MULTISPEC - A TOOL FOR MULTISPECTRAL-HYPERSPECTRAL IMAGE DATA ANALYSIS*

Larry Biehl[1] and David Landgrebe[2]

[1] Research engineer, School of Electrical & Computer Engineering
[2] Professor, School of Electrical & Computer Engineering

Purdue University, West Lafayette IN 47907-1285 USA

ABSTRACT

MultiSpec** is a multispectral image data analysis software application. It is intended to provide a fast, easy-to-use means for analysis of multispectral image data, such as that from the Landsat and SPOT series of Earth observational satellites, hyperspectral data such as that from the Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) or the data that will be produced by the next generation of Earth observational sensors such as the Lewis Hyperspectral Imager (HSI) and Moderate Resolution Imaging Spectrometer (MODIS). The primary purpose for the system was to make new, otherwise complex analysis tools available to the general Earth science community. It has also found use in displaying and analyzing many other types of non-space related digital imagery, such as medical image data and in K-12 and university level educational activities.

MultiSpec satisfies the following design goals:

- The implementation should be on a readily available computer platform which has adequate processing power, but is financially within the reach of any Earth science researcher (i.e., computer platforms < $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.

The more complete system is implemented under the Apple Macintosh® operating system. A version that runs under the Microsoft Windows® operating system is under development specifically for the GLOBE program. The Windows version currently contains a subset of the features that are available in the Macintosh version.

Though copyrighted, MultiSpec with its documentation is distributed without charge. The Macintosh and Windows versions and documentation on its use is available from the World Wide Web at URL:

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

---

* The research leading to the creation of MultiSpec was supported in part by NASA Grants NAGW-925 and NAGW-3924.

INTRODUCTION

MultiSpec is a software application for analyzing multispectral or imaging spectrometer data such as LANDSAT MSS or TM data or the data that will be produced by the next generation Earth observational sensors such as the TRW Lewis HSI and Earth Observational System's MODIS, as well as data from current airborne systems such as AVIRIS. MultiSpec is implemented for personal computers running the Apple Macintosh operating system including PowerPC or 680x0 based machines with or without a math coprocessor. A version of MultiSpec is also implemented for machines running the Microsoft Windows 3.1 operating system. The Windows version currently contains a subset of the features that are available in the Macintosh version.

The Windows version of MultiSpec will run under the Windows 95 and Windows NT operating systems as a 16-bit application. The Macintosh version will run on any Unix workstation that contains the MAE, Macintosh Applications Environment, package.

MultiSpec satisfies the following design goals:

- The implementation should be on a readily available computer platform which has adequate processing power, but is financially within the reach of any Earth science researcher (i.e., computer platforms < $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.

MultiSpec had its origin in the LARSYS multispectral image analysis system (Phillips, 1973) which was one of the first remote sensing multispectral data processing systems, originally created during the 1960's. A number of the systems in government laboratories, university research labs, and several commercially offered products are descendants of this system.

The original purpose for the development of MultiSpec was to provide an easy to use tool that could be used for teaching, research, but especially, to provide the ability for researchers to try new techniques without having to program the algorithms. We use it in our graduate level remote sensing courses to teach signal processing analysis techniques. Our graduate students use it to take advantage of algorithms developed by previous students as they develop and study new algorithms.

The original LARSYS was designed with data in mind that had of the order of 15 or less spectral bands and with the intention of discrimination between as many as 20 or so classes. The new implementation is intended to provide a better degree of interactiveness and ease of use, and also deal effectively with the larger number of spectral bands (of the order of 100 to 200 or more) of future sensor systems and a larger number of classes (of the order of 50 or more). The current version of MultiSpec has been used for displaying and analyzing data of more than 200 spectral bands. However, it is envisioned that new processing algorithms will be needed to deal optimally with this new, more complex data of the future. Thus, this present implementation, while useful in and of itself, is an interim step toward achieving an effective processor for the future, as it currently contains only a few of the algorithms that will be needed to optimally deal with high dimensional multispectral data of the future.

The package has been copyrighted to protect against misuse, but it is freely distributed to requesters by the least expensive and most convenient means available. User interest has been significant. In previous years, several hundred requests for copies were received, and copies were mailed to these requesters. Beginning in 1995, we made the software available via the World Wide Web. Even though the software is free, we do request that users not give it to someone else but rather refer them to us or to the web page. This way we can keep a record of who has the software. In the past 18 months over 1000 copies have been downloaded to users in 31 countries representing universities, government, commercial and K-12 institutions.

An unexpected outcome was that a substantial number of requests for MultiSpec came from teachers of the K-12 level. As a result of this type of interest, a distribution license has been granted to The Consortium for Mathematics and Its Applications (COMAP), a National Science Foundation funded program based in Lexington, Massachusetts, for use in their ARISE (Applications/Reform in Secondary Education) program. This is a 5-year project to generate new mathematics curricula for grades 9-11. They report having incorporated MultiSpec into their new curricula as a motivating tool. Many requests for MultiSpec have also come directly from secondary school teachers.

Development of a Windows 3.1 version of MultiSpec began in 1995 in support of the NASA/NOAA/NSF GLOBE program, a program motivated by Vice President Gore to have K-12 children involved in the collection of environmental data. More than 3000 schools across the U.S. and 30 foreign countries have signed up to participate in the program. MultiSpec is used as a tool in the image processing portion of the program.

TYPICAL PROCESSING SCENARIO

The analysis of (moderate dimensional) multispectral image data may follow any of a number of approaches and processing steps, however, a perhaps typical generic list of steps might be as follows. After appropriate off-line preprocessing steps such as radiometric calibration, geometric or cartographic adjustment, and reformatting, the researcher, in an interactive mode, might proceed through the following steps.

1. **Data Review.** This is to gain general familiarity with a data set, its quality and general characteristics and is usually done, at least in part, by viewing the data in color IR image form or side-by-side channel form. Thus some type of image display is used first.

2. **Class Definition.** Definition of the class of material to be identified or the set of classes to be discriminated between must be carried out. Some means for mathematically defining the specific class characteristics of interest is required here. Often this is accomplished by the researcher labeling a small sample of the pixels from the data itself, as representative of the classes of interest. The analysis process then becomes an extrapolation from these samples, called training samples or design samples, to the entire data set.

3. **Feature Determination.** The specific features to be used in the analysis must be identified or calculated. This may be simply a process of selecting an optimal subset of spectral bands, or there may be some calculation process to combine bands in some useful way.

4. **Analysis.** The specific analysis algorithm is finally applied to the data set to carry out the desired identification or discrimination.

5. **Results Evaluation.** Both quantitative and qualitative means are used to determine the quality and characteristics of the results obtained.

The current implementation of MultiSpec provides at least some means for accomplishing each of these steps. The primary point of departure is that the analysis desired is one that is relative in nature, meaning that rather than identifying a single class of material in a subject vs. background mode, each pixel is to be assigned to a class based upon a judgment criterion relative to all possible classes. This implies that enough classes must be defined so that there is a logical class to which to assign every pixel in the scene, with the possible exception of a small number of pixels that will be determined by a threshold.

Further, the current implementation assumes that the classes are made up of sums of multivariant Gaussian distributions. Thus tools such as histogramming and clustering are provided to assist in determining the modes of the data and to properly define classes and subclasses that reasonably fit this assumption.

Finally, the current implementation is to be regarded as by no means complete. A number of new algorithms are under evaluation or development to broaden the circumstances under which the system will be effective.

CURRENT CAPABILITIES

Capabilities of the current version of MultiSpec, released 7/30/96, include the following.

- **Import data** in either Binary or ASCII format with or without a header, and in Band Interleaved by Line (BIL), Band Sequential (BSQ), or Band Interleaved by Sample (BIS) formats. The data may have either one or two bytes per data value, and may have 4 to 16 bits per data value. In the case of two bytes per sample, the two bytes may be in either order. MultiSpec recognizes several header formats including ERDAS 7.3 and 7.4, IMG, VICAR, Sun "Screen Dump", GAIA, MacSADIE, TARGA uncompressed, TIFF uncompressed, LARSYS Multispectral Image Storage Tape (MIST), MultiSpec ASCII classification and HDF formats.

- **Display multispectral images** in a variety of B/W or color formats using linear or equal area gray scales; display (internally generated) thematic images also in B/W or color, with an ability to control the color used for each theme. The channels in a multispectral image may be displayed as side by side gray scale images. See Figures 1, 2 and 3.

- **Histogram** data for use in determining the gray scale regime for a display or for listing and graphing.

- **Reformat** the data file in a number of ways, e.g., by adding a standard header, changing from any one of the three interleave formats to either of the other two, editing out channels, combining files, adding or modifying channel descriptions, mosaicing data sets, changing the geometry of a data set, and a number of other changes.

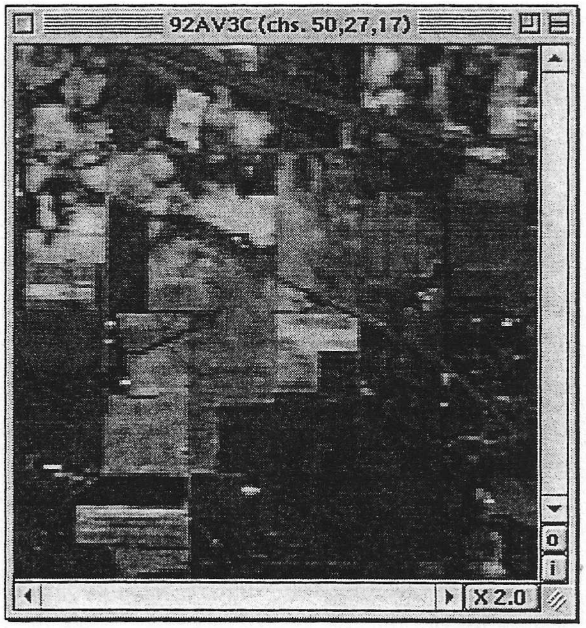

Figure 1. Example 3-channel color multispectral image from a 1992 AVIRIS data set. The 'i' and 'o' buttons are controls for zooming in and out respectively. The 'X 2.0' control indicates the current zoom factor and gives a quick way to go to a X1.0 zoom factor. One can zoom in 0.1 or 1.0 increments.

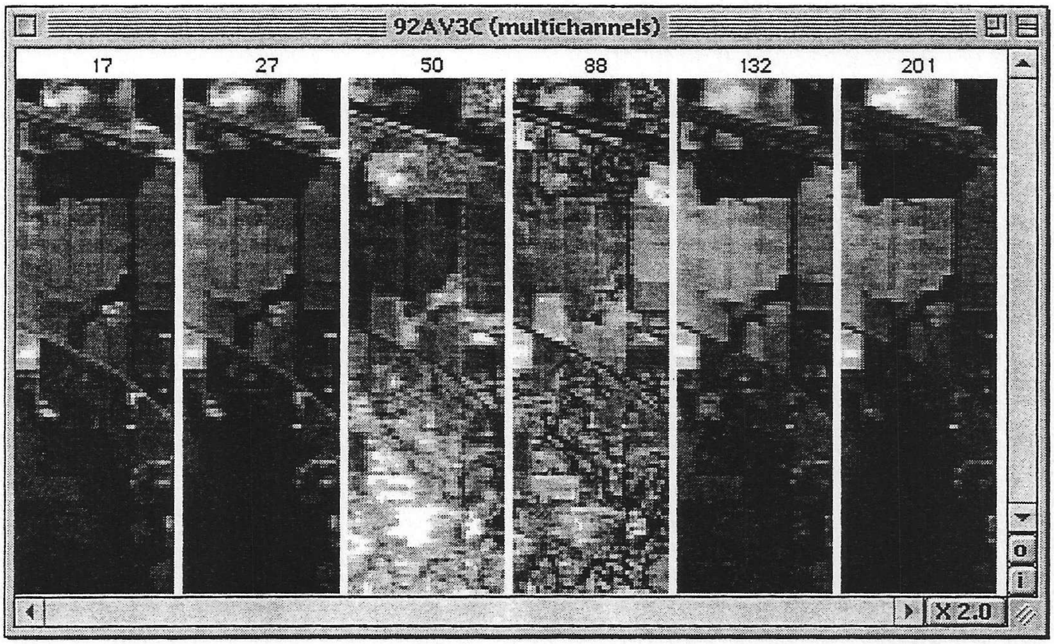

Figure 2. Example side-by-side multispectral image from a 1992 AVIRIS data set. The channel number is listed at the top of the respective image. The channel descriptions are listed at the top in place of the channel number if they are known.

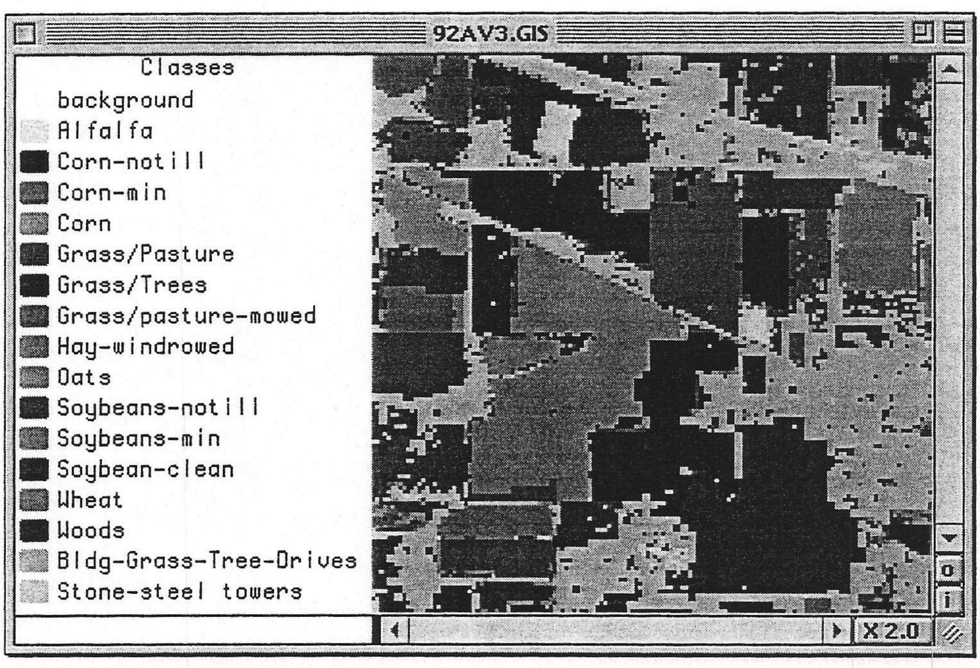

Figure 3. Example thematic image from a classification of a 1992 AVIRIS data set. The legend for the classes (or information groups) is listed to the left of the image.

- **Create new channels** of data from existing channels. The new channels may be the result of a principal components or feature extraction transformation of the existing ones, or they may result from the ratio of a linear combination of existing bands divided by a different linear combination of bands.

- **Cluster** data using either a single pass or an iterative (isodata) clustering algorithm. Save the results for display as a thematic map. Cluster statistics can also be saved as class statistics. Use of clustering followed by ECHO (Extraction and Classification of Homogeneous Objects) spectral-spatial classification provides an effective multivariate unsupervised scene segmentation scheme.

- **Define classes** via designating rectangular or polygonal training fields, compute field and class statistics, and define test fields for use in evaluating classification results quantitatively. A feature called "Enhance Statistics" also allows one to improve the extent to which the defined class statistics fit the composite of all data in the data set.

- **Determine** the **best** subset of **spectral features** to use for a given classification using (a) any of four statistical distance measures or (b) a relative new method based directly upon decision boundaries defined by training samples, or (c) a second method based directly upon the discriminant functions. Also included are methods especially designed to search for narrow spectral features such as spectroscopic characteristics.

- **Classify** a designated area in the data file. Four different classification algorithms are available: use of minimum L1 or L2 distance, the maximum likelihood pixel scheme, or the ECHO spectral-spatial classifier. Save the results for display as a thematic map, with or without training and test fields being shown. Apply a threshold to a classification, and generate a probability map showing the degree of membership of each pixel to the class to which it was assigned.

- **List classification results** of training or test areas in tabular form on a per field, per class, or groups of classes basis.

- Show the coordinates of a currently selected area, Figure 4. Show a **graph** of the **spectral values** of a currently selected pixel or the mean ± one standard deviation for a selected area, Figure 4. Show **scatter diagrams** of data from pairs of bands and ellipses of concentration for training sets or selected areas.

- Show a **color presentation** of the **correlation matrix** for a field or class. See Figure 5.

- Several additional **utility functions** including listing out a subset of the data e.g., for use externally, conducting principal component analysis, etc.

- **Transfer** intermediate or final **results**, be they text, graph, color or B/W image, to other application programs such as word processors, spreadsheet, or graphics program by copying and pasting or by saving and then opening the saved file within another application.

Some new or unique algorithms are:

- ECHO (Landgrebe, 1980)

- Statistics Enhancement (Benediktsson, 1996; Shahshahani, 1994)

- Feature Extraction including decision boundary (Lee, 1993), discriminant analysis and projection pursuit (Jimenez, 1996).

- Statistics Image. See Figure 5.

- Classification Probability Map. See Figure 6.

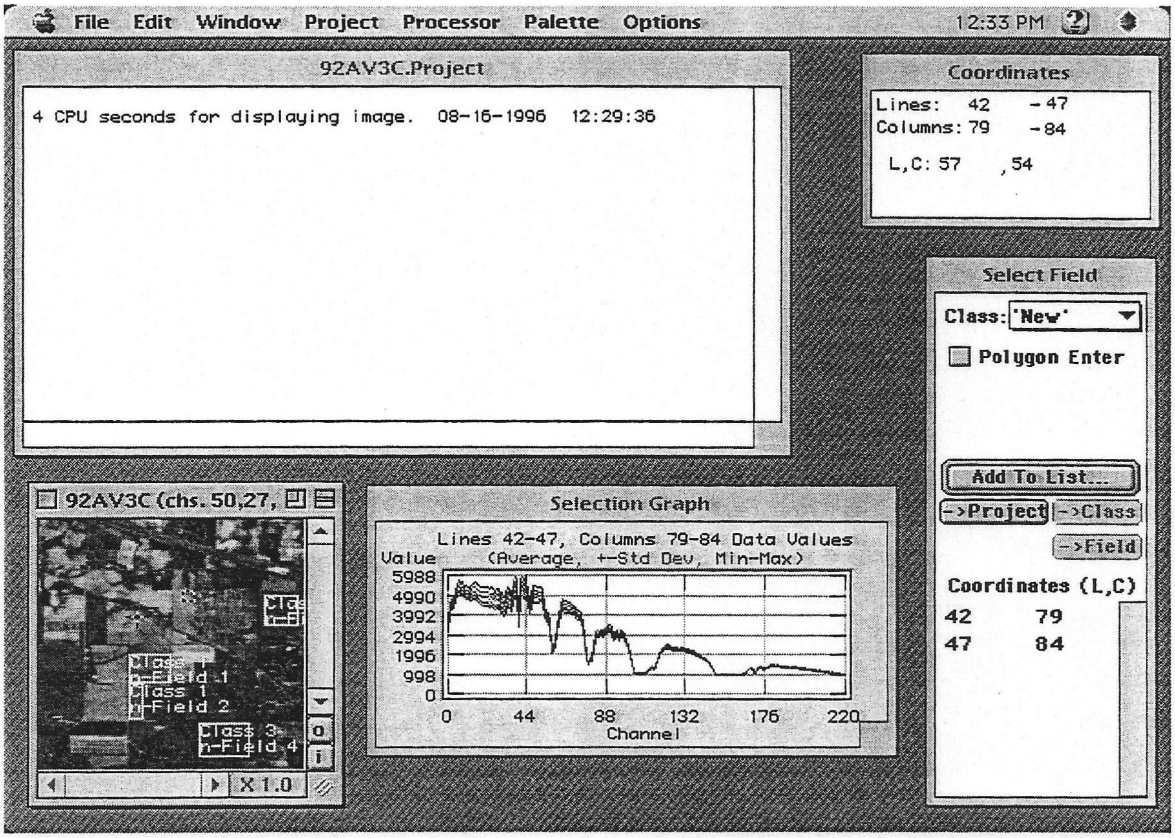

Figure 4. Example screen display illustrating several MultiSpec windows including the text window, an image window, a graph window, the coordinates window and the statistics window.

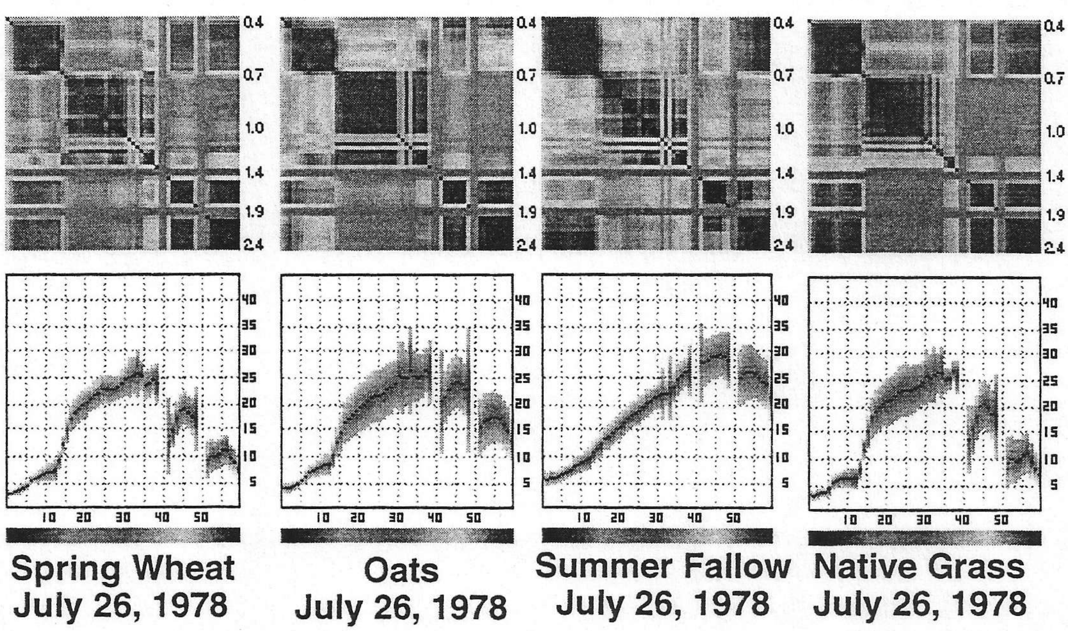

Figure 5. Comparison of four 'Statistics Images' illustrating the differences in the correlation matrices for four crop classes.

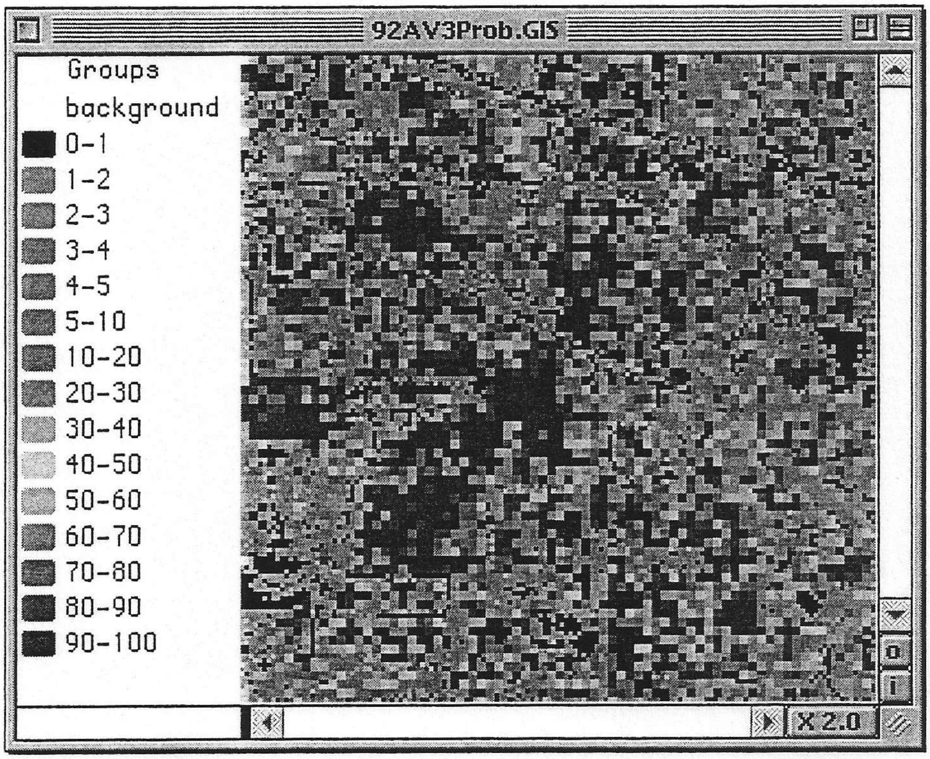

Figure 6. Example of a Classification Probability Map for an ECHO classification. The values in the legend represent the probability, in percent, of belonging to the assigned class.

The MultiSpec implementation is carried out in such a way that effectively the limits on the number of lines or columns of the data, the number of spectral bands, etc., are those determined by the available disk and memory space. See Table 1.

Table 1. Current design limits for the major parameters in MultiSpec

Image File	
lines & columns	99,999
multispectral channels	8,192
thematic classes	32,767
data value bytes	2
Display	
number of lines	32,767
number of columns	4,094-16,376
magnification	0.01-99
number of classes	32,767
Project Statistics	
statistics channels	8,192
statistics classes	255
total number of areas/fields	32,767
fields per class	512

SUMMARY

Taken together, these capabilities provide a state-of-the-art capability to analyze moderate and high dimensional multispectral data sets of practical size. The MultiSpec application along with an eighty page document listing its capabilities in more detail and providing tutorial exercises in its use and additional documentation are available via the World Wide Web at the URL given above in the ABSTRACT.

REFERENCES

Benediktsson, J.A., K. Arnason, A. Hjartarson, and D.A. Landgrebe, 1996. "Classification And Feature Extraction Based On Enhanced Statistics," Proceedings of the IGARSS '96 Symposium, Lincoln, NE, pp 414-416, 27-31 May, 1996.

Jimenez , L.O. and D.A. Landgrebe, 1996. "High Dimensional Feature Reduction Via Projection Pursuit," PhD Thesis and School of Electrical & Computer Engineering Technical Report TR-ECE 96-5, April 1996.

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

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

Lee, C. and D.A. Landgrebe, 1993. "Analyzing High Dimensional Multispectral Data, IEEE Transactions on Geoscience and Remote Sensing, Volume 31, No. 4, July 1993.

Phillips,.T.L., ed., 1973. LARSYS User's Manual. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

Shahshahani, B.M. and D.A. Landgrebe, 1994. "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.