MultiSpec©
Tutorial

January 11, 2001

Program Concept and Introduction Notes by
David Landgrebe and Larry Biehl

MultiSpec Programming by
Larry Biehl

School of Electrical and Computer Engineering
Purdue University

MultiSpec is a data analysis software system implemented for Macintosh computers. A version is also available for PC-Windows systems, though it does not yet have all of the capabilities of the Macintosh version. MultiSpec is intended for the analysis of multispectral image data, such as that from the Landsat series of Earth observational satellites or hyperspectral data such as from AVIRIS, MODIS, and other systems which contain many bands. The primary purpose of the system is to make new algorithms resulting from our research into hyperspectral data analysis conveniently available for others to try, although it has found additional uses in other circumstances, such as university and K-12 education, and in the government and commercial sectors.

The system presumes access to a Macintosh or a PC-Windows machine. A color display is highly desirable. The tutorial in section II requires a copy of the Thematic Mapper data set labeled TIPJUL1.LAN.

Current information, such as the availability of new updates to MultiSpec and a substantial amount of additional documentation on its use is available from the World Wide Web at URL:

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

Questions or comments regarding MultiSpec may be directed to:
Professor David Landgrebe
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA 47907-1285
Internet: landgreb@ecn.purdue.edu

How to use this document. There are three major sections to this document following a brief background in Section I.

- **A Tutorial.** Section II provides a brief tutorial in the use of only the most basic of system capabilities, with program command steps described in a quite detailed fashion. It is intended as a first contact with MultiSpec.

Though attempts have been made to make use of the program as intuitive as possible, it is recommended that the new user read and follow through the steps of Section II before attempting the analysis of new data. Further, a "familiarity reading" of Section IV will be very useful, as there are many additional features and capabilities described in it which do not occur in the earlier sections.

This document focuses on MultiSpec, itself, and how its various options may be invoked. Information on how its various algorithms may best be applied to actually analyze a data set are discussed in documents entitled,

1. On Information Extraction Principles for Hyperspectral Data

This is a white paper that begins by providing a general background to multispectral data. It discusses the various ways multispectral data may be viewed, what such data is like, and general concepts for data analysis. It presents results of research into unique (and perhaps non-intuitive) characteristics of high dimensional data feature spaces, and provides an approach to analyzing hyperspectral data.

2. Multispectral Data Analysis: A Signal Theory Perspective

This document provides a sketch of the theory and concepts implemented into MultiSpec. It also follows step by step through two example analyses, one for a nine-band data set, the second for the 220-band hyperspectral version of the same data. It thus provides a comparative example of the added power of hyperspectral data over conventional multispectral data.

3. An Example Hyperspectral Data Set Analysis

This document provides an example of an analysis of an airborne hyperspectral data set over the Washington DC mall area. The processing steps are listed giving the processor or operator time needed for each in order to illustrate that, though data volumes for hyperspectral data may be large, processing times and analyst times need not be.

These documents may be viewed or download from the Documentation page of the MultiSpec web site:

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

Table of Contents

I. BACKGROUND 4

A. TYPICAL PROCESSING SCENARIO..4
B. THE MULTISPEC IMPLEMENTATION..5

II. A TUTORIAL EXAMPLE 7

A. DISPLAY AND INSPECTION OF THE DATA..7
B. TRAINING SAMPLE SELECTION..9
C. CLASSIFICATION..11
D. OBTAINING A HARD COPY PRINTOUT..12
E. QUANTITATIVE CHECK OF ACCURACY..13
F. CLUSTERING ALGORITHMS..15
G. THE FEATURE SELECTION PROCESSOR..15
H. EXAMPLE USE OF OTHER SOFTWARE..16

AN INTRODUCTION TO MultiSpec

I. BACKGROUND

MultiSpec had its origin in the LARSYS multispectral image data analysis system and is implemented for PowerPC or 680x0 based Macintosh personal computers with or without math coprocessors[1] or PC-Windows based machines[2]. LARSYS 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 purpose of the current system is as a means to allow others to try out the results of our research into how to optimally analyze multispectral and especially hyperspectral image data. New versions are made available periodically as new algorithms emerge from our research. The ultimate goal is to provide a practical, fast, easy-to-use means for analyzing both conventional multispectral data such as LANDSAT MSS or TM data and hyperspectral data such as from MODIS, AVIRIS and other sensors producing high dimensional data.

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 the same degree of interactiveness and ease of use, but dealing effectively with the larger number of spectral bands (of the order of several hundred) of hyperspectral sensor systems and a larger number of classes (of the order of 50 or more). The current version has been used for displaying data of more than 200 spectral bands. However, it is envisioned that additional new processing algorithms will be needed to deal optimally with this new, more complex data. 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.

A. TYPICAL PROCESSING SCENARIO

The analysis of conventional (lower 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, 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 multiband B/W and color IR image form. Thus some type of image display is needed first.

---

[1] A different version is required for PowerPC Macintoshes and 680x0 based machines with a math coprocessor than with 680x0 based machines without a math coprocessor. A

[2] The version for PC-Windows machines does not yet have all of the capabilities of the Macintosh version.

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 quantitatively 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 the available spectral bands, or there may be some calculation process to combine bands in some useful way.

4. **Analysis.** The specific analysis algorithm is 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 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 multivariate Gaussian distributions. 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 and to make the process for user-efficient and convenient.

B. THE MULTISPEC IMPLEMENTATION

For this section, it is assumed that the reader is familiar with the standard user interface, and knows how to use pull-down menus, point, click, and drag with the mouse, etc. If this is not the case, the user should refer to the computer operating system manual or tutorials. The MultiSpec application should also be copied onto the hard disk, if it is not already there, as well as the data to be used. At this point,

- Start the MultiSpec application by double-clicking on it.

Upon opening the MultiSpec application, a Menu Bar and Text Output Window will be displayed as shown in Figure 1. The Text Output Window is the window in which all text output from the program, such as file statistics, histogram data, and initial classification results will appear. Text data appearing there may be copied and pasted into other applications, for example.

Notice the list of menus across the top of the screen and in Figure 1. By pointing with the cursor to each and holding the mouse button down, the various options of each menu item may be observed. In doing so, it is seen that the **File** menu allows one to open an **Image** or **Project Image** window for displaying an image, to **Print** output, to **Save** the text output to a disk file, or to **Change an Image Description**, among other things.

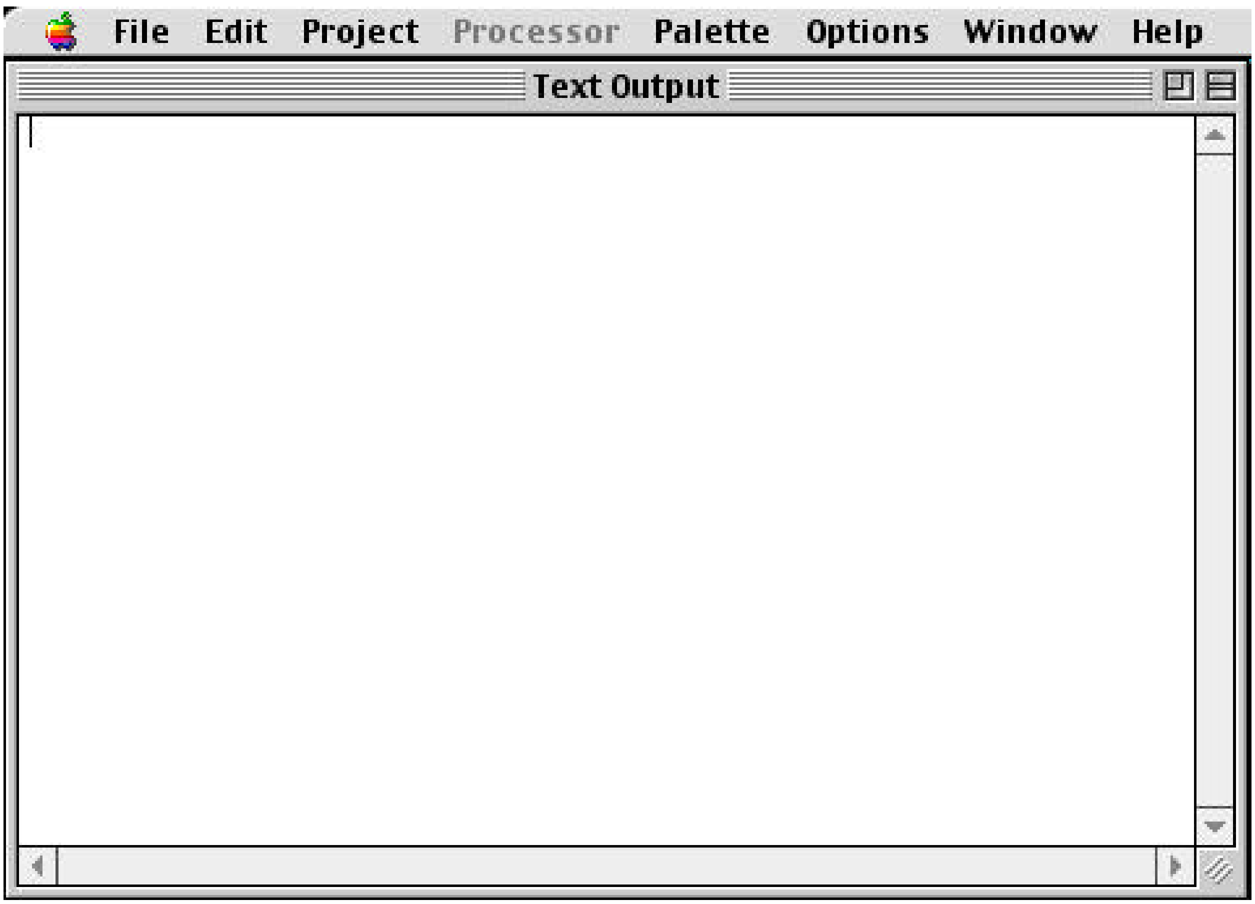


Figure 1. The MultiSpec Menu Bar and Text Window. This is what appears on the screen upon start-up of MultiSpec.

The **Edit** menu is used for standard user interface functions such as cutting, copying, pasting, and clearing. The **Window** menu is used to bring any given window to the front and make it the active window. The **Project** menu is used to start a new **Project** file. The Project file is the file used to store training and test field coordinates, class statistics, etc., so that one may save one's intermediate results and quit before completing an analysis. One can then restart the analysis at a later time without losing any work. The **Processor** menu allows one to choose the MultiSpec processor one wishes to use. The **Palette** and **Options** menus are needed only for advanced purposes and are explained in the Reference section.

II. A TUTORIAL EXAMPLE

A typical analysis task usually involves many classes, however, as a simple example, we will use a small portion of a Spectral Visions ITD data set to carry out a 2-class, 3 band classification. The two classes will be Vegetation and Sparse Vegetation. The data are shown in Figure 2. We will use the three fields indicated in the figure as training data.

A. Display and Inspection of the Data

To begin the example, open MultiSpec if it is not already and,

- From the ***File*** menu choose ***Open Image. . .***.

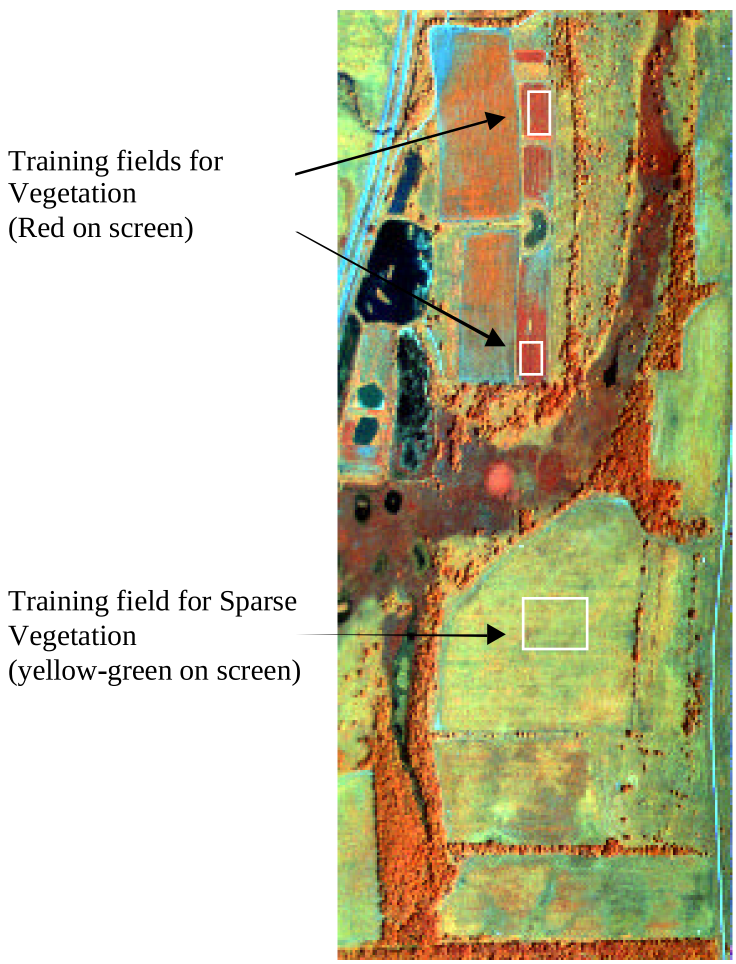


Figure 2. An image presentation of the data in the file BBR-C.lan showing fields to be used for training.

A dialog box will open to allow one to select the data file one wishes to use.

- Select ***BBR-C.lan*** and ***Open,*** or simply double-click on ***BBR-C.lan***

This is a segment (1519 lines x 636 columns of pixels) of a 3 channels of an ITD scene of ____________, Illinois gathered on July ??, 2000. Next a dialog box will appear to allow one to choose among various options for the image display.

- Note that by default the area designated for display is the whole scene and the ***3-Channel Color*** Display Type is selected. If the monitor is set to display 256 or fewer colors, the default settings call for the Red screen color to be derived from band 4 and the Green and Blue screen colors from band 1[1]. Any bands could be chosen, however, since band 3 is from the reflective infrared region and band 2 from the visible red, these particular choices are satisfactory for our use and will cause the screen image to be in a 3-color format approximating Color Infrared film. Therefore, click ***OK.***

If the data histogram has not previously been calculated and stored, another dialog box will next be presented allowing the choice of regions to be histogrammed, so that the various channel gray values can be properly assigned to screen colors. The default options built into this dialog box are satisfactory, so

- Click ***OK*** to begin the histogramming.

After the histograms of the seven channels have been compiled, a dialog box will be presented allowing them to be stored so that they will not have to be re-compiled when this data set is next used. The default file title is satisfactory, so,

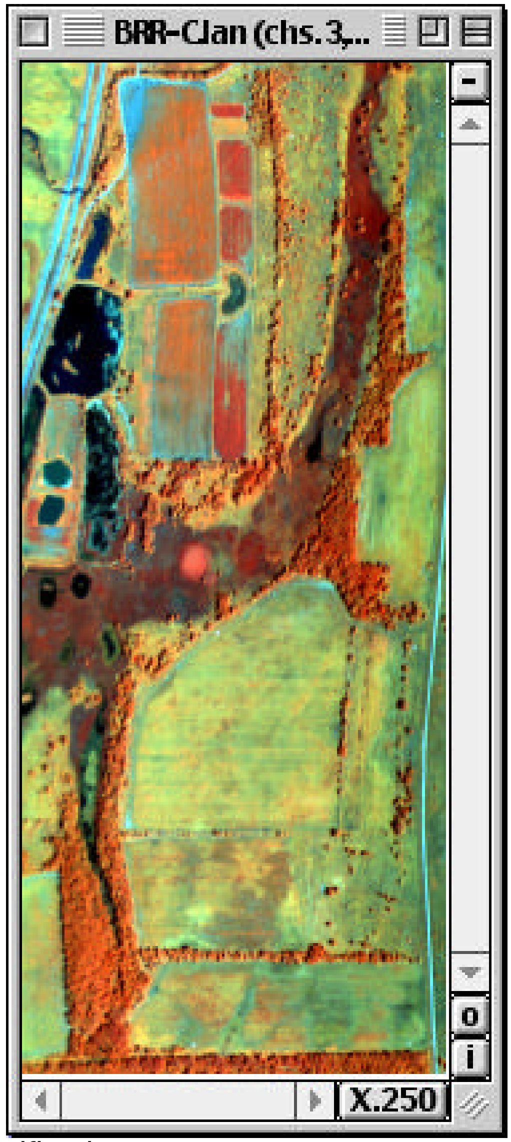
- Click ***Save*** to store the histogram in the image statistics file, BBR-C.sta.

1 Assuming an 8 bit display is being used, setting two colors to the same band results in a larger number of "gray" values per color. This usually results in a better appearing display. If a 24 bit color display is being used then each color is assigned to a different band, providing a display of full quality.

The image of the data will next appear. First, use the window sizing box in the lower right corner of the image window to[1]

- ***Enlarge the window*** to a little more than twice its current size by dragging it to the right and down.

Notice that just above the window sizing box, there are two small boxes labeled **i** and **o**. These are image zooming boxes allowing one to zoom **i**n or **o**ut from the current image scale. Just to the left of the window sizing box is another box which now shows



x .250. This tells the current zoom magnification.

- Click the ***i*** once to magnify the image by a factor of .333.

One can magnify further if one wants just by clicking additional times on the ***i***. Reduce the magnification by clicking on the ***o***. Any time the magnification is different from 1.0, one may click on the magnification size box to return it to x 1.0. After experimenting a bit, return the magnification to ***X 0.333***. By clicking on the Window Zoom Box (upper right corner of the window), one can adjust the window to just the size of the image.

B. Training Sample Selection

The next step is to select training fields. To do so,

1 The implementation for the PC-Windows version varies slightly from that described here. See the MultiSpec Windows portion in section IV. MultiSpec Reference section of these notes for details.

- From the ***Processor*** menu select ***Statistics*** and click ***OK*** in the resulting dialog box as we will use all seven spectral bands in statistics calculation and classification, and the other default values are satisfactory.

A new window labeled ***Select Field*** will appear in the upper right corner of the screen that will be used in a moment. To select training fields for each class, one must simply "drag" a rectangular area on the image (or, with polygon option selected, click on the corners of the desired polygon), and then "Add that field to the list." Thus,

- ***Drag from the upper left corner to the lower right corner*** of the Vegetation training field in the ***image window*** (the bottom one of Figure 2 that shows yellow-green on the screen). If upon inspection, one does not like the exact boundaries resulting, one may immediately repeat the process.

- Note in the ***Select Field dialog box***, that the coordinates (row and column numbers) of the upper left corner and the lower right corner of the selected area appear there. Now,

- Click on the ***Add to list*** button.

A dialog box will appear to allow one to name the class and give the field a special designation, as desired. Thus,

- Type ***Sparse Vegetaton*** into the Class Name box and then click ***OK***.

Since there is to be only one training field for this class, we are ready to select the training for the second training class. Thus next,

- Point to the ***Sparse Vegetation*** class name of the ***Select Field*** window and hold the ***Mouse button down***. A pop-up menu will appear. Without releasing the mouse button, drag to ***New*** and then let up on the Mouse button. Next,

- ***Drag across the second training field*** in the ***Image Window*** shown in Figure 2***.***

- Click on the ***Add to list*** button in the ***Select Field*** window .

Again, a dialog box will appear to allow one to name the class and give the field a special designation, as desired.

- Type ***Vegetation*** into the Class Name box and then click ***OK***.

The class Other is to have two fields in its training set, so we are ready to mark the second of these two.

- ***Drag across the remaining training field*** in the ***Image Window***

Since the class indicated in the Select Field window already is Vegetation, this field will automatically be associated with that class.

- Click on the ***Add to list*** button in the ***Select Field*** and click ***OK*** on the "Field 3" dialog box that results.

Since there are to be only two classes and a total of three training fields, the training process is complete, and we are now ready to evaluate the adequacy of our class definition and training. In more typical circumstances, it would probably be desirable to check its quality in other ways, however, for now we will evaluate it only by classifying the training fields to see that they are appropriately separable.

C. Classification

- From the ***Processor*** menu select ***Classify***.

On the dialog box which appears, under Classify

- Click on the √ near ***Image Selection***

to de-select it as, during this pass, it is desired to classify only the training fields in order to obtain an initial estimate of the quality of the class definition and training. Click on ***Project Text Window*** under ***Write classification results to:*** to select this option. This will cause the results of the classification to be displayed in the text window in alphanumeric form. Note that there is a ***Disk File*** button to cause a disk file version of the results to be written when one is classifying a selected area. Since we are only classifying the training pixels in this case, leave this button unselected.

Since the other default options are satisfactory,

- Click on ***OK*** and then ***Update*** to the "Update Project Statistics" dialog box to begin the classification.

The classification will be complete momentarily.

- From the ***Window*** menu select ***Text Output***,

to bring this window forward and make it active, as it will contain the classification results. If necessary, scroll until the results of the classification can be seen. The class of each pixel is indicated by a "1" or "2," indicating the class to which it has been assigned. Tables after the last field tabulate how the pixels of each field and class were classified. There should be nearly 100% accuracy on the three fields. Assuming satisfactory results, we are ready to classify the whole area. Although it is not really necessary, we could clear the Text Output window to make room for the results. If you wish to do this, with the Text Output window still active,

- From the ***Edit*** menu choose ***Select All*** and then ***Clear***.

To proceed with the classification,

- From the ***Processor*** menu choose ***Classify***.

- Under ***Areas to Classify*** de-select ***Training (resubstitution)*** by clicking on the √ by it, and,

- Select ***Image file***.

- Also click on ***Disk File*** under ***Write classification results to:*** so that a disk file for later use will be created. Then click ***OK***. Also click ***Save*** to the dialog box that follows regarding a file name for the results.

It should take just a few seconds[1] to classify the 966,084 pixels in the data set using all 3 bands. As soon as the classification is complete, one will see a table of the results displayed in the text window and one may want to scroll over the classified area to examine the results to see how many pixels were assigned to each class.

D. Obtaining a Hard Copy Printout

Hard copy in Thematic Map form. A hard copy output in Thematic Map form may be obtained by the following.

- In the ***File*** menu, choose ***Open Image***,

- In the resulting dialog box, select the file ***Untitled Project.GIS*** that was generated by the classifier, and click ***Open***. Click ***OK*** on the resulting dialog box.

One may change from the default color used for each class to more appropriate ones by double-clicking in the legend on the color to be changed, then clicking on the color desired in the color wheel resulting. For example, one may change to green for the Vegetation class. Note that colors of contrasting brightness should be chosen for the two classes if the intent is to subsequently print out a black and white version of the display.

One may enlarge the resulting image as desired using the ***i*** zoom box on the lower right. Then,

- In the ***File*** menu, choose ***Print Image***[2], and ***OK*** on the resulting dialog box.

Figure 3 shows an example of the color form. The image will be printed in gray scale form if the printer is not color.

Note also that one may copy the display to a word processor document by choosing 'Select All' or by dragging over the portion of the display desired, then choosing 'Copy,' and then pasting the result into the word processor document.

1 For a Macintosh G3. Times will vary depending on the computer model.

2 Use the Chooser in the Apple Menu to select a printer, if necessary.

Figure 3. A classification result in color thematic map form.

E. Quantitative Check of Accuracy

To assist in the evaluation of the analysis results, the LIST RESULTS processor may be used. To do so, first,

- Make certain that the Untitled Project.GIS window is active; (if it is not, click on any part of it showing or select it in the ***Windows*** menu),

- From the ***Project*** menu, select ***Add Associated Image.***

The outlines and names of the training fields will immediately be shown on the image. Next,

- From the ***Processors*** menu, select ***List Results***.

A dialog box providing several options will appear. These options are all described in the Reference Section. For now click on **Field** under Summarize by.(train/test only), and then

- click ***OK***.

This will cause the system to generate in the text window two tables of results for the accuracy of the training samples, similar to the following.

TRAINING FIELD PERFORMANCE (Resubstitution Method)

Project Field Name	Class	Reference Number	Accuracy+ (%)	Number Samples	Number of Samples in Thematic Image Class: 0 background	1 Vegetation	2 Sparse Veget
Field 1		1	99.8	1769	0	1766	3
Field 3		1	100.0	1305	0	1305	0
Field 2		2	100.0	7469	0	0	7469
TOTAL				10543	0	3071	7472

TRAINING CLASS PERFORMANCE (Resubstitution Method)

Project Class Name	Reference Class Number	Accuracy+ (%)	Number Samples	Number of Samples in Thematic Image Class: 0 background	1 Vegetation	2 Sparse Veget
Vegetation	1	99.9	3074	0	3071	3
Sparse Vegetation	2	100.0	7469	0	0	7469
TOTAL			10543	0	3071	7472
Reliability Accuracy (%)*					100.0	100.0

OVERALL CLASS PERFORMANCE (10540 / 10543) = 100.0%
Kappa Statistic (X100) = 99.9%. Kappa Variance = 0.000000.

+ (100 - percent omission error); also called producer's accuracy.
* (100 - percent commission error); also called user's accuracy.

The first table provides an accuracy assessment for the training fields. Each line of the table contains the results for one of the training fields. The second table combines the fields of each class to show the overall training class results. If Test Fields had been defined, similar tables could be calculated for these as well. A capability to Group classes, i.e., combine classes that are actually subclasses of a more general class, is described in the Reference section of this document, under the Open Image command of the File menu. If such Grouping had been done, tables as above could be determined for Groups as well.

F. Clustering Algorithms

Two Clustering Algorithms are contained in MultiSpec. They are useful in the classifier training step in determining how many classes might be separable in a given data set, and in finding the modes of data. One algorithm implemented is a simple one-pass type. The second is of the iterative type. More information on these algorithms is contained in the Using MultiSpec and Reference sections of this document.

G. The Feature Selection Processor

A Feature Selection Processor is included to assist in choosing a best subset of spectral features to be used for a specific classification. For example, instead of using all seven TM bands in the above tutorial example, it may have been better to choose the best three of the seven bands for the particular pair of classes. This would certainly have saved computation time, and in a more complex classification task, it may have provided higher classification accuracy.

The Feature Selection Processor produces a table showing the statistical distance between each of the class pairs for each of the possible subsets of features. In a case of three classes, the best three of seven features and the Bhattacharyya statistical distance measure, the results that appear in the text window are as follows:

Feature Selection 02-05-1999 14:33:45 (MultiSpec1.29.99)
Project = 'TIPJUL1.LAN'
Original class statistics are used.
Base image file = 'TIPJUL1.LAN'
List best 12 combinations.
Minimum value to be listed: 0
Maximum value to be listed: 30000
List separability table.
1 contiguous channels in each channel combination group
All possible channel combinations will be searched

Classes used:	Symbol
1: Veg	1
2: Other	2
3: 2nd other	3

Channels used:
1: 0.45-0.52 µm
2: 0.52-0.60 µm
3: 0.63-0.69 µm
4: 0.76-0.90 µm
5: 1.55-1.75 µm
6: 2.08 - 2.35 µm
7: 10.40 - 12.50 µm

Start Bhattacharyya Feature Selection

There are 3 class combinations.
There are 35 channel combination(s) for 3 group(s) of 1 contiguous channel(s).

	class pair symbols >			12	13	23
			weighting factor >	(10)	(10)	(10)
	Channels	Min.	Ave.	Weighted Interclass Distance Measures		
1.	5 6 7	1.07	8.42	15.2	8.95	1.07
2.	2 6 7	1.05	7.12	10.1	10.1	1.05
3.	3 6 7	1.01	7.01	8.35	11.6	1.01
4.	4 6 7	0.99	7.01	7.32	12.7	0.99

5.	1 6 7	0.94	6.33	6.72	11.3	0.94
6.	4 5 7	0.79	10.55	15.3	15.5	0.79
7.	1 2 7	0.70	5.99	10.0	7.20	0.70
8.	1 4 7	0.69	6.20	7.96	9.94	0.69
9.	1 3 7	0.65	5.20	7.77	7.17	0.65
10.	2 5 7	0.59	6.90	10.0	10.0	0.59
11.	2 4 7	0.58	7.99	12.3	11.0	0.58
12.	3 5 7	0.55	6.97	9.83	10.5	0.55

0 CPU seconds for feature selection. 02-05-1999 14:33:46

By choosing the correct option, the subsets of channels have been listed in descending order according to the minimum of the Weighted Interclass Distance values. (Another option would have the subsets listed in descending order according to the average of the Weighted Interclass Distance values.) Note from the table, this algorithm suggests that Channels 5, 6, and 7, would be the best three for discriminating among these three classes. However, though the minimum interclass separability for n-tuple 2, 4, 5 is the largest, if one were more concerned about good separation between classes 1 and 3, channels 4, 5, and 7 might be a better choice, since it has a maximum interclass distance of 15.5.

In general, in using this processor, one would want to scan both the Average column and the Minimum column of the processor output, perhaps in addition to reviewing the individual entries appearing in the table of Weighted Interclass Distance Measures. One could also use the Weights option to vary the interclass distance weighting in the averaging process to assist evaluation when there are a large number of classes. Alternatively, the Feature Selection Processor output may either be copied and pasted to a spreadsheet, or saved on disk as a text file and opened from a spreadsheet for more complex study of the table. Other example uses of commercially available software are contained in the following section.

H. Example Use of Other Software.

The use of commercial applications such as a graphing program can be very useful in extending the capabilities of MultiSpec, and indeed, such use is an intention of the design of it. As another example, Figure 4 shows a histogram of every fifth line and every fifth column of the data set for bands 3 and 4. This figure was obtained by choosing the Histogram Processor with the List Histogram option selected, and then copying the result from the Text Output window and pasting it into a commercially available graphing application program.

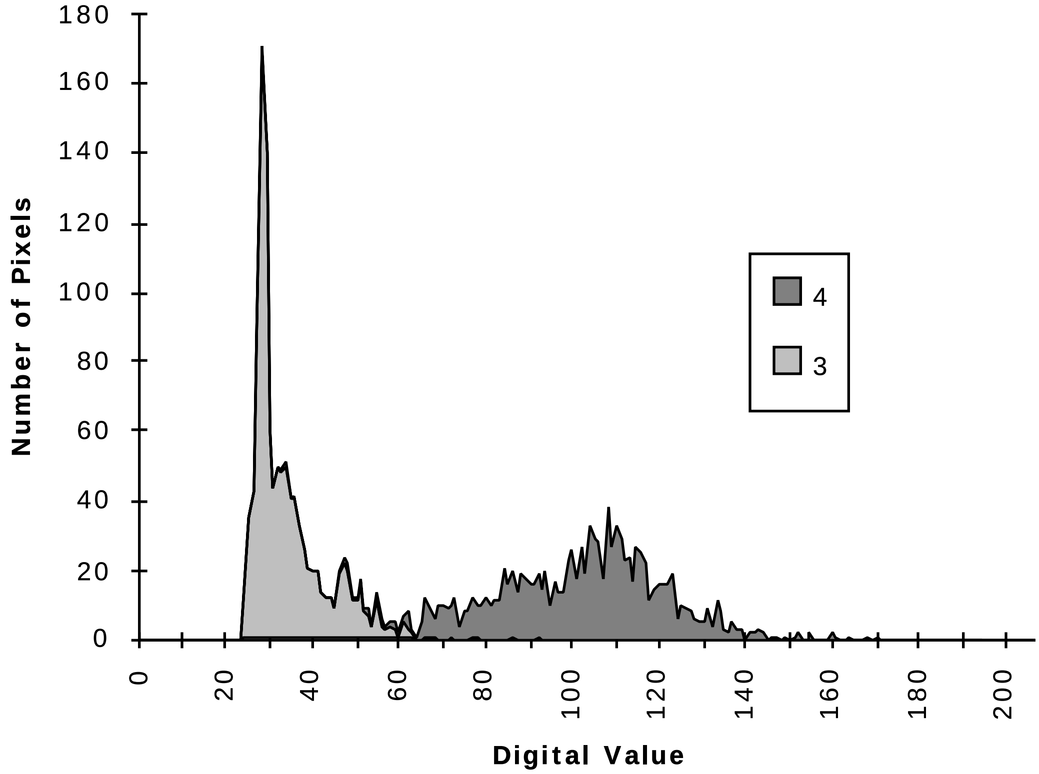


Figure 4. Histogram of TIPJUL1.LAN in bands 3 and 4