRIS-like data sets. The system has also been tested with synthetic high spectral resolution radiance data in a phenomenology study.

6. REFERENCES

[1] S. C. Feldman and J. V. Taranik, "Comparison of Techniques for Discriminating Hydrothermal Alteration Minerals with Airborne Imaging Spectrometers Data", Remote Sensing of Environment 24 (1988).

[2] A. Hutsinpiller, "Discrimination of Hydrothermal Alteration Mineral Assemblages at Virginia City, Nevada, Using the Airborne Imaging Spectrometer", Remote Sensing of Environment 24 (1988).

[3] G. Vane and A. F. H. Goetz, "Terrestrial Imaging Spectroscopy", Remote Sensing of Environment 24 (1988).

[4] S. S. Shen and B. Y. Trang, "A Feature Selection Technique for Classification of Hyper-Spectral AVIRIS Data", Proceedings of SPIE on Applications of Digital Image Processing, Vol. 1567 (1991).

[5] A. S. Mazer, M. Martin, M. Lee and J. E. Solomon, "Image Processing Software for Imaging Spectrometry Data Analysis", Remote Sensing of Environment 24 (1988).

[6] S. S. Shen, B. Y. Trang and M. T. Noga, "Hyper-Spectral Signature Correlation Via Tree Matching", in preparation.

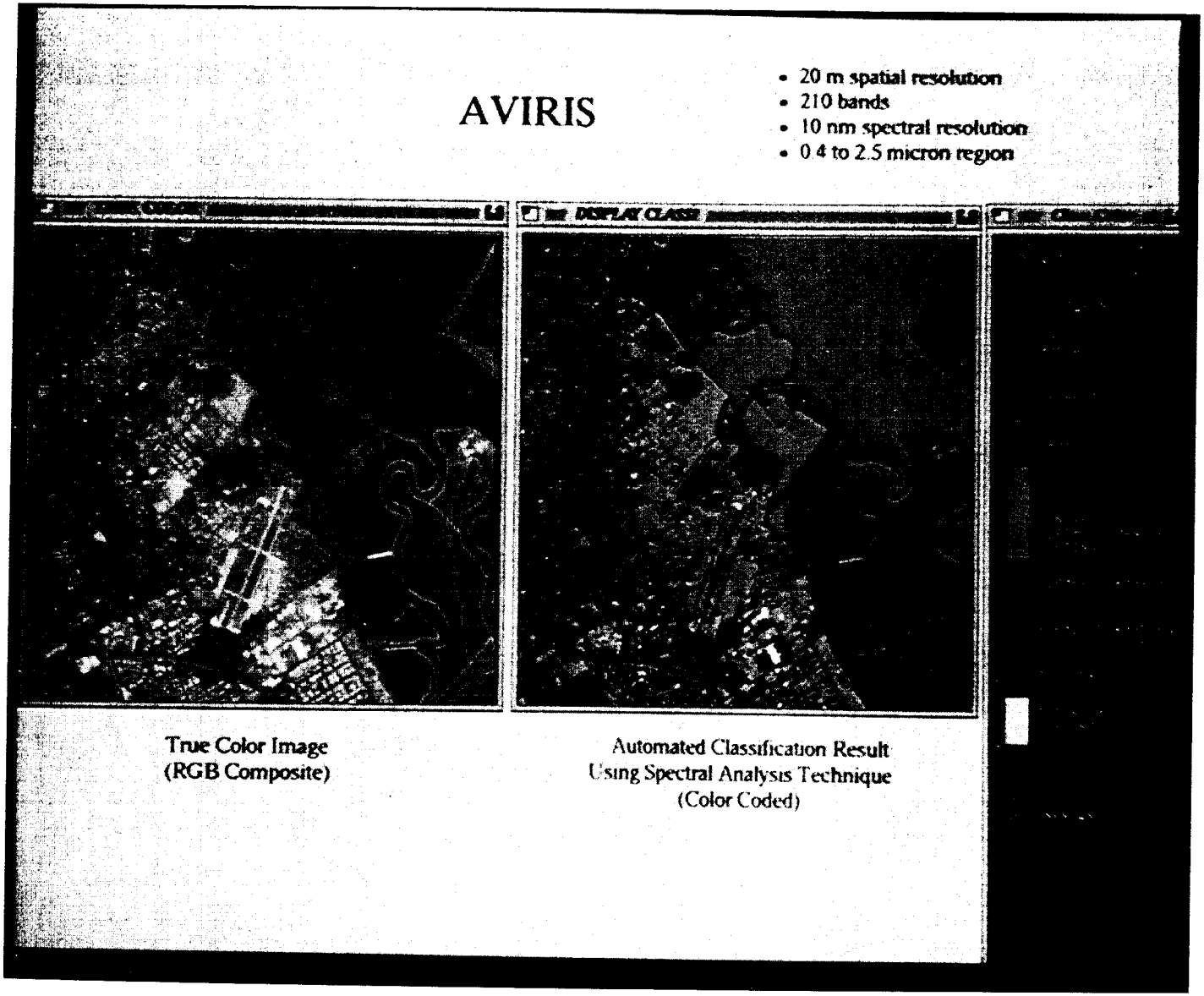

Figure 6: Automated Classification Results [see slide 13]

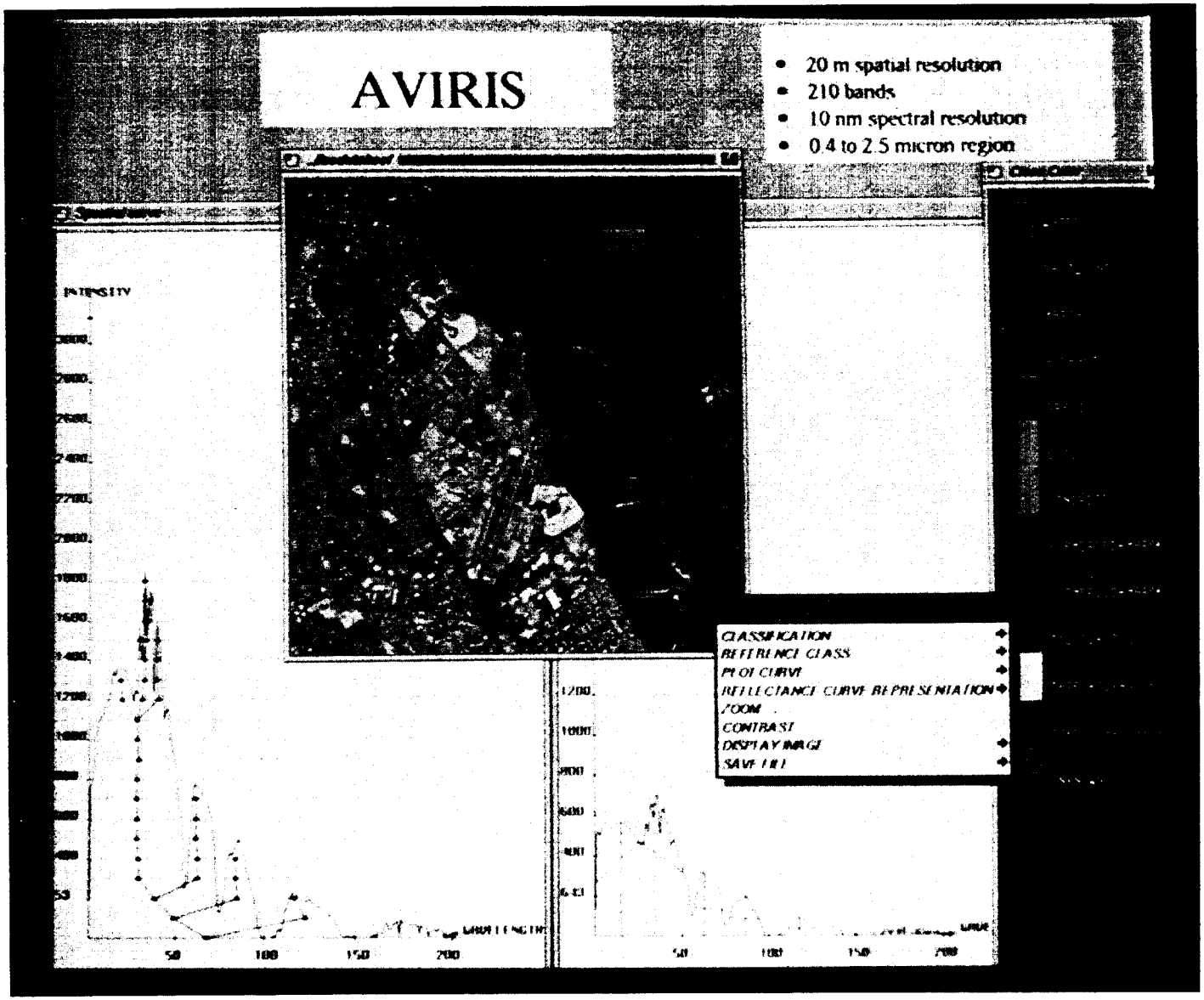

Figure 7: Advanced Hyper-Spectral Image Analysis Software System [see slide 14]