12-13-2002

Dimensionality Reduction of Hyperspectral Signatures for Optimized Detection of Invasive Species

Abhinav Mathur

DIMENSIONALITY REDUCTION OF HYPERSPECTRAL SIGNATURES

FOR OPTIMIZED DETECTION OF INVASIVE SPECIES

By

Abhinav Mathur

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in Electrical Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2002

DIMENSIONALITY REDUCTION OF HYPERSPECTRAL SIGNATURES

FOR OPTIMIZED DETECTION OF INVASIVE SPECIES

By

Abhinav Mathur

Approved:

Lori M. Bruce Assistant Professor of Electrical and Computer Engineering (Director of Thesis)	Roger L. King Professor of Electrical and Computer Engineering (Committee Member)
Nicholas H. Younan Professor and Electrical and Computer Engineering Graduate Program Coordinator (Committee Member)	John D. Byrd Jr. Professor of Plant and Soil Sciences (Committee Member)

A. Wayne Bennett
Dean of the College of Engineering

Name: Abhinav Mathur

Date of Degree: December 13, 2002

Institution: Mississippi State University

Major Field: Electrical Engineering

Major Professor: Dr. Lori M. Bruce

Title of Study: DIMENSIONALITY REDUCTION OF HYPERSPECTRAL SIGNATURES FOR OPTIMIZED DETECTION OF INVASIVE SPECIES

Pages in Study: 62

Candidate for Degree of Master of Science

The aim of this thesis is to investigate the use of hyperspectral reflectance signals for the discrimination of cogongrass (*Imperata cylindrica*) from other subtly different vegetation species. Receiver operating characteristics (ROC) curves are used to determine which spectral bands should be considered as candidate features. Multivariate statistical analysis is then applied to the candidate features to determine the optimum subset of spectral bands. Linear discriminant analysis (LDA) is used to compute the optimum linear combination of the selected subset to be used as a feature for classification. Similarly, for comparison purposes, ROC analysis, multivariate statistical analysis, and LDA are utilized to determine the most advantageous discrete wavelet coefficients for classification. The overall system was applied to hyperspectral signatures collected with a handheld spectroradiometer (ASD) and to simulated satellite signatures

(Hyperion). A leave-one-out testing of a nearest mean classifier for the ASD data shows that cogongrass can be detected amongst various other grasses with an accuracy as high as 87.86% using just the pure spectral bands and with an accuracy of 92.77% using the Haar wavelet decomposition coefficients. Similarly, the Hyperion signatures resulted in classification accuracies of 92.20% using just the pure spectral bands and with an accuracy of 96.82% using the Haar wavelet decomposition coefficients. These results show that hyperspectral reflectance signals can be used to reliably detect cogongrass from subtly different vegetation.

DEDICATION

I would like to dedicate this thesis to my parents Dr. (Mrs.) Manju Mathur and Mr. Vijay Kumar Mathur, whose guidance and support have allowed me to complete my every plan and excel in whatever I have attempted to do, and to my brother Raghav Mathur, whose critical eye has always kept me on my toes. I would like to thank them for creating an environment in which following this path seemed so natural.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor and director of thesis, Dr. Lori Mann Bruce. Her constant guidance and unconditional support throughout the process of completion of this thesis have been immeasurable. I thank Dr. John D. Byrd, Dr. Nicholas H. Younan and Dr. Roger L. King for serving as my thesis committee members.

I would like to take this opportunity to thank Mr. Louis L. Wasson for promptly helping me out to find specifications regarding the ASD equipment, Mr. Valentine G. Anantharaj and Mr. Donald B. Mask for supplying me with the ASD data, and the Remote Sensing Technologies Center for funding my research. Without their help it would have been a very long walk.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

The problem of detecting invasive species, particularly noxious weeds, in remotely sensed images, is of great importance in many agricultural and environmental applications. One of the most challenging tasks is to accurately distinguish the weeds of interest from neighboring, subtly different vegetation. The first step is to investigate species-related differences in the reflection spectra of plant leaves, to determine if the plant species of interest can be discriminated spectrally. There are six major wavelength-dependent components of a typical remotely sensed spectrum. These include solar flux (F), atmospheric transmission (T), surface reflectance (R), atmospheric scattering ($A\uparrow$ and $A\downarrow$, irradiance at the top and bottom of the atmosphere respectively), the instrument gain (G), and the dark current (D) [1]. Most healthy vegetation looks green in color to the human eye, indicating that the differences in spectral signatures of different plant species are quite subtle. Thus, to remotely distinguish plant species spectrally has been a challenge and has been made more difficult by the lack of adequate instrumentation.

With the advent of hyperspectral sensors providing reflectance measurements in the visible and infrared regions of the spectrum, more pertinent information has been gathered about the spectral behavior of plant species.

Spectral bands are much narrower and more in number as compared to the previous generation instruments like the multispectral Landsat Thematic Mapper (TM) and Multispectral Scanner (MSS). These multispectral systems can distinguish general brightness and slope differences in the reflectance spectrum of the target material. Hyperspectral systems, however, provide detailed reflectance measurements that allow one to resolve the absorption bands in the spectrum that can be used to identify specific species. This is due to the fact that the correlation between biophysical factors and spectral reflectance is influenced by the width and location of absorption bands within the electromagnetic spectrum. Spectroscopy allows any material to be mapped, whether natural or synthetic mineral, vegetation, water, snow, or other, if it displays unique absorption features in the measured spectral region.

The goal of this thesis is to create an algorithm that extracts only those features that are most pertinent for the discrimination and classification of two subtly different vegetation classes. The idea behind this feature reduction is not only to reduce computation time but also to improve the overall classification accuracy. This is done by using multivariate statistical analysis. Some labeled data can be used to train the system offline to locate and process these features. Then an online system uses this information to extract the “best features” and perform an automated classification.

This system will be applied to the problem of automated detection of cogongrass among other subtly different species like bahiagrass, bermudagrass, broomsedge, centipedegrass and vaseygrass. Cogongrass is a weed that spreads by

seed and vegetatively. Cogongrass produces numerous underground horizontal stems, or rhizomes, which are capable of rooting at each node and producing a new stem [2]. It has several common names, including japgrass, Japanese bloodgrass, Red Baron, or speargrass but its scientific name is Imperata cylindrica (L.) Beauv. Cogongrass looks different at different times in the year making it even more difficult to detect using a general off-the-shelf method. This is shown in Figures 1, 2 and 3. Different stages of growth of Cogongrass is shown in Figures 4, 5, 6 and 7.

Figure 1: Cogangrass in spring, photo courtesy of Dr. John D. Byrd Jr., Professor of Plant and Soil Sciences

Figure 2: Cogangrass in summer, photo courtesy of Dr. John D. Byrd Jr., Professor of Plant and Soil Sciences

Figure 3: Cogangrass in fall, photo courtesy of Dr. John D. Byrd Jr., Professor of Plant and Soil Sciences

Figure 4: Field of Cogongrass [2]

Figure 5: Mature Seedhead [2]

Figure 6: Immature Seedhead [2]

Figure 7: Off-set Midriff on Leaf [2]

Cogongrass has been designated the seventh worst weed in the world. It was both accidentally and purposely introduced into the southern United States (Alabama, Florida, and Mississippi) in the teens and early 1920's. Many farmers planted cogongrass for pastures and erosion control. Cogongrass was not a good livestock feed, and it was too weedy for erosion control. Currently, cogongrass

occurs as a weed in Alabama, Florida, Georgia, Louisiana, Mississippi, South Carolina, Texas, and Virginia, and it continues to spread. More than 1.2 billion acres worldwide are infested with cogongrass (Figure 8). Cogongrass is included on the Federal Noxious Weeds List, because of its aggressive, weedy habit. Cogongrass usually chokes out existing vegetation, due to the dense stems and rooting system. In Mississippi and other southern states, cogongrass usually occurs in non-cultivated sites, including pastures, orchards, fallow fields, forests, parks, natural areas, highway, medians, alongside pipeline, and alongside railroads (Figure 9 and 10).

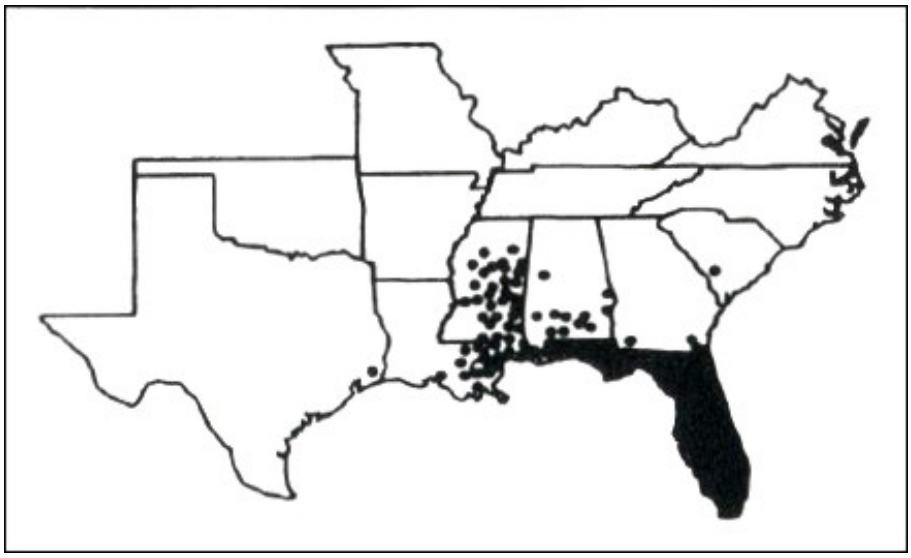

Figure 8: United States Distribution Map of Cogongrass [2]

Figure 9: Cogan grass spreading on highway median, photo courtesy of Dr. John D. Byrd Jr., Professor of Plant and Soil Sciences

Figure 10: Cogan grass spreading in a field, photo courtesy of Dr. John D. Byrd Jr., Professor of Plant and Soil Sciences

CHAPTER II

CURRENT STATE OF KNOWLEDGE

Methods for automated species detection based on spectral information of plant canopies have been investigated in the past. Four main approaches that were studied are the blue shift of the red edge, vegetation indices, optimal spectral subset extraction and wavelet based features.

2.1 Blue shift of the red edge

In 1995, a method was presented that analyzed the position of the blue shift of the chlorophyll absorption red edge [3]. The red edge generally refers to the suddenly increasing value in the reflectance signature as the wavelength increases over the interval of approximately 650-800 nm. This shift was found to have varying levels for different fields of vegetation under observation. The "Tricorder" algorithm was used to distinguish between different species based on the shape differences of their spectral signatures. This algorithm utilized continuum-removed spectra. Continuum removal is a procedure that facilitates the distinction of similar absorption bands in hyperspectral curves. The continuum consists of the so-called "background absorption," which is, in essence, an extrapolation of the baseline of the general curve (fits a smoothed curve to the general trend extended across the base of absorption bands). This local reduction

specifies the continuum and is determined by mathematical manipulation of the absorption coefficients by a subtraction process [4]. The "Tricorder" algorithm compares the continuum-removed spectral features from the imaging spectroradiometers data set to a corresponding set of continuum-removed spectral features from a library of reference spectral signatures [5]. Multiple features from different materials are compared, and the best match is selected as the class of the test sample.

2.2 Vegetation Indices

A great deal of research has also been done to determine optimal vegetation indices and band combinations to discriminate different vegetation species [6]. Among these are the Normalized Difference Vegetation Index (NDVI),

$$NDVI = (R_{nir} - R_{red})/(R_{nir} + R_{red}) \tag{1}$$

the Modified Soil and Atmospherically Resistant Vegetation Index (MSARVI) [6],

$$MSARVI = \{2\rho^*_{nir} + 1 - [(2 * \rho^*_{nir} + 1)^2 - 8(\rho^*_{nir} - \rho^*_{rb})]^{0.5}\}/2 \tag{2}$$

the Photochemical Reflectance Index (PRI) [6],

$$\mathrm{PRI} = (\mathrm{R}_{531} - \mathrm{R}_{570}) / (\mathrm{R}_{531} + \mathrm{R}_{570}) \tag{3}$$

and the Normalized Pigments Chlorophyll Ratio Index (NPCI) [6].

$$\mathrm{NPCI} = (\mathrm{R}_{680} - \mathrm{R}_{430}) / (\mathrm{R}_{680} + \mathrm{R}_{430}) \tag{4}$$

The most frequently used vegetation index is the NDVI. It was first described by Rouse et al. (1974) [6]. Studies have shown a strong correlation

between NDVI and plant primary productivity (PPP), biomass, and leaf area index (LAI) [6].

The MSARVI normalizes the spectral data for atmospheric attenuation and soil background influences, and the PRI and NPCI indices are sensitive to plant pigments (chlorophylls a and b, and carotenoid), which allow a more direct estimate of photosynthesis and biomass accumulation.

While these indices are appropriate to a specific problem, they cannot be generally applied to any classification problem. For example, though the NDVI performs very well in the classification problem of vegetation versus non-vegetation, it loses the information that distinguishes different vegetation species. When using NDVI as a classification feature, different vegetation species look spectrally similar, hence making inter-vegetation species classification a difficult task.

2.3 Optimal Spectral Subset

Attempts have also been made to find the location of specific wavelengths that are affected differently by various plant species. The hyperspectral reflectance at these specific wavelengths could be used as classification features. In 1999, Compact Airborne Spectrographic Imager (CASI) sensor data was analyzed to extract subsets of spectral bands in the electromagnetic spectrum [7]. Areas of greatest difference in amplitude between the spectral signatures for individual species, as recorded by the handheld spectral radiometer, were identified. Various algorithms were examined to compare their ability to extract

the best spectral bands for target material classification. Three such algorithms were the Best Individual Feature algorithm, the Forward Sequential or (1, 0) algorithm, and the Greedy (2, 1) algorithm [8]. These methods were studied to determine the fewest bands to achieve a specified performance goal for pixel classification. The basic need for this was to reduce computation time and not necessarily to increase classification accuracies.

2.4 Wavelet Based Features

More recently, wavelet-based features have been investigated to be used as features for classification [9, 10, 11]. In order to study the partition of the energy of the original signal according to scale or resolution, the energies of wavelet decomposition coefficients were used as features. The energy at a particular scale was calculated by taking the mean square value of the wavelet coefficients at that scale. Although this method is advantageous in reducing the features to a very small number, it loses the translation information. There could be vital information residing in a particular scale at a particular translation. The above mentioned approaches do not exploit the individual coefficients of a particular scale.

There still is a need to extract only those features that contribute toward the differences between the electromagnetic spectrum of different plant species and a system that can use them collectively to accurately classify subtly different vegetation species.

CHAPTER III

BACKGROUND

The main building blocks in a discrimination and classification algorithm are feature extraction, feature reduction and classification, as shown in Figure 11. For feature extraction in this thesis, two primary techniques were used: reflectance amplitudes and Discrete Wavelet Transform (DWT) coefficients. For feature reduction, Receiver Operating Characteristic (ROC) curves and Linear Discriminate Analysis (LDA) were used. For classification, the nearest mean classifier and the nearest neighbor classifier were used. Details of these methods are explained in the following sections. Also included in this chapter is information on two hyperspectral sensors (ASD and Hyperion).

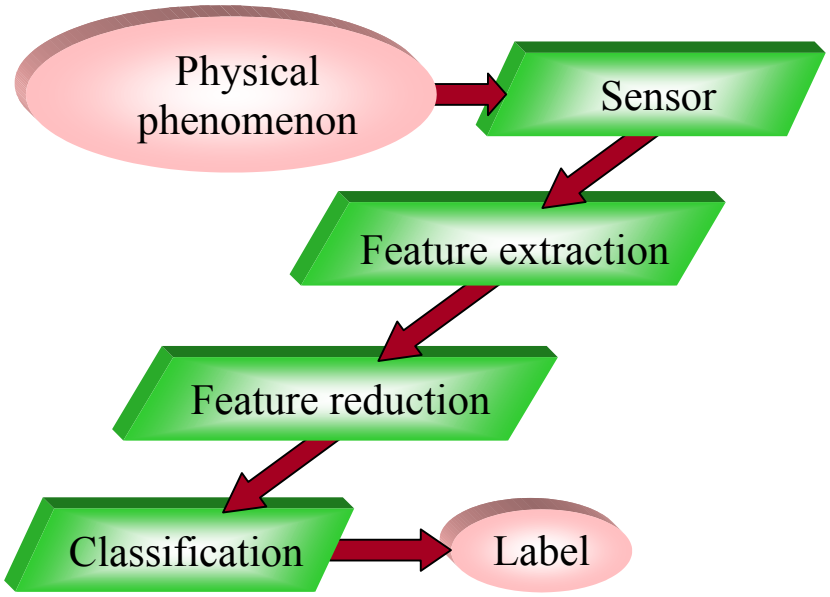

Figure 11: Block diagram for a pattern recognition system.

3.1 Discrete Wavelet Transform

Multi resolution transform can provide a domain in which both time and scale information can be studied simultaneously giving a time-scale representation of the signal under observation. This is of great advantage in case a non-stationary component of the signal is of particular interest. In these cases, it may be very beneficial to know the time intervals these particular components occur. A point to be noted here is that generally, wavelet transforms are studied in the time – scale context. In this study, however, the analyzed signals are in the wavelength (λ) domain rather than the time domain. However, the signals are analyzed using the convenient time-scale approach.

A wavelet transform can be obtained by projecting the signal onto shifted and scaled versions of a basic function. This function is known as the mother wavelet, $\Psi(\lambda)$, and the shifted and scaled versions of the mother wavelet form a basis of functions. These basis functions can be represented as

$$\psi_{a,b}(\lambda) = \frac{1}{\sqrt{a}}\psi\left(\frac{\lambda - b}{a}\right) \tag{5}$$

where, $a > 0$ is the scaling variable and b is the shift variable, both being real. The multiplier ($1/\sqrt{a}$) normalizes the energy of the wavelets. The wavelet expands, or dilates, if $a > 1$ and contracts, or shrinks, if $a < 1$. For the dyadic DWT, the scale variables are powers of 2 and the shift variables are non overlapping and discrete. The input signal is also discrete. There are certain conditions that must be fulfilled before a function can be deemed a "mother wavelet". These are known as the "admissibility conditions" and require that (*i*) the function must be oscillatory in

nature such that it has a zero average value and (*ii*) it must have finite support, i.e. its extremities must be finite and of zero value. One property that most wavelet systems satisfy is the multiresolution analysis (MRA) property. This is not a necessary condition for wavelet transforms. However, it provides a detail and approximation form of decomposition of the signal, which can be implemented using a recursive filter tree. In the case of the DWT, the signal decomposition can be implemented with a dyadic filter tree [12]. The dyadic filter tree is shown in Figure 12.

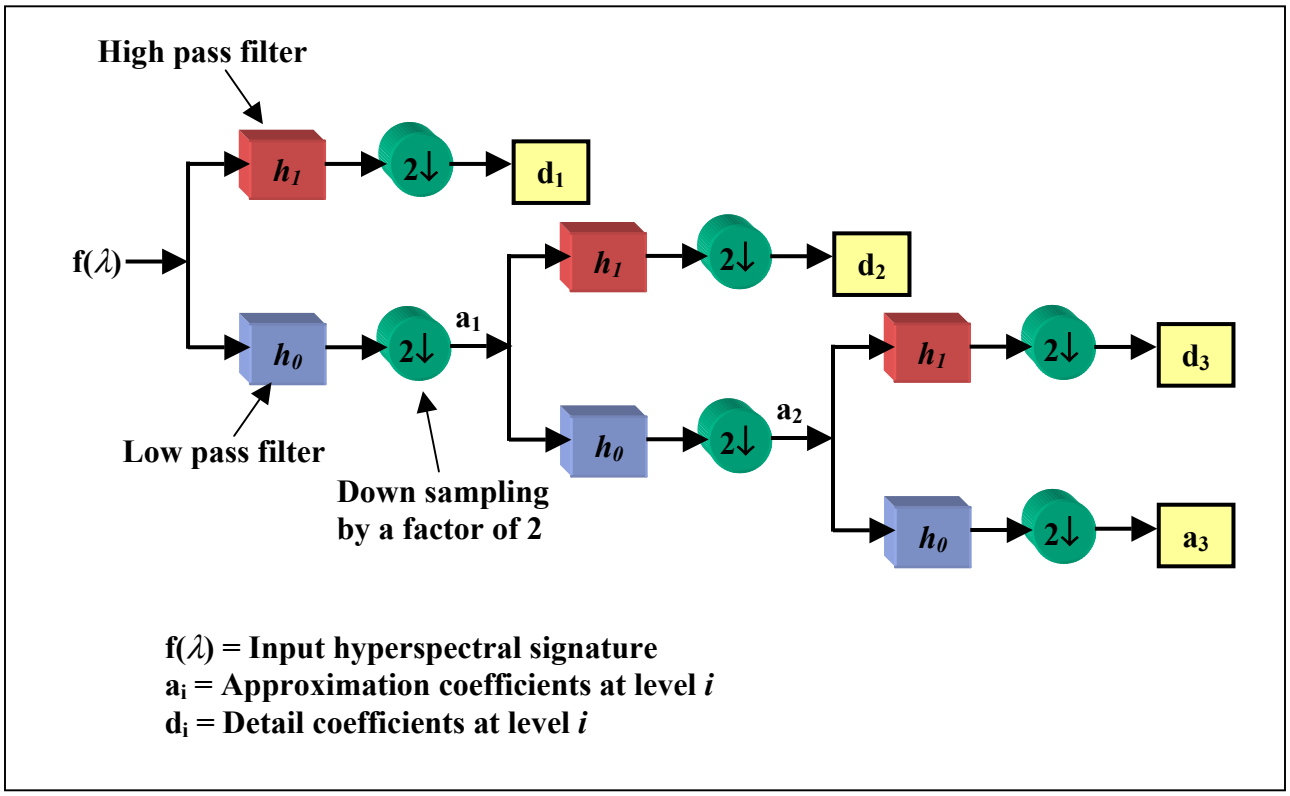

Figure 12: The dyadic filter tree implementation for a level-3 DWT.

This form of DWT allows the use of fast algorithms (like the ones used for the fast Fourier transform). The highpass and lowpass filters of this tree are designed according to the selection of the mother wavelet. As the signal passes through each stage of the filter tree, a set of detail and approximation coefficients are produced, corresponding to the small and large-scale behavior of the input signal, respectively. In theory, the decomposition can be repeated to $p = \log_2(N)$ levels, where N is the length of the input signal. In practice, however, the maximum number of decomposition levels also depends on the choice of a mother wavelet, *i.e.* $p = \log_2(N/l)$, where l is the length of the lowpass and highpass filters. This can be explained as follows. Let the length of the lowpass and highpass filters be *l*. Let the length of the input signal be *N*. If the input signal is decomposed once, the length becomes *N/2*. Also, if the input signal is decomposed p times, the length becomes $N/(2^p)$. For effective filtering, the length of the incoming signal should be greater than or equal to the length of the filter. In the limiting case, the two lengths should be the same. Thus:

$$N/(2^p) = l \tag{6}$$

or

$$p = \log_2(N/l) \tag{7}$$

For this study, the number of levels is chosen such that p is maximized for each mother wavelet investigated.

In this study the mother wavelets used for the univariate feature extraction and classification approach (Section 4.2) were Haar, Daubechies2 (Db2), Daubechies3 (Db3), Daubechies5 (Db5), Daubechies7 (Db7), Daubechies8 (Db8),

Daubechies15 (Db15), Symlet2 (Sym2), Symlet5 (Sym5), Symlet8 (Sym8), Coiflet1 (Coif1), Coiflet3 (Coif3), Coiflet5 (Coif5) and Discrete Meyer wavelet (Dmey). These mother wavelets are displayed in Figure 13. Among these, the 'Haar' and the 'Daubechies15' were found to perform best for the inter-plant species classification problem. These were further investigated using the multivariate feature extraction and classification approach (Sections 4.3 and 4.4).

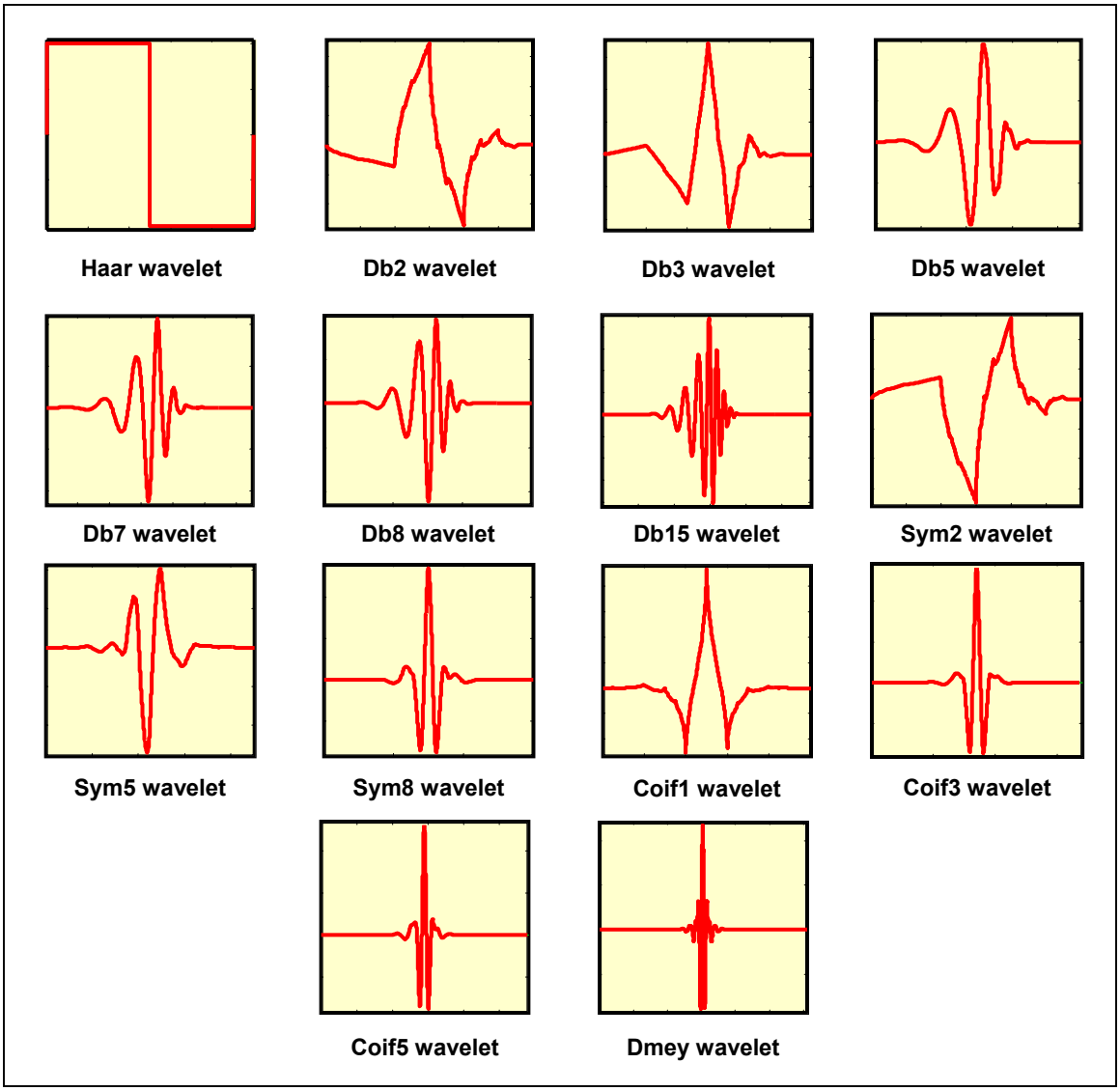

Figure 13: Various mother wavelet functions used for the univariate method.

The shape of the reflectance spectra is a key aspect in the classification problem as different materials have hyperspectral curves with different shapes. For example, the sharpness of the red-edge in a hyperspectral signature is dependent on the particular material under observation. The advantage of performing a wavelet transform as a preprocessing stage is that the scale and shift are both preserved. Thus the localized behavior and the overall shape of the signal can both be analyzed in one single domain.

3.2 Receiver Operating Characteristics

ROC curves are utilized to estimate the ability of a feature to discriminate among the data samples of two classes (w_1, w_2). The classifier employs a hard threshold to classify this data.

Let there be some data samples (features) for two classes, where w_2 is the target materials' class and w_1 is the class representing all other materials. Let x^* be the threshold for the classification, and x be a randomly selected sample (feature value) from the data. Then, the following four probabilities can be defined as shown in Figure 14:

$P(x > x^* \mid x \in w_2)$: a hit, *i.e.* a data sample of w_2 is classified as that of w_2.

$P(x > x^* \mid x \in w_1)$: a false alarm, *i.e.* a data sample of w_1 is classified as that of w_2.

$P(x < x^* \mid x \in w_2)$: a miss, *i.e.* a data sample of w_2 is classified as that of w_1.

$P(x < x^* \mid x \in w_1)$: a correct rejection, *i.e.* a data sample of w_1 is classified as that of w_1.

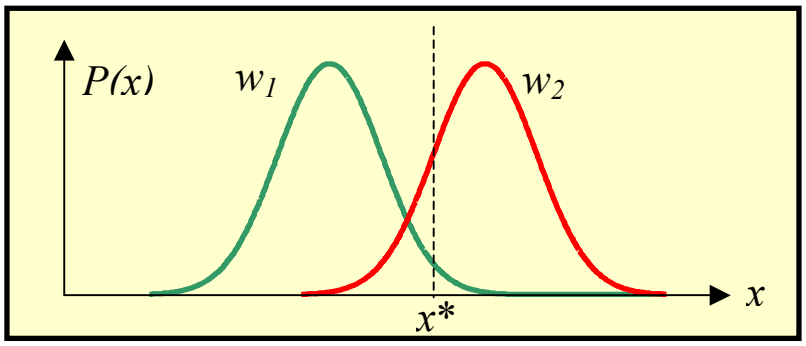

Figure 14: Probability density functions for the two classes

If a large data set is available, one can experimentally determine these probabilities. To do so, increase x^* from the start of w_2 to the end of w_2 and plot the correct rejection rate against the miss rate. This gives a smooth curve known as the ROC curve. This is shown in Figure 15.

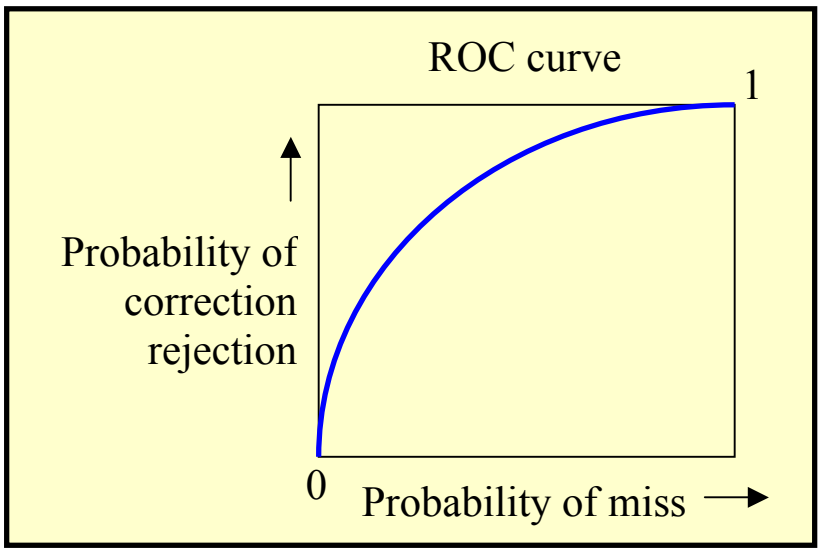

Figure 15: ROC curve.

In the worst case, the two probability density functions will be in complete overlap with each other, as shown in Figure 16. In this case, there is no means to separate the two classes. Also, one can easily see that in this case the ROC curve

will be a straight line as the two probability density functions are the same. Hence the area under the curve will be 0.5, and it will be the minimum possible.

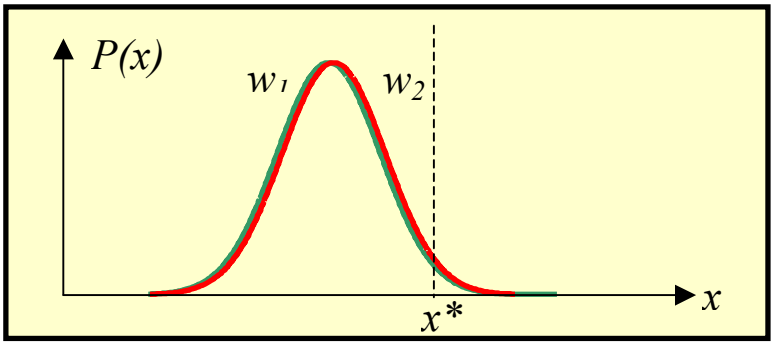

Figure 16: Probability density functions for the two classes with complete overlap

In the best case, the two probability density functions will be well separated, with no overlap, as shown in Figure 17. Hence, the correct rejection rate will always be 1.0 thus making the ROC curve a constant function equal to 1.0. As a result, the area under the ROC curve, A_z, will be 1.0 and will be the maximum that is possible.

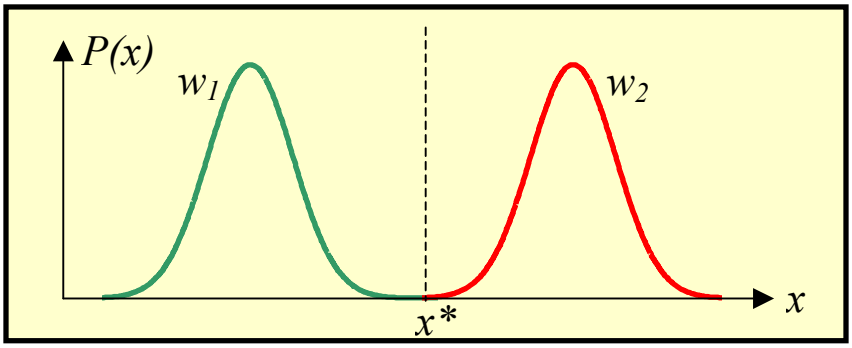

Figure 17: Probability density functions for the two classes with no overlap

Thus the area under the ROC curve ranges between 0.5 (no class separation – worst case) to 1.0 (complete class separation – best case) and gives

an estimate of the ability of the feature under observation to discriminate between the two classes.

3.3 Linear Discriminant Analysis

LDA is a commonly used technique to find the optimum surface on which to project the features of different classes to obtain maximum separation between the classes. Fisher's LDA for the two class problem is described below [13].

Let there be dimensional data (multiple features) for two classes that is projected on to a subspace of any arbitrary direction, the outcome usually will be a confused mixture of samples from all classes. The goal is to separate the two classes. Thus, there exists a need to find the optimal direction of that subspace so as to obtain well separated projections.

Let there be a set of *n* feature vectors (x1, x2, …, xn), each of dimension [*1* x *d*]. Assume n_1 of the feature vectors are in the subset D_1 and n_2 of the feature vectors are in the subset D_2. A linear combination of the components of $\vec{x}_j$ is formed to obtain the scalar dot product

$$y_j = \vec{w}^t \vec{x}_j \tag{8}$$

where, $\vec{w}$ is a (1xd) set of weights and y_j is scalar. If $\|\vec{w}\| = 1$, y_j becomes the projection of the corresponding $\vec{x}_j$ onto a line in the direction of $\vec{w}$. While the magnitude of $\vec{w}$ is not of great significance as it just scales the values of y, the direction of $\vec{w}$ is important. This is because as $\vec{w}$ is varied, the subspace is rotated, and the relative positions of the projections vary. It is desired that the

projections falling on to the subspace be well separated. Figures 18 and 19 show the effect of choosing two different directions of $\vec{w}$ for a $d = 2$ case.

To find this optimal value of $\vec{w}_j$, first the difference of the sample means is considered. Let $\vec{m}_j$ be the (*1 x d*) mean vector for class *i* given by:

$$\vec{m}_i = \frac{1}{n_i} \sum_{\vec{x} \in Di} \vec{x}_j \,. \tag{9}$$

then, the sample mean of the projected data is given by

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y_j \in Yi} y_j \,. \tag{10}$$

$$= \frac{1}{n_i} \sum_{x_j \in Di} \vec{w}_j{}^t \vec{x}_j = \vec{w}^t \vec{m}_i \,. \tag{11}$$

This is the projection of $\vec{m}_i$. This leads to the distance between the projected means given by

$$| \tilde{m}_1 - \tilde{m}_2 | = | \vec{w}^t (\vec{m}_1 - \vec{m}_2) | \,. \tag{12}$$

This difference should be large relative to some measure of the standard deviation of each class. Thus, scatter for the projected samples can be defined as:

$$\tilde{s}_i^2 = \sum_{y_j \in Y_i} (y_j - \tilde{m}_i)^2 \,. \tag{13}$$

Thus, the estimate of the variance of all the data is given by

$$(1/n)(\tilde{s}_1^2 + \tilde{s}_2^2) \tag{14}$$

and $(\tilde{s}_1^2 + \tilde{s}_2^2)$ is termed the within class scatter of the projected samples. The *Fisher's linear discriminant* maximizes the criterion function given by:

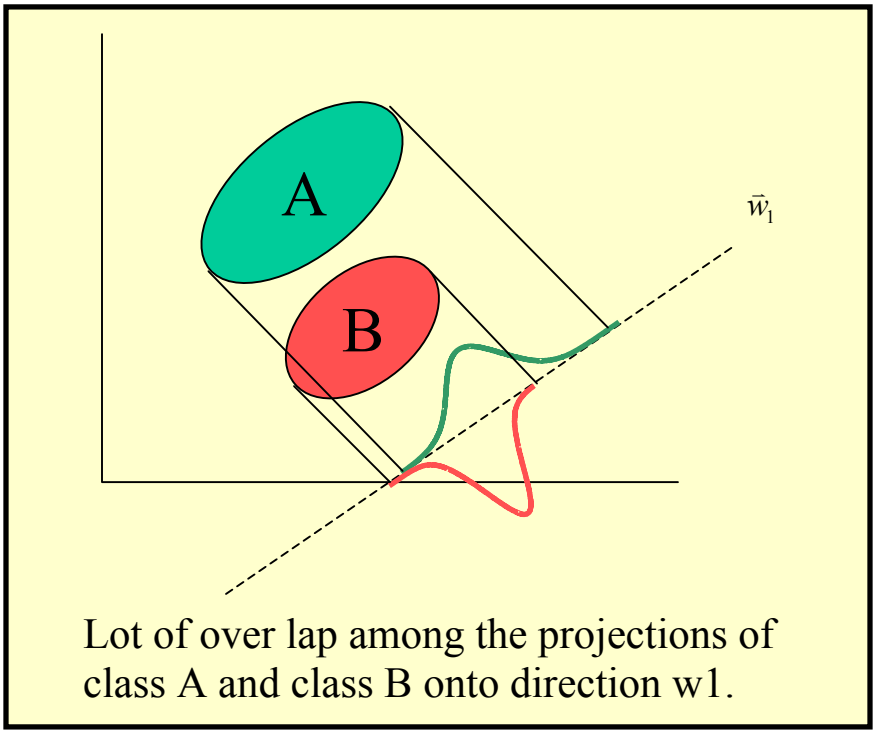

Figure 18: Two classes with much overlap when projected onto $\vec{w}_1$.

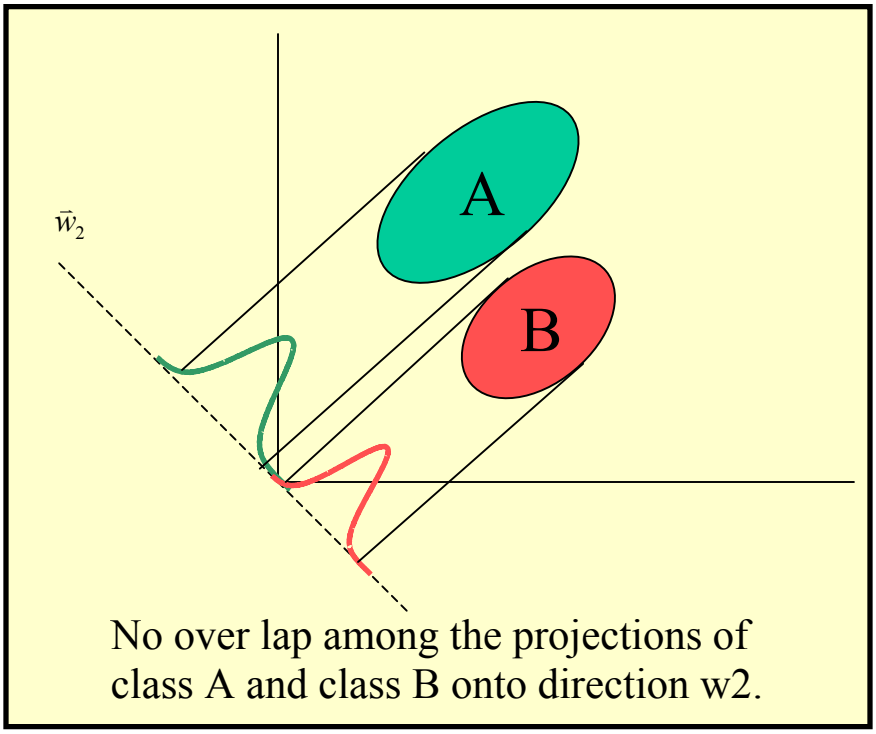

Figure 19: Two classes with no overlap when projected onto $\vec{w}_2$.

$$J(w) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}. \tag{15}$$

To obtain $J(w)$, the scatter matrices S_i and S_w are defined as

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^t \tag{16}$$

and,

$$S_W = S_1 + S_2. \tag{17}$$

Then,

$$\tilde{s}_i^2 = \sum_{\vec{x}_j \in D_i} (\vec{w}^t \vec{x}_j - \vec{w}^t \vec{m}_i)^2 \tag{18}$$

$$= \sum_{x_j \in D_i} \vec{w}^t (\vec{x}_j - \vec{m}_i)(\vec{x}_j - \vec{m}_i)^t \vec{w} \tag{19}$$

$$= \vec{w}^t S_i \vec{w}. \tag{20}$$

Therefore, the sum of these scatters can be given by

$$\tilde{s}_1^2 + \tilde{s}_2^2 = \vec{w}^t S_W \vec{w}. \tag{21}$$

Similarly, the difference between the projected means is given by

$$(\tilde{m}_1 - \tilde{m}_2)^2 = (\vec{w}^t \vec{m}_1 - \vec{w}^t \vec{m}_2)^2 \tag{22}$$

$$= \vec{w}^t (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^t \vec{w} \tag{23}$$

$$= \vec{w}^t S_B \vec{w} \tag{24}$$

where,

$$S_B = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^t. \tag{25}$$

S_w is known as the within class scatter matrix, and S_B is called the between class scatter matrix. For any $\vec{w}$, $S_B \vec{w}$ is in the direction of $(\vec{m}_1 - \vec{m}_2)$. Now the criterion function becomes

$$J(w) = \frac{\vec{w}^t S_B \vec{w}}{\vec{w}^t S_W \vec{w}}. \quad (26)$$

This is of the form of the generalized Rayleigh quotient. It follows that a vector *w* that maximizes the above criterion must satisfy

$$S_B \vec{w} = \lambda S_W \vec{w} \quad (27)$$

for some constant λ. For a non singular S_W,

$$S_W^{-1} S_B \vec{w} = \lambda \vec{w}. \quad (28)$$

Now $S_B \vec{w}$ is in the direction of $(\vec{m}_1 - \vec{m}_2)$ and that direction is of more interest than the scale of $\vec{w}$. Therefore, it is not necessary to actually solve for the eigenvalues and eigenvectors. Hence,

$$\vec{w} = S_W^{-1}(\vec{m}_1 - \vec{m}_2) \quad (29)$$

For this value of $\vec{w}$ for Fisher's Linear Discriminant, the maximum ratio of between class scatter to the within class scatter is obtained.

3.4 Analytical Spectral Devices (ASD) spectroradiometers

Field spectroradiometers can be used to quantitatively measure *in situ* radiance, irradiance, reflectance or transmission as a function of wavelength for a given target. Portable, battery-powered spectroradiometers are used to make these measurements. An example is shown in Figure 20.

Figure 20: Example usage of a field spectroradiometer.

One example of a spectroradiometer is the ASD. The ASD spectroradiometers can be used to measure reflectance for wavelengths in the range of 350 to 2500 nm. The measurements can be used to examine geological, man-made, and vegetative materials. The individual bands of the hyperspectral signature have a bandwidth of 1.4 nm [14]. This gives a very finely resolved and detailed version of the reflectance spectra that can be exploited to improve target discrimination and classification methods. More detailed specifications of the ASD spectroradiometers used to collect data for this study are given in Table 1.

Table 1

Specifications for the ASD FieldSpec® Pro FR spectroradiometers

Name	FieldSpec® Pro FR
Spectral Range	350-2500 nm
Spectral Resolution	3 nm @ 700 nm 10 nm @ 1400 & 2100nm
Sampling Interval	1.4 nm @ 350-1050 nm 2 nm @ 1000-2500 nm
Scanning time	100 milliseconds
Detectors	One 512 element Si photodiode array 350-1000 nm Two separate, TE cooled, graded index InGaAs photodiodes 1000-2500 nm
Input	1.4 m fiber optic (25° field of view Optional foreoptics available
Calibration	Wavelength, reflectance, radiance*, irradiance*. All calibrations are NIST traceable (*radiometric calibrations are optional)
Noise Equivalent Radiance (NeDL)	UV/VNIR 1.4 x 10^{-9} W/cm^2/nm/sr @ 700nm NIR 2.4 x 10^{-9} W/cm^2/nm/sr @ 1400nm NIR 8.8 x 10^{-9} W/cm^2/nm/sr @ 2100nm
Notebook Computer	Pentium processor, 800 MB hard disk, 16 MB Ram, 3.5" floppy disk drive, battery, AC power supply
Weight	7.2 kg or 15.8 lbs

Two terms associated with the design of spectroradiometers are "spectral resolution" and "spectral sampling interval". While the two terms are often used interchangeably, they refer to very different characteristics of a spectroradiometer. Spectral resolution is a measure of the narrowest spectral band that can be resolved by a spectroradiometer. It is also defined as the full width at half maximum (FWHM) response to a spectral line source. The spectral sampling interval of a spectroradiometer on the other hand is the interval, in wavelength units, between data points in the measured spectrum. This is shown in Figure 21.

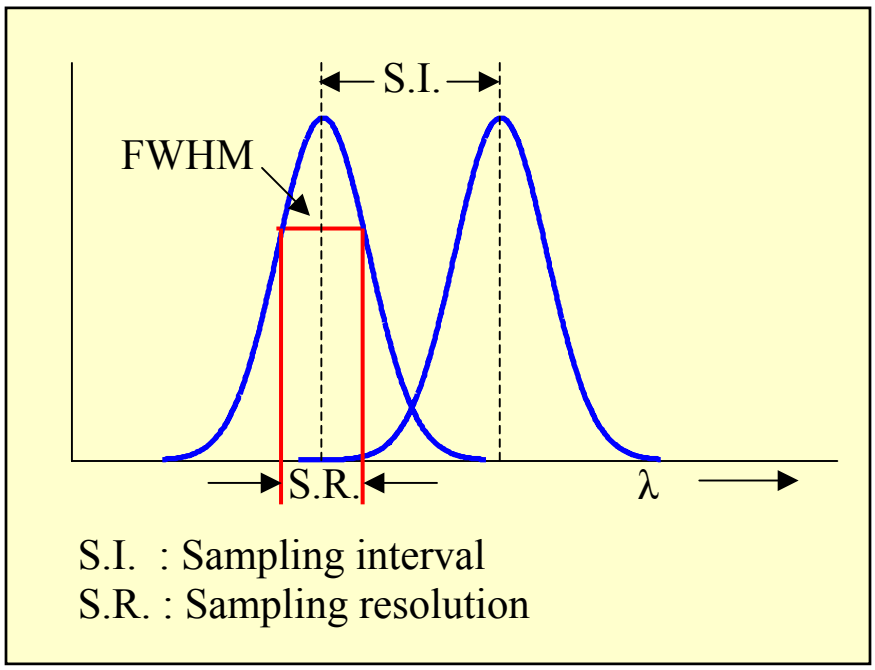

Figure 21: Sampling interval and sampling resolution.

Currently available spectroradiometers are based upon either (*i*) a fixed grating and an array detector or (*ii*) a single element detector and a scanning grating. The drawback of an array-based spectroradiometer is that the signal-to-noise ratio (SNR) is dependent on the sampling interval as well as the spectral resolution (Smith, 1992). In a scanning spectroradiometer, the SNR is independent of the spectral sampling interval.

In existing field spectroradiometers, two basic approaches are used to collect light energy and deliver it to the spectroradiometers. In some spectroradiometer designs, foreoptics are used to form an image of the target on the entrance slit of the spectroradiometers [14]. This approach results in a ground field of view (GFOV) that has the same shape as the spectroradiometer entrance slit (often a rectangle with a height to width ratio of more than 10:1). If more than

one sensor is used in an instrument, it becomes difficult to ensure that both sensors are viewing the same GFOV. A circular GFOV is obtained in case optical fibers are used to deliver the light to the spectroradiometer. Splitting the optical fiber bundle within the instrument to deliver light to the various sensors accommodates the use of more than one sensor. The size of the GFOV is determined by the angular field of view of the instrument and distance to the target. While optics can be added to modify the angular field of view of a field instrument, this is usually practical only for instruments that use optical fibers for light collection.

3.5 Hyperion sensor simulation

The Hyperion imaging spectroradiometer is the next generation of satellite imaging spectroradiometers. Compared to currently used instruments like Landsat, the Hyperion sensor has a very finely resolved spectral resolution and spectral sampling interval. This gives it the ability to have a large number of spectral bands in the observed wavelength range. Hyperion data can be used to exploit the reflectance spectrum in greater detail than previous satellite sensors. Remote sensing images will no longer be restricted to just a few bands as in the case of multispectral images. The specifications of the Hyperion sensor are given in Table 2 [15].

Table 2

Specifications of the Hyperion sensor

GSD* at 705 km Altitude	30 +/- 1 m
Swath Width (km)	7.5 km minimum
Spectral Coverage	0.4 - 2.5 ηm
Imaging Aperture	12.5 +/- 0.1 cm diameter
On-orbit Life	1 year (2 years goal)
Instantaneous Field of View	42.5 +/- 3.0 ηrad
Number of Spectral Channels	220 minimum
SWIR** Spectral Bandwidth	10 +/- 0.1 nm
VNIR*** Spectral Bandwidth	10 +/- 0.1 nm
Cross-track Spectral Error	<1.5 nm (VNIR), <2.5 nm (SWIR)
Spatial Co-registration	<20% of Pixel
Absolute Radiometric Accuracy	<6% (1 sigma)
Data Quantization	12-bit
Operability (SWIR, VNIR)	>98% each*
Grating Type	Convex
Detectors	CCD, VNIR and HgCdTe SWIR (60μm pixels)
Calibration options	Lamps, lunar, solar, ground imaging and laboratory
Sensor type	Push broom
Absolute radiometric accuracy	6%
Cycle orbit	16 day (8 day repeat for selected locations)
Radiometric resolution	12 bits

* GSD = Ground sample distance
** SWIR = Short-Wave-Infrared
*** VNIR = Visible and near-infrared

The available ASD data was used to simulate Hyperion sensor data taking the above specifications into account. This simulated data was analyzed with the proposed feature extraction and classification method to estimate performance on satellite data that will be available in the future.

3.6 Classifiers

In this study, two types of basic classifiers were implemented and used. These are the nearest mean classifier and the nearest neighbor classifier, which are briefly explained in the following sections.

3.6.1 Nearest Mean Classifier

The nearest mean classifier is a supervised method. Training data is used to calculate the class means for each class. Test data is then classified by comparing it to the class means. Various distance measures can be used for this method; in this thesis, Euclidian distance is used. The Euclidian distance between the test data and each class mean is computed using:

$$d = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(x_{0i} - \bar{x}_{ji})^2} \tag{30}$$

where,

d is the Euclidian distance,

x_0 is the test data,

x_{0i} is the i^{th} element of the test data,

$\bar{x}_j$ is the class mean of the j^{th} class,

$\bar{x}_{ji}$ is the i^{th} element of the class mean of the j^{th} class, and

M is the dimension of x_0 and $\bar{x}$.

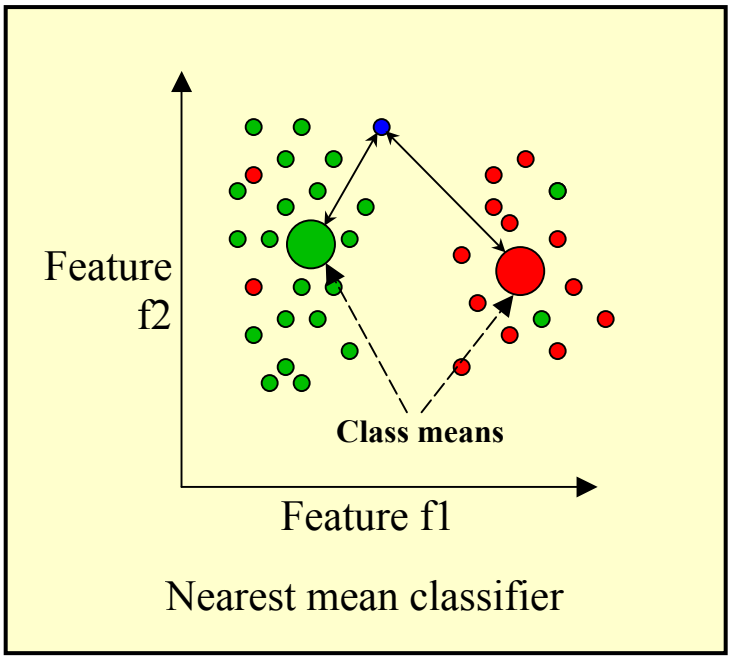

Figure 22: Nearest mean classifier for a two-feature problem.

Test data is then assigned to the class with the minimum distance, as illustrated in Figure 22.

3.6.2 Nearest Neighbor Classifier

The nearest neighbor classifier is a supervised method. The training data samples are directly used for classification of the test data. The Euclidian distance from the test data to each of the training data samples is computed. The test data sample is then assigned to the class whose training data sample has the shortest distance from it. This is illustrated in Figure 23.

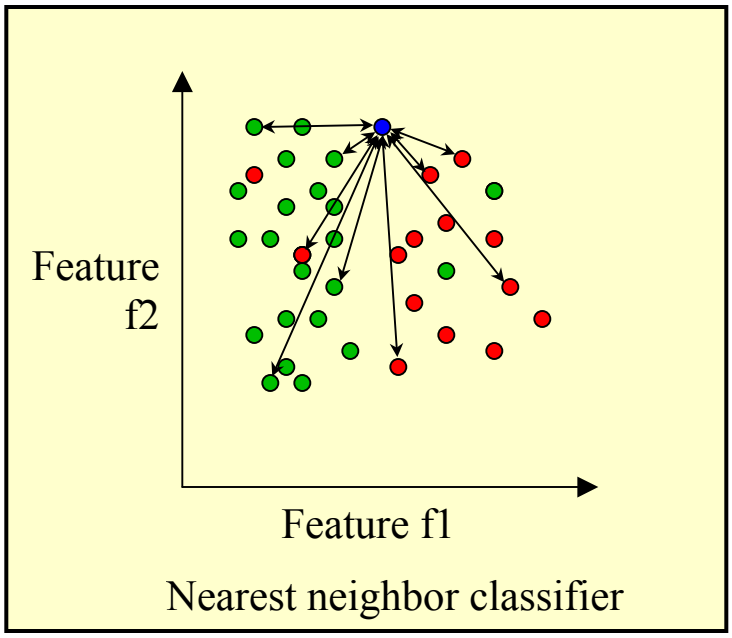

Figure 23: Nearest neighbor classifier for a two-feature problem.

3.7 Testing techniques

Two types of testing techniques were used to calculate the system accuracy: jack-knife testing and leave-one-out testing. These methods are briefly described in the following section.

3.7.1 Jack-knife testing

In the jack-knife testing, the entire data is divided into two groups: the training set and the testing set. This partitioning needs to be done in a random fashion, *i.e.* the data samples selected for training and testing are chosen in no specific order. This is done to eliminate any intrinsic biasing that the data may have simply as a result of the way it is collected and stored. Once the data is segregated into the two groups, the training set is used to determine the best

classification features. The training set is also used to train the classifiers. The testing set is then classified, and the accuracy is calculated as the percentage of the test data that is correctly classified. The advantage of this method is that it simulates a practical situation and is very fast. Also, it is absolutely unbiased as the training and testing sets are absolutely separate. The disadvantage of this method occurs when a small number of labeled data samples are available. Since only half of the available data is used for training, the results might not be statistically very significant.

3.7.2 Leave-one-out testing

In the leave-one-out testing, all but one sample of the data is used as the training set. This set is used to determine the best features and train the classifiers. The data sample that was left out is then classified. This process is repeated until each of the data samples has been 'left out' and classified. The final accuracy is calculated as the percentage of the test data that is correctly classified. The advantage of this method occurs when only a small number of labeled data samples are available. The method maximizes the size of the training set while producing an unbiased classification for the testing set. The disadvantage of this method is its relative computational complexity.

CHAPTER IV

METHODOLOGY

4.1 Data Collection

Analysis and classification were conducted on two types of hyperspectral data: ASD sensor data and simulated-Hyperion sensor data. The ASD data was collected for different species like Cogongrass, Bahiagrass, Bermudagrass, Broomsedge, Centipedegrass and Vaseygrass at various sites in Mississippi in late-spring and mid-summer of the year 2000, specifically, on March 23, April 18 and July 20, 2000. The water bands were corrected using a cubic spline interpolation to avoid discrepancies in training and testing of the system due to unreliable measurements in the water bands (highly variable large values). These large values arise due to the following reason. The earth's atmosphere acts as a transfer function for the reflected light reaching the sensor. There are many random factors that contribute towards this transfer function, such as water vapor, CO_2, CO, N_2, O_2, O_3, etc. that add environmental error to the data and make it a very complex problem to handle. To nullify these effects, a white reference reading is taken along with other observations. That is, the sensor is used to collect a reading of the reflectance of a pure white surface. A special white material is chosen for the reference, as it is known to reflect all wavelengths

equally. The resulting signature can be treated as analogous to the atmosphere's transfer function. Thus, the reflectance values of the target material are divided by the atmosphere's transfer function. Most of the energy in the water bands is absorbed by moisture in the atmosphere, hence the name. Therefore, for the white reference, the reflectance values in the water bands are very small. When the observations are divided by the white reference, the reflectance values in the water bands get divided by very small values. Any slight change in the water bands' reflectance results in very large values. Examples of the collected ASD hyperspectral signals are shown in Figure 24.

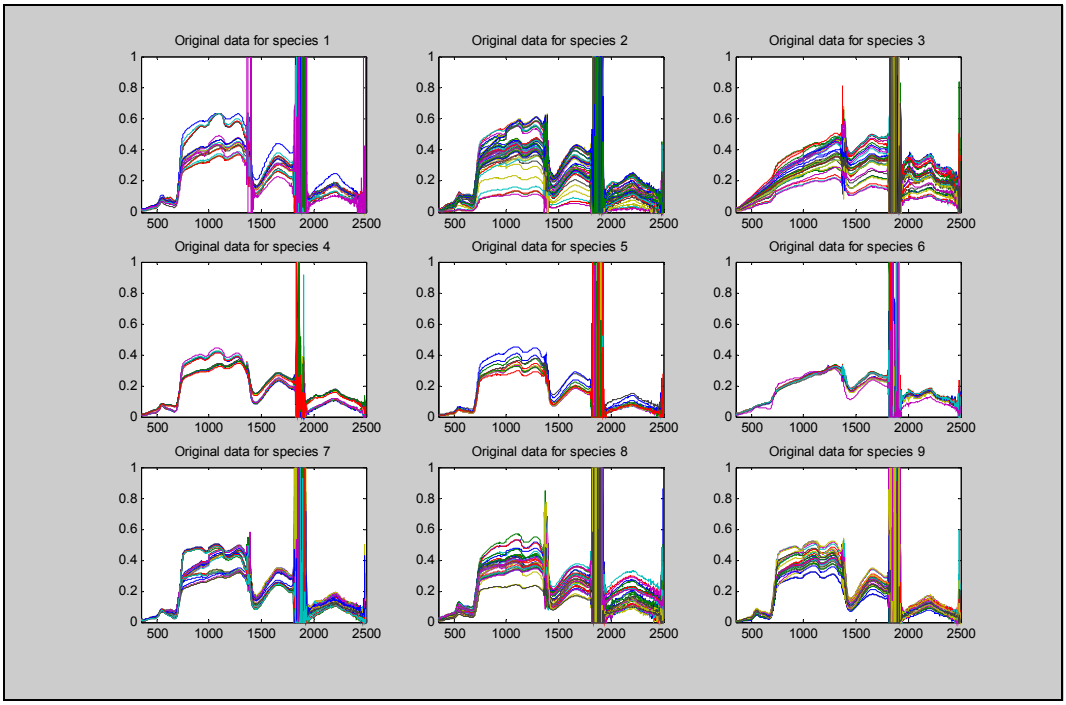

Figure 24: Original unprocessed data collected using an ASD sensor.

Correction for the water bands is done by either removal (replacing samples with zero value) or by interpolation. For this thesis, cubic spline interpolation was used. The resulting signatures are shown in Figure 25.

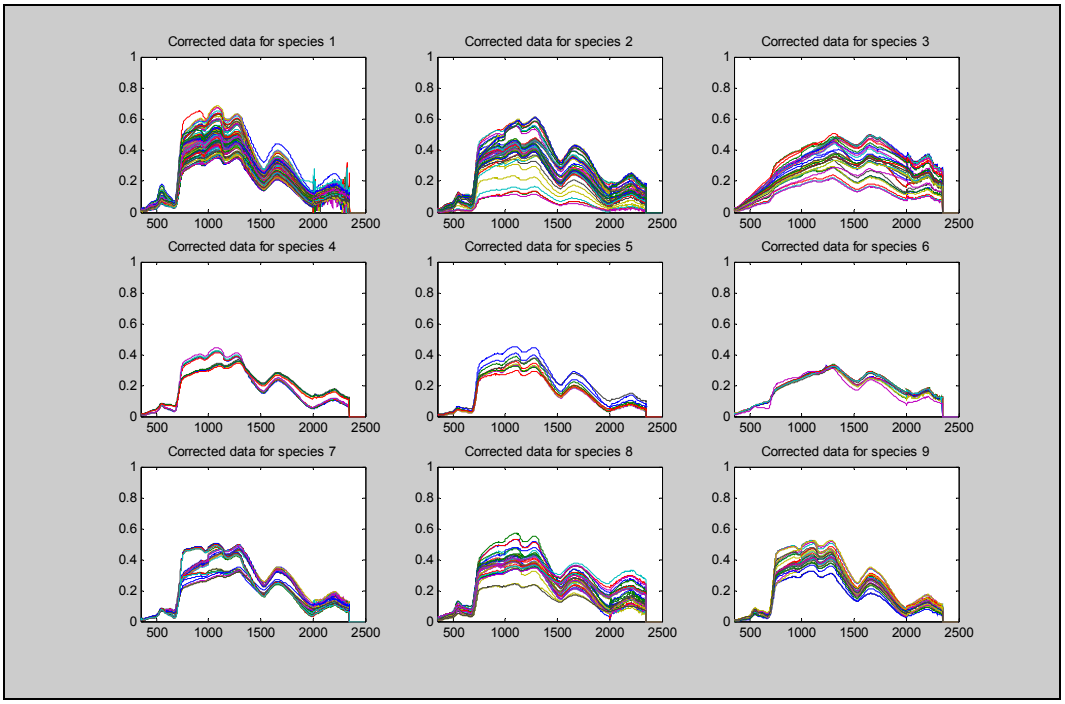

Figure 25: Data corrected for waterbands and other bad data removed.

Next, the ASD signatures were used to create simulated Hyperion data. Every ten bands of the ASD signatures were averaged, taking into account the exact band limits of the Hyperion sensor. Analysis was conducted on this dataset to estimate the future performance of the algorithm using Hyperion data. Example simulated Hyperion signatures are shown in Figure 26.

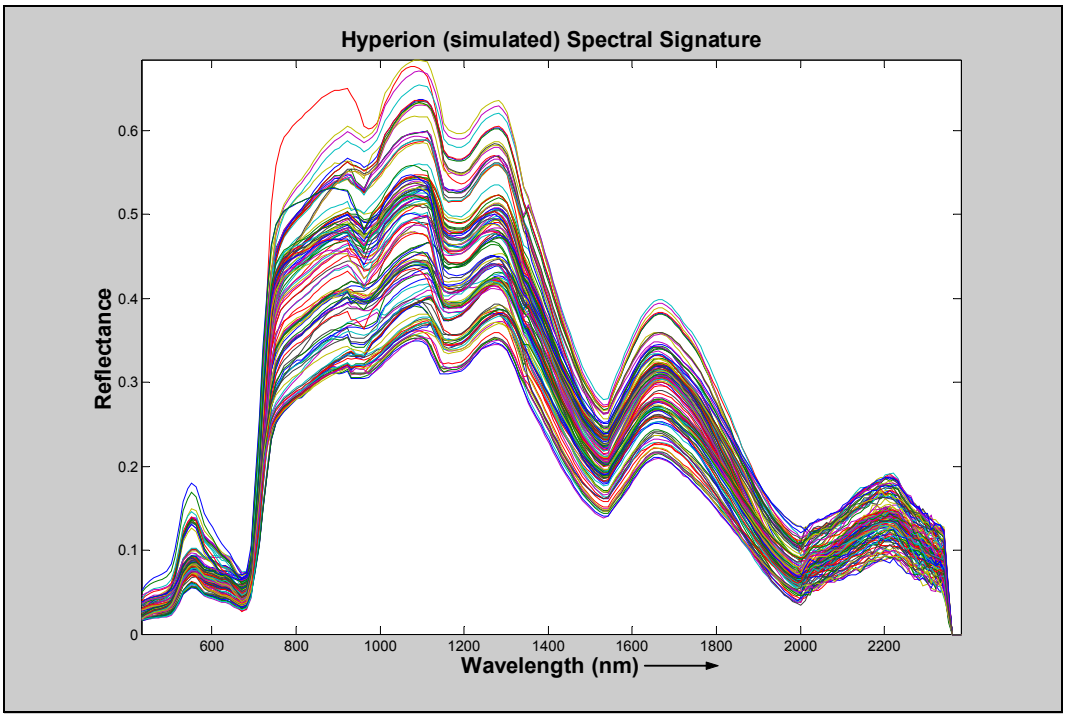

Figure 26: Data simulated according to the specifications of the Hyperion sensor.

4.2 "Best feature" selection – Univariate approach

There is a need to identify and extract the most useful information from a hyperspectral signal, where utility is measured in terms of signal classification. For this reason, instead of using the entire hyperspectral signature for cogongrass detection, pertinent features were extracted to reduce the dimensionality of the signature. The classification accuracy increases if the distributions of the classes are statistically more separate. One method to measure the ability of a feature to discriminate between two classes is to calculate the area under the feature's ROC curves. To estimate the classification capabilities of the different hyperspectral bands, ROC areas were calculated for each spectral band in the spectrum. This gave a plot as shown in Figure 27.

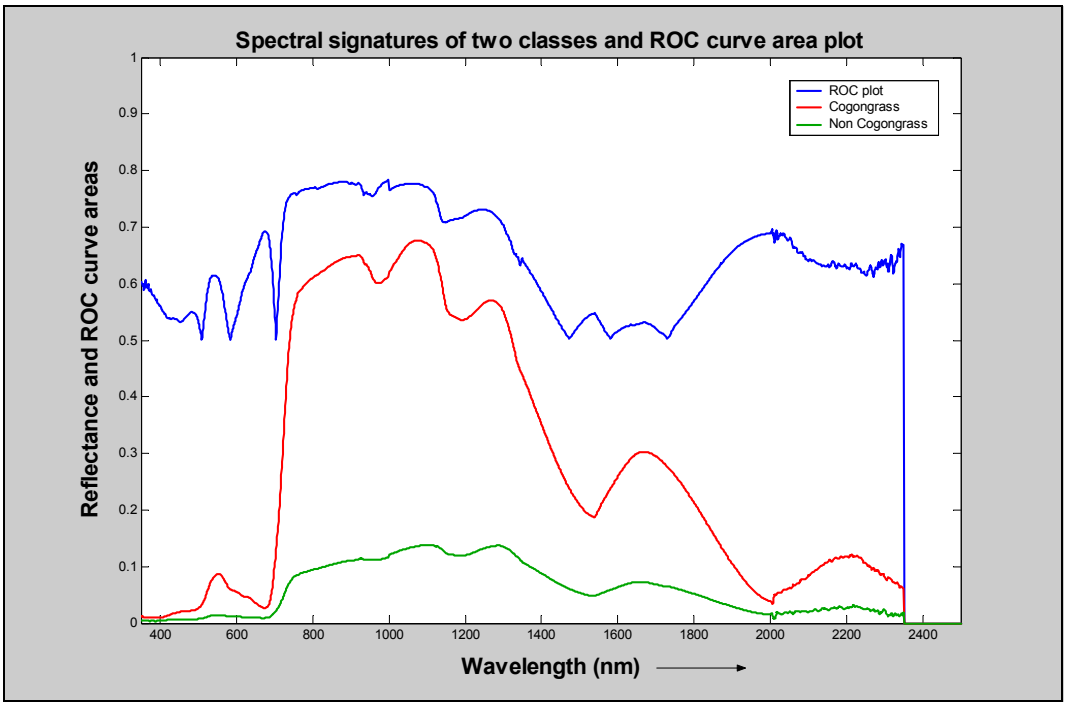

Figure 27: ASD signatures of two classes and their ROC plot.

Spectral bands were sorted in decreasing order of the ROC areas. Top *N* bands with the largest ROC areas were chosen as the best features. These bands were then combined using LDA to obtain an optimum scalar quantity. This one scalar quantity was expected to have a very high ROC area. For different values of *N*, the classification accuracy was calculated. It was observed that the results were not drastically better than the case when all the bands were used for classification. It was also noted that the ROC values of the combined reduced *N*-features were not much higher than the highest individual band ROC curve area. Moreover it was observed that there was not a regular pattern in the final ROC area values as *N* increased. These observations motivated the use of different combinations of spectral bands and not necessarily just the top N bands. Thus, a multivariate approach was investigated.

4.3 Off-line "best feature" selection – multivariate approach

For the multivariate approach, two types of features were separately investigated: (*i*) amplitude of critical spectral bands and (*ii*) pertinent discrete wavelet transform (DWT) coefficients. To perform DWT, a mother wavelet must first be selected. These mother wavelets were selected based on some preliminary experiments. These experiments were direct classifications without any feature extraction or reduction. These classification results are shown in Table 3. This led to the selection of the Haar and the Daubechies15 mother wavelets for further analysis.

In each case, (*i*) and (*ii*), an optimum subset of the initial features was determined. The following description of this offline "best feature" selection process was completed for each case, (*i*) and (*ii*), separately. The overall method is shown in Figure 28. First, ROC curves were obtained for all of the initial features for the two-class system, the two classes being $D_2 \equiv$ cogongrass and $D_1 \equiv$ non-cogongrass. A plot of the ROC area, A_Z, versus each individual feature is shown in Figures 29, 30 and 31.

Table 3

Preliminary classification results using wavelet decomposition coefficients

Wavelet decomposition and Jack-knife nearest mean classifier Confidence interval calculated at 99.5 % accuracy					
Mother Wavelet	**Level of decomposition**	**Class A %**	**Class B %**	**Overall %**	**Confidence interval**
haar	11	92.42	92.45	92.44	5.20
db2	9	86.36	94.34	91.28	5.55
db3	8	81.82	91.51	87.79	6.44
db5	7	74.24	94.34	86.63	6.70
db7	7	81.82	96.23	90.70	5.71
db8	7	69.70	97.17	86.63	6.70
db15	6	83.33	96.23	91.28	5.55
sym2	9	72.73	93.40	85.47	6.93
sym5	7	81.82	94.34	89.53	6.02
sym8	7	66.67	99.06	86.63	6.70
coif1	8	77.27	92.45	86.63	6.70
coif3	6	92.42	92.45	92.44	5.20
coif5	6	90.91	93.40	92.44	5.20
dmey	5	80.30	95.28	89.53	6.02

An algorithm was designed and implemented to obtain the best combination of the individual features to serve as a "best feature". First, the initial features were sorted in descending order of the ROC curve areas, A_Z. The best individual feature, the one with the largest A_Z, is placed into the "best features" vector, where its corresponding ROC area is A_{Z1}. The second best feature is appended to the feature vector. LDA is used to reduce the feature vector down to a scalar, and A_{Z2} is computed for the new scalar. If and only if $A_{Z2} > A_{Z1}$, the second best feature is also retained and added to the list of best features. Next, the third best feature is appended to the feature vector, and A_{Z3} is computed. If and only if $A_{Z3} > A_{Z2}$, then the third best feature is retained. This is continued until all

features have been tested for their ability to add to the discriminating power of the feature vector. In the end, we have the “best features”.

Various subsets of the data were used to find the best features to determine the consistency in the selection of the features, or robustness of the feature selection method. These subsets were created using cross validation, or leave-one-out techniques. This technique generated a large number of different data sets to be used for the above algorithm. A best band selection histogram was created that showed the probability of a particular feature being selected as one of the best features. If a feature had a value of 1.0 on the distribution, it had been selected in the list of best features every time. If a feature showed up as 0.5, it had been selected as one of the best bands only half of the time. These distributions are shown in Figures 32, 33 and 34.

A threshold was then manually selected for the probability distribution. The features, whose probability is greater than the threshold, are selected to form the reduced set of optimal features. LDA was then performed on the training data set using the final set of best features. The LDA weights were saved to be used for the classification of an incoming test signature. Note that the entire process of the best feature selection is conducted on the training data before a test hyperspectral signal is input to the system. Thus, the best feature selection is referred to as “off-line”.

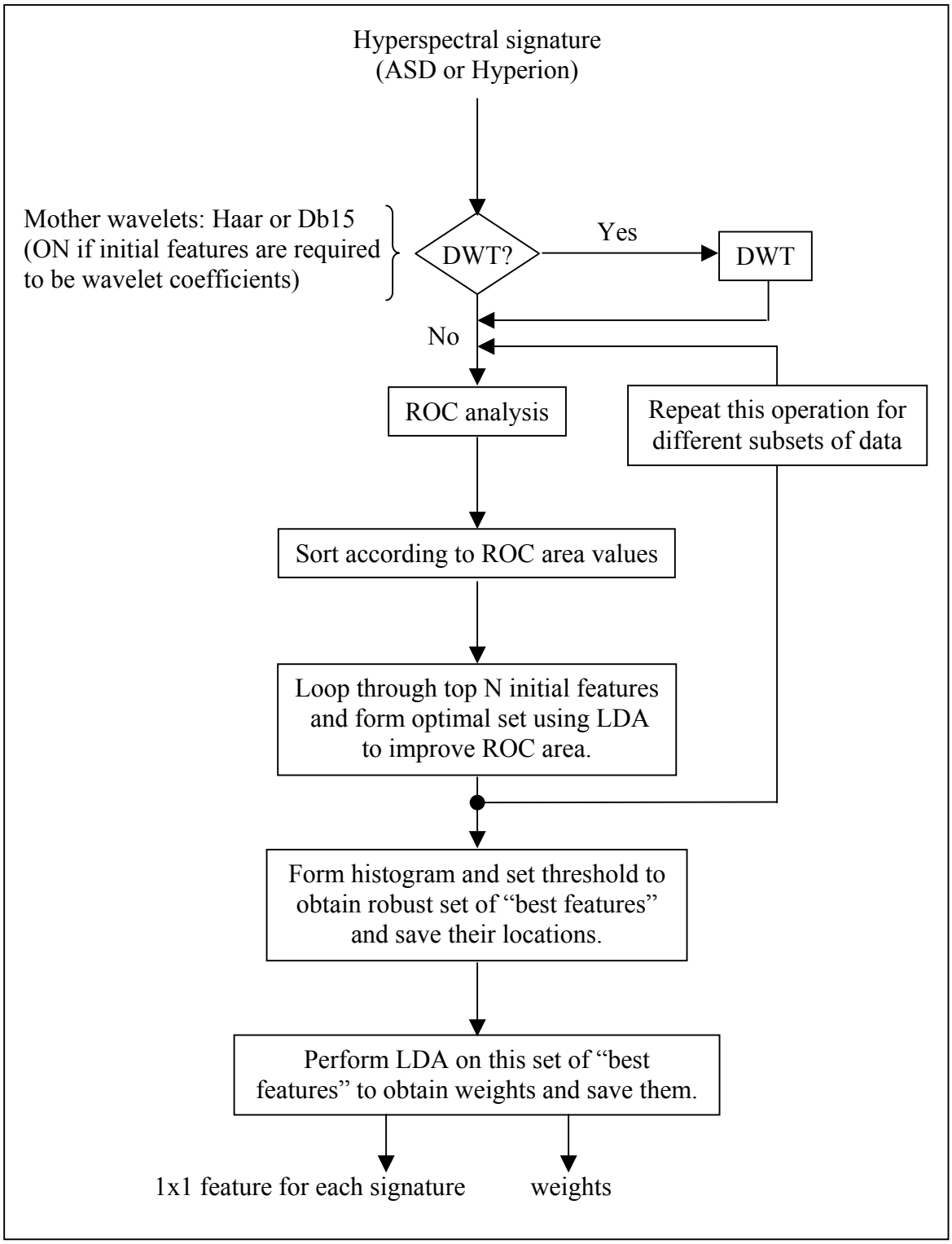

Figure 28: Multivariate best band selection algorithm.

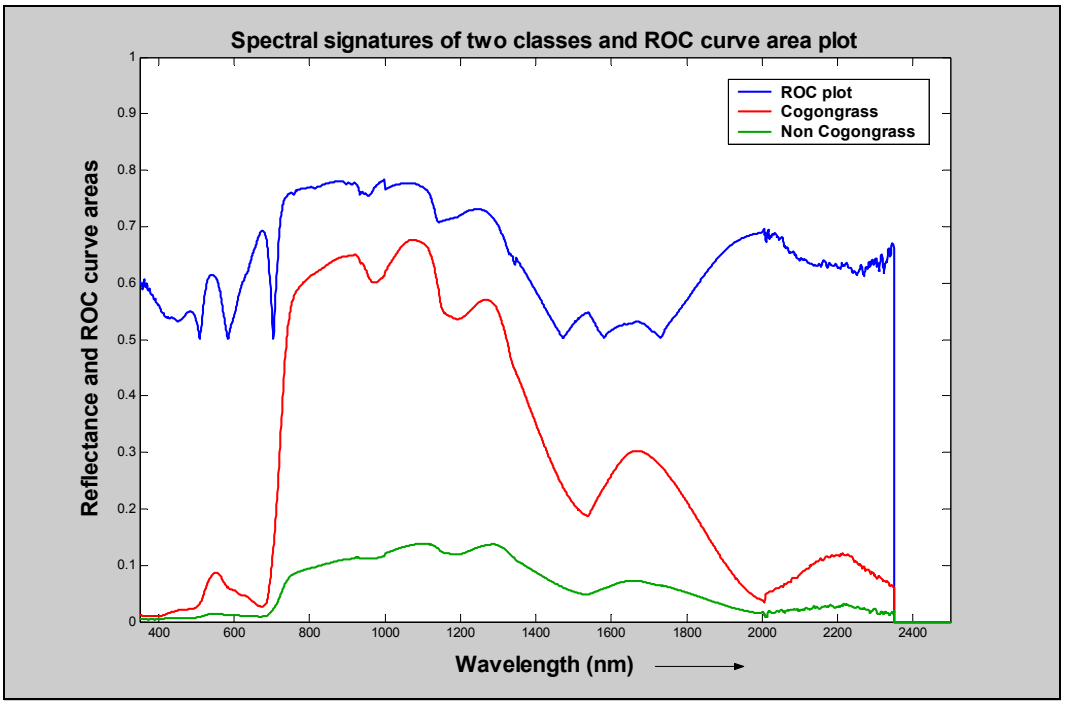

Figure 29: Example ASD signatures of cogongrass and non cogongrass along with a plot of ROC areas for each individual spectral band.

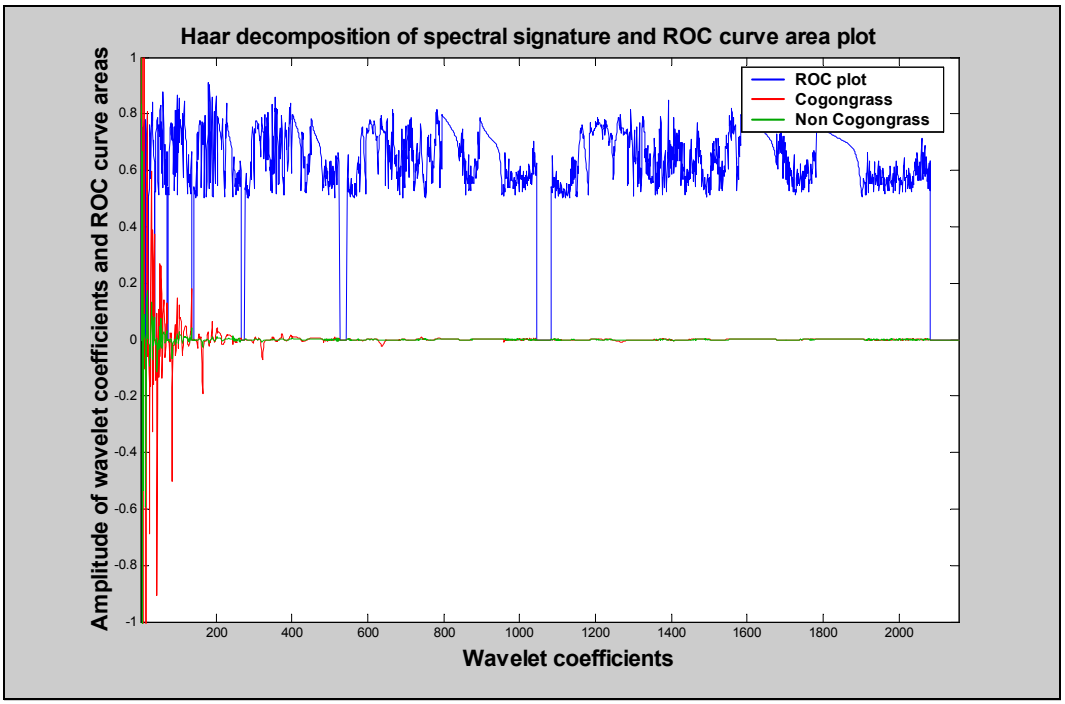

Figure 30: Example Haar decomposition coefficients of ASD signatures of cogongrass and non cogongrass along with a plot of ROC areas for each individual coefficient.

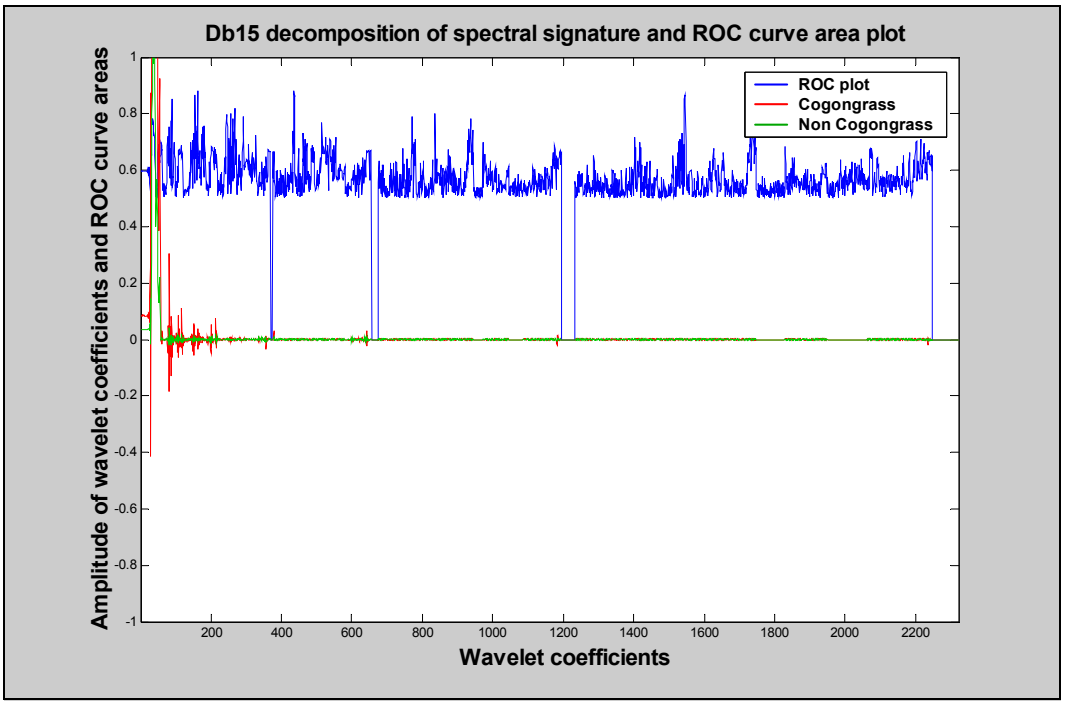

Figure 31: Example Daubachies15 decomposition coefficients of ASD signatures of cogongrass and non cogongrass along with a plot of ROC areas for each individual coefficient.

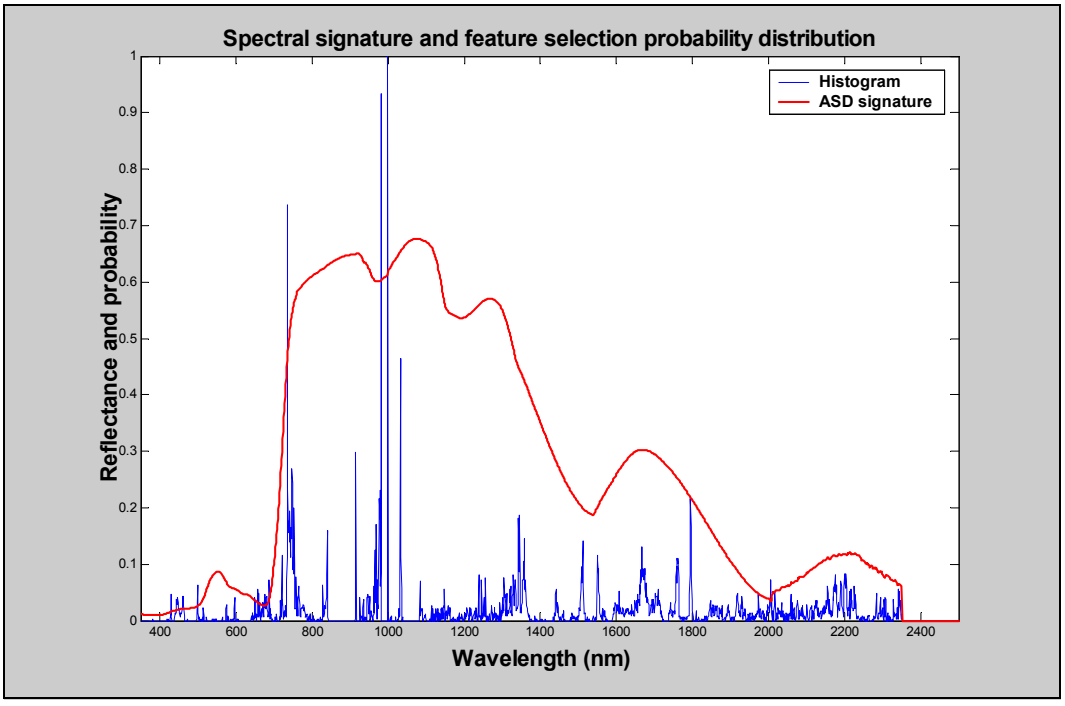

Figure 32: Example ASD signature and best band selection histogram.

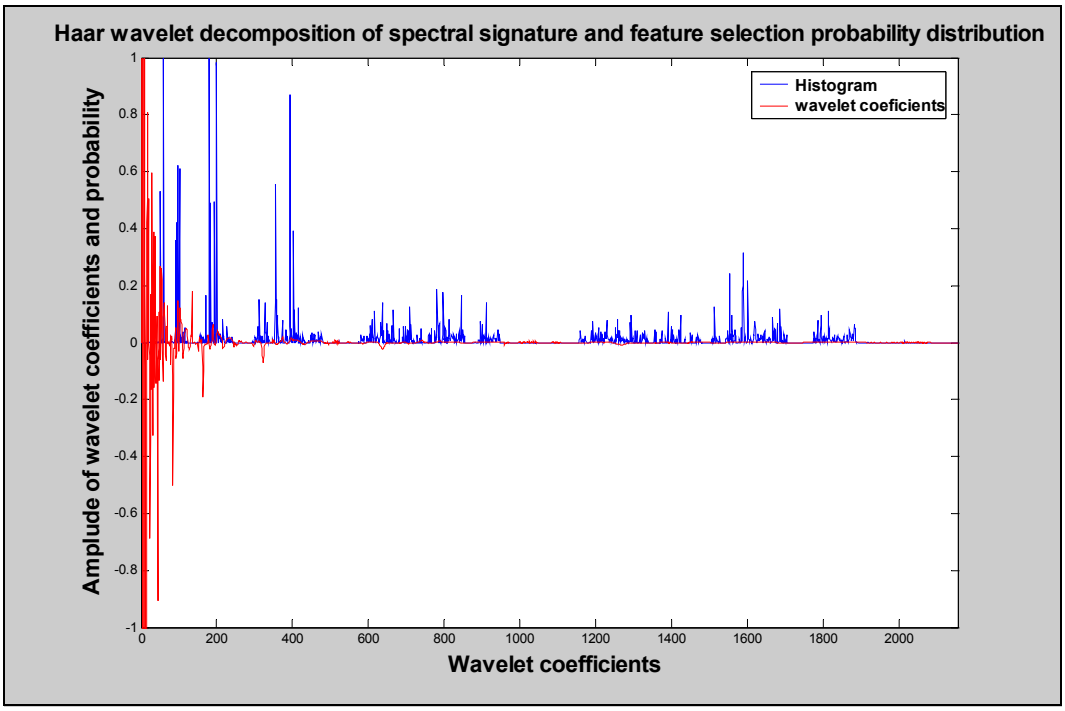

Figure 33: Example set of Haar wavelet decomposition coefficients and best band selection histogram.

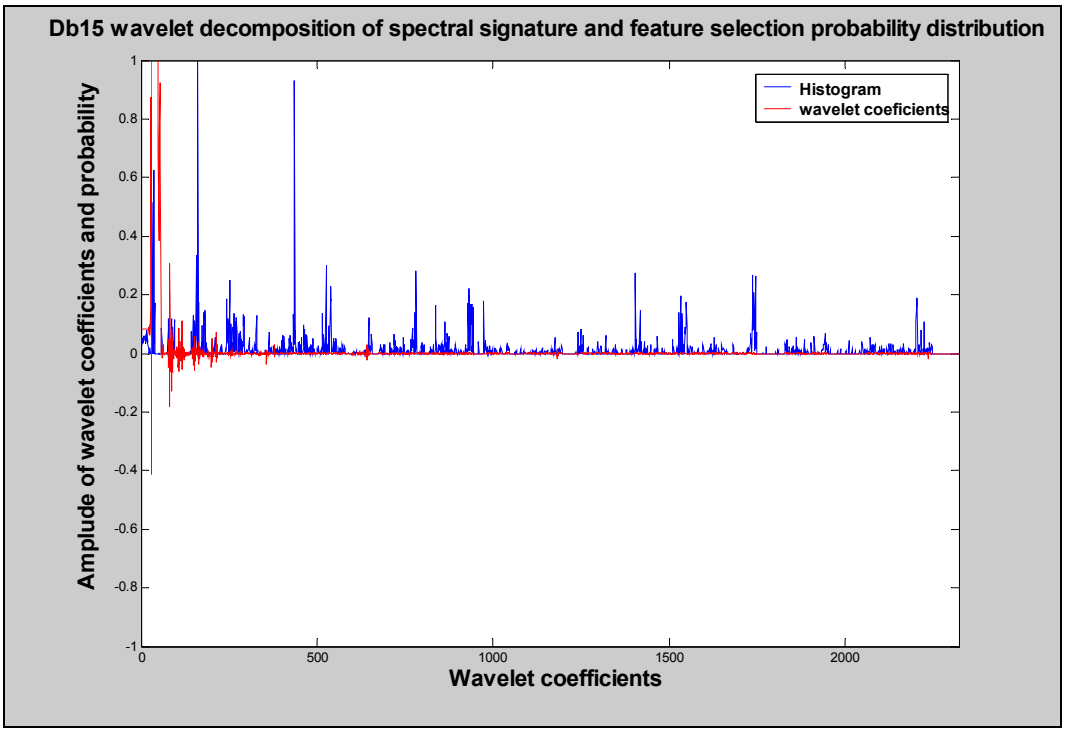

Figure 34: Example set of Daubachies15 wavelet decomposition coefficients and best band selection histogram.

4.4 On-line feature reduction and classification.

Once the best features as well as the appropriate LDA weights (for linearly combining them into a scalar) are determined, the system is tested for accuracy. To separate training data from testing data, two testing techniques were used: jack-knife and leave-one-out. Also, the classification was done by two classifiers: nearest mean and nearest neighbor. The final target detection system is illustrated in Figure 35.

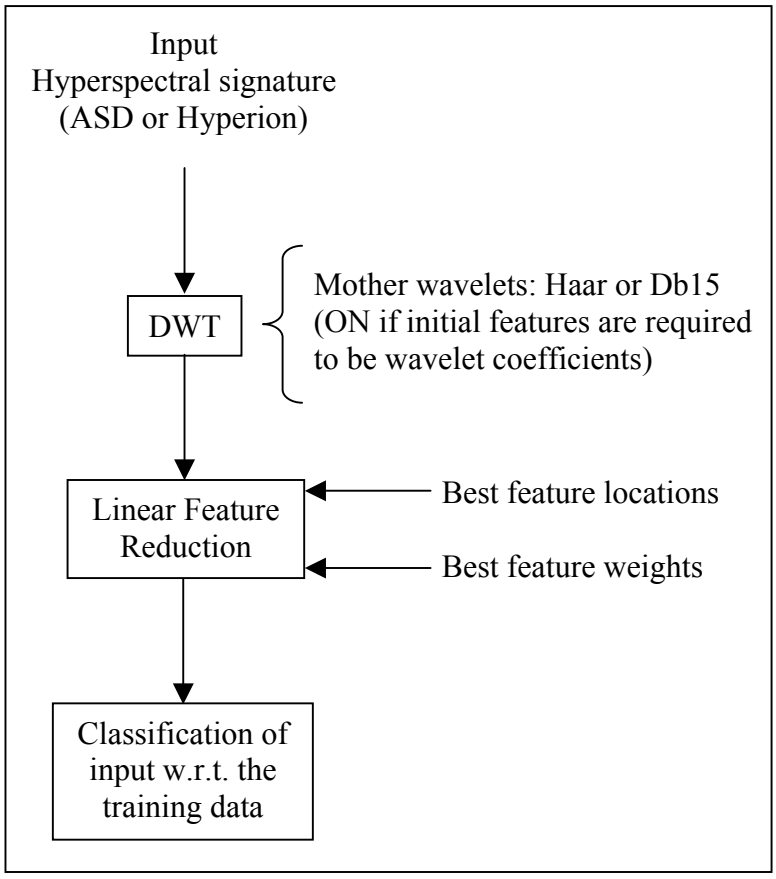

Figure 35: Offline target detection system testing.

The incoming test signature was first subjected to feature reduction by extracting the best features (spectral bands or wavelet coefficients), where the best feature locations were determined in the off-line feature selection process. Then, this set of features was weighted with the pre-determined weights to reduce it to a scalar quantity. Next, this scalar was subjected to nearest mean and nearest neighbor classifiers to determine if the input sample is cogongrass, the target material, or some other species.

CHAPTER V

RESULTS AND CONCLUSIONS

In this thesis, two types of data were used: ASD signatures and simulated Hyperion signatures. Three different types of initial features were considered to discriminate between cogongrass and non-cogongrass: pure spectral values, Haar wavelet decomposition coefficients of spectral values, and Daubachies15 wavelet decomposition coefficients of spectral values. Two types of testing techniques were used: jack-knife and leave-one-out. Two types of classifiers were used: nearest mean and nearest neighbor. For each combination of the above cases, the thresholds used for the feature selection probability distribution were 0.1 to 1.0 with increments of 0.1. The classification results for each of the above 24 cases are given in Tables 4, 5, 6, 7, 8 and 9.

In all, there were 133 hyperspectral data samples of cogongrass and 213 hyperspectral data samples of non-cogongrass that were used to compute classification accuracies. It can easily be seen from the tables that as the threshold is increased, the accuracies generally decrease. As the threshold is increased, the number of features selected also decreases. This explains the decrease in the classification accuracies as the representation of the signal is done by fewer features, and there is a loss of information. Yet this will be a better representation than selecting a number of features at random because as we raise the threshold,

the features passing through are the ones that have a higher ability to discriminate between the classes under observation. Therefore, there is a trade off in the ability of a feature set to discriminate and the number of features in the set. Also it is observed that the accuracies remain fairly constant in a particular range of thresholds. This is backed by the fact that as the threshold is raised, the features that are lost are not very important in terms of classification. Thus, by looking at these results, we can establish an optimum threshold for the application that will produce a desired level of accuracy and have a low number of features extracted to reduce the computation time. For example, Table 4 shows that as soon as the threshold is raised beyond 0.2, the number of features selected drops to as low as just 4 out of a total of 2150 bands. Also, for the cases where the nearest mean classifier is used, the overall accuracy actually increases when only these 4 bands are selected for classification, proving that the presented method not only drastically reduces the dimensionality of the problem, but also selects an optimum set of features. In the case where the nearest neighbor classifier is used, it is observed that beyond a threshold of 0.2, the overall accuracy falls by just 2 – 4% where as the dimensionality reduces from 52 to 4. Another point worth noting in Table 4 is that when the number of selected features drops from 3 to 2, the accuracy drops suddenly. This can be observed in all the four cases of Table 4, suggesting that the third last feature to be dropped has a large contribution towards the overall classification and should be kept in the feature set.

A similar observation in Table 5 shows that until the number of features drops to 2, the overall accuracy remains in a small range, after which it falls. This

shows that the features being dropped till then do not contribute enough to make a large difference to the accuracy.

Looking at Tables 6 and 9 (Daubachies-15 decomposition coefficients used as initial features), it is observed that accuracies fall much faster as compared to those in Tables 5 and 8 (Haar decomposition coefficients used as initial features). This suggests that for the similar dimensionality (4 to 6 coefficients), the Haar decomposition coefficients perform much better than Daubachies-15 decomposition coefficients.

It is also observed that for the jack-knife testing, the overall accuracy may change every time classification is done, even if the selected bands are the same, whereas in the case of leave-one-out, the accuracy is always the same when the same bands are selected. The reason is as follows: during jack-knife testing, the training and testing sets are initially randomized, which leads to different results every time. In the case of the leave-one-out testing, there is only one way to select the training and testing set, hence the same accuracies.

The results for the Hyperion sensor simulated data (Tables 7, 8 and 9) are comparable to those for the actual ASD data (Tables 4, 5 and 6). This indicates that simulated Hyperion data yields classification results as good as the ASD data. This is favorable as the initial number of bands to be handled are about one tenth the number of bands as that of ASD. This will reduce the computational complexity without affecting the overall classification accuracy. Also, Hyperion data is collected via satellite, making it much more practical. A very important point to be noted here is that the simulation of the Hyperion data has been done

under very ideal conditions and with limited specifications. The system may not perform as well on actual Hyperion data for the following reasons. First, the true specifications of the band pass filters of the Hyperion sensor are not available, they are assumed to have ideal rectangle transfer functions. Secondly, the ASD data is taken using handheld sensors. Therefore, the error introduced due to the atmosphere is minimum. Compared to ASD, the Hyperion sensor will contain much more atmospheric noise as it is a satellite based. Thirdly, because the ASD sensor is handheld, it collects reflectance spectra of true pixels as it is held close to the material under observation. Much more mixing of pixels will be introduced in the case of the Hyperion sensor, as the spatial resolution of the Hyperion is 30 meters. These three factors will hinder the overall classification accuracies of the actual Hyperion data.

The classifiers, nearest mean and nearest neighbor, have very similar classification results and are within the confidence interval. This can be explained as follows. All the species under observation are vegetation species and are very similar in features. Thus, they lie very close to each other in the feature space. Nearest mean and nearest neighbor classifications would differ if the mean of class A lies outside the class and closer to class B. In that case, the nearest neighbor classifier would give a more accurate result than the nearest mean classifier, whereas in this case, the class means lie within the class boundary and there are virtually no outliers, making both nearest mean and nearest neighbor classifiers perform equally well.

From the above discussion, it can thus be concluded that Hyperspectral signatures can be used to detect cogongrass among other subtly different vegetation species, reliably and practically.

Table 4

Classification results for amplitude of ASD data.

Hyperspectral Jack-Knife - Nearest Mean	Threshold	Number of Best Bands Selected	Cogon %	Non-Cogon %	Overall %	Confidence Interval
	0.1	52	72.73	94.34	86.05	6.82
	0.2	11	75.76	89.62	84.30	7.16
	0.3	4	86.36	92.45	90.12	5.87
	0.4	4	60.61	99.06	84.30	7.16
	0.5	3	74.24	93.40	86.05	6.82
	0.6	3	75.76	92.45	86.05	6.82
	0.7	3	66.67	92.45	82.56	7.47
	0.8	2	46.97	46.23	46.51	9.81
	0.9	2	50.00	46.23	47.67	9.83
	1.0	1	68.18	71.70	70.35	8.98
Hyperspectral Jack-Knife - Nearest Neighbor	**Threshold**	**Number of Best Bands Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	52	77.27	90.57	85.47	6.93
	0.2	11	86.36	90.57	88.95	6.17
	0.3	4	80.30	86.79	84.30	7.16
	0.4	4	81.82	85.85	84.30	7.16
	0.5	3	75.76	82.08	79.65	7.92
	0.6	3	74.24	86.79	81.98	7.56
	0.7	3	77.27	82.08	80.23	7.83
	0.8	2	45.45	62.26	55.81	9.77
	0.9	2	40.91	66.98	56.98	9.74
	1.0	1	51.52	75.47	66.28	9.30
Hyperspectral Leave-one-out - Nearest Mean	**Threshold**	**Number of Best Bands Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	52	78.95	93.43	87.86	4.53
	0.2	11	72.93	97.18	87.86	4.53
	0.3	4	79.70	95.77	89.60	4.23
	0.4	4	79.70	95.77	89.60	4.23
	0.5	3	78.20	90.14	85.55	4.88
	0.6	3	78.20	90.14	85.55	4.88
	0.7	3	78.20	90.14	85.55	4.88
	0.8	2	50.38	46.95	48.27	6.93
	0.9	2	50.38	46.95	48.27	6.93
	1.0	1	71.43	72.30	71.97	6.23
Hyperspectral Leave-one-out - Nearest Neighbor	**Threshold**	**Number of Best Bands Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	52	83.46	89.67	87.28	4.62
	0.2	11	86.47	91.55	89.60	4.23
	0.3	4	82.71	88.73	86.42	4.75
	0.4	4	82.71	88.73	86.42	4.75
	0.5	3	78.95	88.73	84.97	4.96
	0.6	3	78.95	88.73	84.97	4.96
	0.7	3	78.95	88.73	84.97	4.96
	0.8	2	46.62	63.85	57.23	6.86
	0.9	2	46.62	63.85	57.23	6.86
	1.0	1	58.65	73.24	67.63	6.49

Table 5

Classification results for Haar wavelet decomposition coefficients of ASD data.

Hyperspectral Haar Jack-Knife - Nearest Mean	Threshold	Number of Best Coefficients Selected	Cogon %	Non-Cogon %	Overall %	Confidence Interval
	0.1	40	90.91	91.51	91.28	5.55
	0.2	18	78.79	99.06	91.28	5.55
	0.3	16	74.24	97.17	88.37	6.31
	0.4	13	80.30	100.00	92.44	5.20
	0.5	9	77.27	97.17	89.53	6.02
	0.6	6	69.70	97.17	86.63	6.70
	0.7	4	62.12	45.28	51.74	9.83
	0.8	4	84.85	94.34	90.70	5.71
	0.9	3	65.15	96.23	84.30	7.16
	1.0	2	42.42	60.38	53.49	9.81
Hyperspectral Haar Jack-Knife - Nearest Neighbor	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	40	86.36	100.00	94.77	4.38
	0.2	18	95.45	91.51	93.02	5.01
	0.3	16	90.91	93.40	92.44	5.20
	0.4	13	84.85	91.51	88.95	6.17
	0.5	9	86.36	85.85	86.05	6.82
	0.6	6	80.30	96.23	90.12	5.87
	0.7	4	84.85	88.68	87.21	6.57
	0.8	4	87.88	87.74	87.79	6.44
	0.9	3	78.79	91.51	86.63	6.70
	1.0	2	39.39	69.81	58.14	9.70
Hyperspectral Haar Leave-one-out - Nearest Mean	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	40	82.71	99.06	92.77	3.59
	0.2	18	79.70	100.00	92.20	3.72
	0.3	16	80.45	99.53	92.20	3.72
	0.4	13	80.45	99.53	92.20	3.72
	0.5	9	72.93	95.31	86.71	4.71
	0.6	6	75.94	97.65	89.31	4.29
	0.7	4	75.94	91.55	85.55	4.88
	0.8	4	75.94	91.55	85.55	4.88
	0.9	3	75.94	88.26	83.53	5.15
	1.0	2	45.86	54.93	51.45	6.93
Hyperspectral Haar Leave-one-out - Nearest Neighbor	**Threshold**	**Number of Best Coefficientss Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	40	89.47	93.90	92.20	3.72
	0.2	18	87.97	91.08	89.88	4.18
	0.3	16	88.72	93.43	91.62	3.84
	0.4	13	87.22	93.90	91.33	3.90
	0.5	9	84.21	88.73	86.99	4.67
	0.6	6	81.95	90.61	87.28	4.62
	0.7	4	81.20	86.85	84.68	5.00
	0.8	4	81.20	86.85	84.68	5.00
	0.9	3	81.20	81.69	81.50	5.39
	1.0	2	36.84	59.62	50.87	6.93

Table 6

Classification results for Daubachies15 wavelet decomposition coefficients of ASD data.

Hyperspectral_db15 Jack-Knife - Nearest Mean	Threshold	Number of Best Coefficients Selected	Cogon %	Non-Cogon %	Overall %	Confidence Interval
	0.1	66	89.39	99.06	95.35	4.14
	0.2	18	83.33	96.23	91.28	5.55
	0.3	6	54.55	44.34	48.26	9.83
	0.4	4	46.97	40.57	43.02	9.74
	0.5	4	56.06	55.66	55.81	9.77
	0.6	3	46.97	46.23	46.51	9.81
	0.7	2	54.55	44.34	48.26	9.83
	0.8	2	60.61	99.06	84.30	7.16
	0.9	2	54.55	99.06	81.98	7.56
	1.0	1	53.03	100.00	81.98	7.56
Hyperspectral_db15 Jack-Knife - Nearest Neighbor	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	66	83.33	94.34	90.12	5.87
	0.2	18	72.73	83.96	79.65	7.92
	0.3	6	43.94	70.75	60.47	9.62
	0.4	4	43.94	63.21	55.81	9.77
	0.5	4	37.88	61.32	52.33	9.83
	0.6	3	71.21	88.68	81.98	7.56
	0.7	2	75.76	79.25	77.91	8.16
	0.8	2	45.45	60.38	54.65	9.79
	0.9	2	43.94	65.09	56.98	9.74
	1.0	1	65.15	85.85	77.91	8.16
Hyperspectral_db15 Leave-one-out - Nearest Mean	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	66	89.47	98.59	95.09	3.00
	0.2	18	54.14	61.03	58.38	6.84
	0.3	6	51.13	62.44	58.09	6.84
	0.4	4	57.14	58.69	58.09	6.84
	0.5	4	57.14	58.69	58.09	6.84
	0.6	3	62.41	69.95	67.05	6.52
	0.7	2	61.65	98.59	84.39	5.03
	0.8	2	61.65	98.59	84.39	5.03
	0.9	2	61.65	98.59	84.39	5.03
	1.0	1	60.15	100.00	84.68	5.00
Hyperspectral_db15 Leave-one-out - Nearest Neighbor	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	66	93.98	96.71	95.66	2.82
	0.2	18	60.15	75.12	69.36	6.39
	0.3	6	49.62	66.67	60.12	6.79
	0.4	4	54.14	65.26	60.98	6.77
	0.5	4	54.14	65.26	60.98	6.77
	0.6	3	61.65	71.36	67.63	6.49
	0.7	2	76.69	84.98	81.79	5.35
	0.8	2	76.69	84.98	81.79	5.35
	0.9	2	76.69	84.98	81.79	5.35
	1.0	1	66.92	80.75	75.43	5.97

Table 7

Classification results for amplitude of simulated Hyperion data.

Hyperion Jack-Knife - Nearest Mean	Threshold	Number of Best Bands Selected	Cogon %	Non-Cogon %	Overall %	Confidence Interval
	0.1	30	89.39	97.17	94.19	4.60
	0.2	24	75.76	94.34	87.21	6.57
	0.3	22	75.76	88.68	83.72	7.26
	0.4	20	86.36	93.40	90.70	5.71
	0.5	17	78.79	94.34	88.37	6.31
	0.6	9	78.79	84.91	82.56	7.47
	0.7	8	74.24	81.13	78.49	8.08
	0.8	7	81.82	77.36	79.07	8.00
	0.9	4	65.15	84.91	77.33	8.24
	1.0	2	68.18	75.47	72.67	8.77
Hyperion Jack-Knife - Nearest Neighbor	**Threshold**	**Number of Best Bands Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	30	89.39	93.40	91.86	5.38
	0.2	24	77.27	84.91	81.98	7.56
	0.3	22	80.30	82.08	81.40	7.66
	0.4	20	92.42	87.74	89.53	6.02
	0.5	17	78.79	82.08	80.81	7.75
	0.6	9	72.73	79.25	76.74	8.31
	0.7	8	71.21	84.91	79.65	7.92
	0.8	7	60.61	81.13	73.26	8.71
	0.9	4	68.18	73.58	71.51	8.88
	1.0	2	51.52	74.53	65.70	9.34
Hyperion Leave-one-out - Nearest Mean	**Threshold**	**Number of Best Bands Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	30	82.71	98.12	92.20	3.72
	0.2	24	75.19	96.71	88.44	4.44
	0.3	22	74.44	96.24	87.86	4.53
	0.4	20	73.68	97.18	88.15	4.48
	0.5	17	76.69	94.84	87.86	4.53
	0.6	9	75.19	83.10	80.06	5.54
	0.7	8	78.95	84.04	82.08	5.32
	0.8	7	78.20	82.63	80.92	5.45
	0.9	4	75.94	83.57	80.64	5.48
	1.0	2	69.17	72.30	71.10	6.29
Hyperion Leave-one-out - Nearest Neighbor	**Threshold**	**Number of Best Bands Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	30	88.72	92.49	91.04	3.96
	0.2	24	84.21	91.55	88.73	4.39
	0.3	22	84.96	89.67	87.86	4.53
	0.4	20	82.71	87.79	85.84	4.84
	0.5	17	81.20	85.45	83.82	5.11
	0.6	9	74.44	84.04	80.35	5.51
	0.7	8	66.17	82.16	76.01	5.92
	0.8	7	69.17	77.00	73.99	6.08
	0.9	4	63.16	78.40	72.54	6.19
	1.0	2	60.90	77.46	71.10	6.29

Table 8

Classification results for Haar wavelet decomposition coefficients of simulated Hyperion data.

Hyperion_haar Jack-Knife - Nearest Mean	Threshold	Number of Best Coefficients Selected	Cogon %	Non-Cogon %	Overall %	Confidence Interval
	0.1	47	87.88	93.40	91.28	5.55
	0.2	29	90.91	95.28	93.60	4.81
	0.3	21	71.21	96.23	86.63	6.70
	0.4	16	71.21	95.28	86.05	6.82
	0.5	13	87.88	94.34	91.86	5.38
	0.6	11	87.88	90.57	89.53	6.02
	0.7	11	78.79	92.45	87.21	6.57
	0.8	11	66.67	100.00	87.21	6.57
	0.9	10	83.33	93.40	89.53	6.02
	1.0	3	78.79	77.36	77.91	8.16
Hyperion_haar Jack-Knife - Nearest Neighbor	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	47	93.94	94.34	94.19	4.60
	0.2	29	90.91	93.40	92.44	5.20
	0.3	21	86.36	91.51	89.53	6.02
	0.4	16	84.85	88.68	87.21	6.57
	0.5	13	87.88	93.40	91.28	5.55
	0.6	11	87.88	84.91	86.05	6.82
	0.7	11	89.39	88.68	88.95	6.17
	0.8	11	89.39	87.74	88.37	6.31
	0.9	10	81.82	92.45	88.37	6.31
	1.0	3	65.15	82.08	75.58	8.45
Hyperion_haar Leave-one-out - Nearest Mean	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	47	92.48	99.53	96.82	2.43
	0.2	29	80.45	99.53	92.20	3.72
	0.3	21	81.95	96.71	91.04	3.96
	0.4	16	71.43	98.59	88.15	4.48
	0.5	13	74.44	95.31	87.28	4.62
	0.6	11	72.93	97.18	87.86	4.53
	0.7	11	72.93	97.18	87.86	4.53
	0.8	11	72.93	97.18	87.86	4.53
	0.9	10	70.68	97.18	86.99	4.67
	1.0	3	75.19	67.61	70.52	6.32
Hyperion_haar Leave-one-out - Nearest Neighbor	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	47	96.24	99.06	97.98	1.95
	0.2	29	84.96	92.49	89.60	4.23
	0.3	21	85.71	92.96	90.17	4.13
	0.4	16	84.96	89.67	87.86	4.53
	0.5	13	83.46	92.49	89.02	4.34
	0.6	11	88.72	91.55	90.46	4.07
	0.7	11	88.72	91.55	90.46	4.07
	0.8	11	88.72	91.55	90.46	4.07
	0.9	10	86.47	91.08	89.31	4.29
	1.0	3	65.41	81.22	75.14	5.99

Table 9

Classification results for Daubachies15 wavelet decomposition

Hyperion_db15 Jack-Knife - Nearest Mean	Threshold	Number of Best Coefficients Selected	Cogon %	Non-Cogon %	Overall %	Confidence Interval
	0.1	46	80.30	96.23	90.12	5.87
	0.2	12	69.70	99.06	87.79	6.44
	0.3	6	83.33	92.45	88.95	6.17
	0.4	5	54.55	50.94	52.33	9.83
	0.5	5	50.00	48.11	48.84	9.83
	0.6	5	30.30	53.77	44.77	9.78
	0.7	1	62.12	86.79	77.33	8.24
	0.8	1	60.61	87.74	77.33	8.24
	0.9	1	65.15	86.79	78.49	8.08
	1.0	1	56.06	92.45	78.49	8.08
Hyperion_db15 Jack-Knife - Nearest Neighbor	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	46	92.42	86.79	88.95	6.17
	0.2	12	90.91	93.40	92.44	5.20
	0.3	6	75.76	87.74	83.14	7.37
	0.4	5	77.27	83.96	81.40	7.66
	0.5	5	43.94	58.49	52.91	9.82
	0.6	5	36.36	67.92	55.81	9.77
	0.7	1	65.15	83.02	76.16	8.38
	0.8	1	57.58	82.08	72.67	8.77
	0.9	1	53.03	79.25	69.19	9.08
	1.0	1	60.61	84.91	75.58	8.45
Hyperion_db15 Leave-one-out - Nearest Mean	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	46	80.45	99.06	91.91	3.78
	0.2	12	74.44	97.65	88.73	4.39
	0.3	6	74.44	76.53	75.72	5.95
	0.4	5	64.66	79.34	73.70	6.11
	0.5	5	64.66	79.34	73.70	6.11
	0.6	5	64.66	79.34	73.70	6.11
	0.7	1	63.91	88.26	78.90	5.66
	0.8	1	63.91	88.26	78.90	5.66
	0.9	1	63.91	88.26	78.90	5.66
	1.0	1	63.91	88.26	78.90	5.66
Hyperion_db15 Leave-one-out - Nearest Neighbor	**Threshold**	**Number of Best Coefficients Selected**	**Cogon %**	**Non-Cogon %**	**Overall %**	**Confidence Interval**
	0.1	46	90.23	94.37	92.77	3.59
	0.2	12	85.71	90.14	88.44	4.44
	0.3	6	59.40	75.59	69.36	6.39
	0.4	5	60.90	80.75	73.12	6.15
	0.5	5	60.90	80.75	73.12	6.15
	0.6	5	60.90	80.75	73.12	6.15
	0.7	1	64.66	77.93	72.83	6.17
	0.8	1	64.66	77.93	72.83	6.17
	0.9	1	64.66	77.93	72.83	6.17
	1.0	1	64.66	77.93	72.83	6.17

CHAPTER VI

SUGGESTIONS FOR FUTURE RESEARCH

In this study, a method has been proposed and successfully been implemented to extract an optimum set of features and weight them to form a robust scalar feature for discrimination of subtly different vegetation using hyperspectral reflectance signals.

This method finds use in many applications. For example, we can obtain some labeled reflectance data and train the system offline for a particular season and/or set of weather conditions. Then, the online algorithm can be used to solve the classification problem at hand. We not only get the best features, but we also do not have to completely redesign the system for another similar case.

This method can be extended to any problem where we can obtain some labeled data that adequately represents the incoming test signals. This procedure can also be used to locate the regions of interest in the entire initial features' space. For example, if this is used on hyperspectral signatures to distinguish between two types of vegetation, the final distribution of the best features will provide regions in the spectrum for best discrimination. Further analysis should be done at these locations in order to link the effect of these wavelengths to the different physiological properties of the two plant species under observation. This

study can thus be used to better define vegetation indices, or similar features, for specific applications.

The algorithm can also find use in the design of application-specific multispectral sensors. This method can provide the “best bands” in the entire spectrum for a particular classification problem. Then, the multispectral sensor can be better designed such that the reflectances at these wavelengths are recorded with the least amount of error. Once such a multispectral sensor is designed, the data it provides will produce results as good as if the entire hyperspectral signature were used and passed through the optimal best bands selection algorithm. Likewise, an adaptive hyperspectral sensor could be designed and used to collect and/or transmit bands pertinent to a particular problem.

One of the aspects that still need to be dealt with is the selection of the threshold for the best bands distribution histogram. In this thesis, the threshold has been selected manually from 0.1 to 1.0 with increments of 0.1 to quantify the performance and to experimentally find the optimal threshold. The results show that there is a tradeoff between the ability of a feature to discriminate between two classes and the number of such features. Thus, there is a need to investigate automated methods that can be applied for calculation of the threshold for a particular case and for a desired performance and computation time.

REFERENCES

[1] Lisa Gaddis, Laurence Soderblom, Hugh Kieffer, Kris Becker, Jim Torson, and Kevin Mullins, "Spectral Decomposition of AVIRIS Data," *United States Geological Survey*, 2255 N. Gemini Drive, Flagstaff, AZ 86001

[2] John D. Byrd, Jr. and Charles T. Bryson, "Biology, Ecology, and Control of Cogongrass," *Mississippi Department of Agriculture and Commerce*.

[3] Clark, R.N., King, T.V.V., Ager, C., and Swayze, G.A., H. H. Posey, J.A. Pendelton, and D. Van Zyl Eds., "Initial vegetation species and senescence/stress mapping in the San Luis Calley, Colorado using imaging spectrometer data," *Proceedings: Summitville Forum '95, Colorado Geological Survey Special Publication* 38, p. 64-69, 1995.

[4] Nicholas M. Short "Principles of Spectroscopy" *RST Section 13* http://rst.gsfc.nasa.gov/Sect13/Sect13_6.html

[5] R.N. Clark, and T.L. Roush, "Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications," *J. Geophys.* Res., 89, p. 6329-6340, 1984.

[6] Price, K.P. X. Guo and J.M. Stiles, "Optimal Landsat TM Band Combinations and Vegetation Indices for Discrimination of Six Grassland Types in Eastern Kansas," *PECORA 14 Land Satellite Information III*, Denver, Colorado. Dec. 6-10. P347-357, 1999.

[7] Ralph Garano, Greg Gaston and Ofer Heyman, "Remote Sensing Systems for Identifying and Mapping Aquatic Vegetation in Estuaries and Other Near-shore Environments Along the Pacific Coast," *NASA Commercial Remote Sensing Program Office, John C. Stennis Space Center*, MS 39529, 1999.

[8] Stephen D. Stearns, Bruce E. Wilson, and James R. Peterson, "Optimal Band Selection for Dimensionality Reduction of Hyperspectral Imagery," *ESL Incorporated.*

[9] L.M. Bruce, C. Morgan, S. Larsen, “Automated detection of subpixel targets with continuous and discrete wavelet transforms,” *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2217-2226, 2001

[10] J. Li, L.M. Bruce, J. Byrd, J. Barnett, “Automated detection of Pueraria montana (kudzu) through Haar analysis of hyperspectral reflectance data,” *Proc. IEEE Geoscience and Remote Sensing Symposium*, pp. 2247–2249, Sydney, Australia, July 2001.

[11] Yan Huang; Bruce, L.M.; Koger, T.; Shaw, D., "Analysis of the effects of cover crop residue on hyperspectral reflectance discrimination of soybean and weeds via Haar transform," *Geoscience and Remote Sensing Symposium, 2001. IGARSS '01. IEEE 2001 International*, vol. 3, pp. 1276-1278, 2001.

[12] C.S. Burrus, R.A. Gopinath, H. Guo, “Introduction to Wavelets and Wavelet Transforms: A Primer”, *Prentice-Hall, Inc.*, New Jersey, 1998.

[13] R. O. Duda, P. E. Hart and David G. Stork “Pattern Classification (2nd ed)” Wiley, New York, NY, 2000.

[14] D.C. Hatchell, Managing Editor, “Analytical Spectral Devices, Inc. (ASD) Technical Guide, third edition”, 1999.

[15] Steve Carman, "Hyperion Grating Imaging Spectrometer," *TRW Space & Electronics* http://eo1.gsfc.nasa.gov/overview/Workshop/06.pdf