HYPERSPECTRAL DATA ANALYSIS PROJECT

This project will provide a realistic opportunity to explore the methodology of hyperspectral data analysis. You will use several of the algorithms studied in the textbook and available in MultiSpec to analyze a 191-band airborne multispectral scanner data set.

The Data Set. The figure here shows a simulated color IR view of an airborne hyperspectral data flightline over the Washington DC Mall provided with the permission of Spectral Information Technology Application Center of Virginia who was responsible for its collection. The sensor system used in this case measured pixel response in 210 bands in the 0.4 to 2.4 µm region of the visible and infrared spectrum. Bands in the 0.9 and 1.4 µm region where the atmosphere is opaque have been omitted from the data set, leaving 191 bands. The data set contains 1208 scan lines with 307 pixels in each scan line. It totals approximately 150 Megabytes. The image at left was made using bands 60, 27, and 17 for the red, green, and blue colors respectively.

Data Analysis - Part 1. Download the data set for the DC Mall data (labeled dc.tif) and a file labeled dctest.projec*t* from MultiSpec web site to the computer you intend to use. Carry out a carefully designed quadratic maximum likelihood supervised classification with the goal of constructing an accurate thematic map of the area showing the following ground cover types: *Roofs, Street, Path (graveled paths down the mall center), Grass, Trees, Water, and Shadow.* A copy of the file labeled *dctest.project* should be used for entering your training fields and already contains test fields for the above classes for determining a quantitative accuracy figure. Do not include any test field pixels in your training set, in order to obtain a better evaluation of the classifier's ability to generalize. You will need to use a feature extraction algorithm for this analysis, and you should carry out the analysis using DAFE for this part.

Draft a report covering at least the following:

1. One or more thematic images of your results, along with tables showing the accuracies obtained on your training samples and the test sample set provided in the *dctest.project* file and the information classes listed above.

2. The list of information classes indicated above, showing the spectral subclasses that form these

information classes.

3. The procedure you used to identify and select training fields, and a very brief explanation of why you chose this specific procedure.

4. Compare the final results obtained by using several different numbers of DAFE features.

Part 2 - Algorithms and Training Samples

The purpose of this portion of the project is to explore the use of analysis algorithms of different degrees of complexity and the relationship between that complexity and training set size. Fill out the following table with classifier algorithm accuracy performance data, based on your training samples and the standard set of test samples used in part 1. Enter the accuracy figures as Training/Test (e.g. 95/90) in each cell.

A. Use the set of classes and training samples devised in Part 1 as the baseline set of classes for this part of the project, and fill out the lines of results marked "1. Standard," LOOC, and Enhanced below by classifying the entire data set with each of the algorithms, grouping the subclasses into information classes as you did in part 1, and determining the training and test set accuracies.

Classifier	Min Dist.	Fisher Lin. Disc	Quadratic ML	ECHO	Corr(SAM)	Matched Filter(CEM)
Baseline 100-200 pixels/class						
1. Standard						
LOOC						
Enhanced						
1 Pixel/Training Field						
2. Standard						
LOOC						
Enhanced						

B. Next, complete the lines marked "2. Standard," LOOC, and Enhanced " by reducing the training set to include only the pixel in the upper left corner of each training field and, if need be, two of its neighbors to achieve a minimum of 3 pixels per subclass, re-computing the training statistics and optimal features, and classify the data as before. Where non-rectangular training fields were used, pick a single, typical pixel from within the training field for this purpose. For algorithms utilizing second order statistics (maximum likelihood and ECHO), this may lead to singular covariance matrices that prevent the algorithm from being used. Mark such results accordingly.

There is a significant relationship between the complexity of the algorithm used and the precision with which the classes are defined. This precision is directly related to the size of the training set used to estimate the training class statistics. Do your results to this point demonstrate this? Add the completed table above to your report of Part 1 and comment on the relationship between algorithms and class description that the results display.

Data Analysis - Part 3

The purpose of this section is to compare the use of Discriminate Analysis Feature Extraction (DAFE), Decision Boundary Feature Extraction (DBFE), and Nonparametric Weighted Feature Extraction (NWFE).

1. Use the Feature Extraction Processor of MultiSpec to explore the effect of using DAFE, DBFE and NWFE (We will not be using a Preprocessing Transformation available with DBFE at this time). Using the same training statistics and test samples as Part 1 above, apply DAFE, then classify using the transformed statistics. Use the first one, the first 2, ... up to 15 transformed features and plot the accuracy obtained for the test sample set. Save a Thematic Image and Probability Results Image for the case of the first 10 features.

2. Plot the magnitude of the first 5 eigenfunctions resulting from the transformation vs. the feature number. Comment on what implications you can draw from this plot.

3. Repeat 1 and 2 but using DBFE.

4. Repeat 1 and 2 but using NWFE.

Add to your previous report draft by adding comments consisting of appropriately annotated versions of the above results and graphs and a brief summary providing conclusions.