1-1-2009

Incorporating Multiresolution Analysis With Multiclassifiers And Decision Fusion For Hyperspectral Remote Sensing

Terrance Roshad West

INCORPORATING MULTIRESOLUTION ANALYSIS WITH MULTICLASSIFIERS AND DECISION FUSION FOR HYPERSPECTRAL REMOTE SENSING

By

Terrance R. West

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Electrical Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2009

INCORPORATING MULTIRESOLUTION ANALYSIS WITH MULTICLASSIFIERS AND DECISION FUSION FOR HYPERSPECTRAL REMOTE SENSING

By

Terrance R. West

Approved:

Lori M. Bruce Associate Dean of Research and Professor of Electrical and Computer Engineering (Director of Dissertation)	Nicolas H. Younan Department Head of Electrical and Computer Engineering and James Worth Bagley Chair (Committee Member)
Jenny Q. Du Associate Professor of Electrical and Computer Engineering (Committee Member)	Daniel Reynolds Professor of Plant and Soil Sciences (Committee Member)
James E. Fowler Professor of Electrical and Computer Engineering Graduate Program Director, Electrical and Computer Department	Sarah A. Rajala Dean of the Bagley College of Engineering (Dean of the College of Engineering)

Name: Terrance R. West

Date of Degree: December 11, 2009

Institution: Mississippi State University

Major Field: Electrical Engineering

Major Professor: Dr. Lori Bruce

Title of Study: INCORPORATING MULTIRESOLUTION ANALYSIS WITH MULTICLASSIFIERS AND DECISION FUSION FOR HYPERSPECTRAL REMOTE SENSING

Pages in Study: 123

Candidate for Degree of Doctor of Philosophy

The ongoing development and increased affordability of hyperspectral sensors are increasing their utilization in a variety of applications, such as agricultural monitoring and decision making. Hyperspectral Automated Target Recognition (ATR) systems typically rely heavily on dimensionality reduction methods, and particularly intelligent reduction methods referred to as feature extraction techniques. This dissertation reports on the development, implementation, and testing of new hyperspectral analysis techniques for ATR systems, including their use in agricultural applications where ground truthed observations available for training the ATR system are typically very limited.

This dissertation reports the design of effective methods for grouping and down-selecting Discrete Wavelet Transform (DWT) coefficients and the design of automated Wavelet Packet Decomposition (WPD) filter tree pruning methods for use within the framework of a Multiclassifiers and Decision Fusion (MCDF) ATR system. The efficacy of the DWT MCDF and WPD MCDF systems are compared to existing ATR methods

commonly used in hyperspectral remote sensing applications. The newly developed methods' sensitivity to operating conditions, such as mother wavelet selection, decomposition level, and quantity and quality of available training data are also investigated.

The newly developed ATR systems are applied to the problem of hyperspectral remote sensing of agricultural food crop contaminations either by airborne chemical application, specifically Glufosinate herbicide at varying concentrations applied to corn crops, or by biological infestation, specifically soybean rust disease in soybean crops. The DWT MCDF and WPD MCDF methods significantly outperform conventional hyperspectral ATR methods. For example, when detecting and classifying varying levels of soybean rust infestation, stepwise linear discriminant analysis, results in accuracies of approximately 30%-40%, but WPD MCDF methods result in accuracies of approximately 70%-80%.

DEDICATION

I dedicate this dissertation to my mother, who always believed and always had faith in me when others doubted. I also dedicate this dissertation to minorities in the fields of science, technology, engineering and mathematics, who wish to obtain a Masters degree or Doctoral degree in their fields. To them I say, stay the course because success it is not measured by how fast one finishes but is measured by if one continues to completion. "Look Onward and Upward Toward the Light"

ACKNOWLEDGEMENTS

I would like to thank Dr. Lori Mann Bruce for being an excellent dissertation advisor and committed and providing encouragement and outstanding guidance over the course of my Master and Doctoral program. I thank my dissertation committee members, Dr. Jenny Du, Dr. Nicolas Younan, Dr. Daniel Reynolds, for providing guidance. I thank Dr. Lori Bruce and Dr. David Shaw (Director of the GeoSystems Research Institute at Mississippi State University) for ensuring continual financial and technological support during my doctoral program. I thank all my colleagues which aided in the success of this dissection. I would like to thank Mr. Louis Wasson for assisting me with the experimental hyperspectral data collection.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

With the ongoing development and increased affordability of a diverse array of sensors, many of today's sensing systems produce huge amounts of raw data. Automated pattern recognition systems typically rely heavily on dimensionality reduction methods, and particularly intelligent reduction methods referred to as feature extraction techniques. Feature extraction, in general, is a procedure that reduces the dimensionality of a data set while selecting or constructing features that describe the observation in a meaningful way. Typically, the term "meaningful" relates to an ability to detect a given target or discriminate between particular classes of observations. From a statistical perspective, the goal of feature extraction often is to select features leading to large between-class variances and small within-class variances within the feature space [1]. Feature extraction in pattern recognition systems is an essential element in numerous applications, including speech recognition [2], remotely sensed target recognition [3], and computer aided diagnosis (CAD) medical systems [4].

In the field of remote sensing hyperspectral sensors have the ability to produce 100's to 1000's of contiguous spectral bands that normally range from the visible to the thermal infrared (IR) portions of the electromagnetic spectrum. Typically, each band conveys the percentage of incident light that is reflected over a specified narrow range of

the electromagnetic spectrum. Hyperspectral sensors have become an attractive method of collecting data for Automated Target Recognition (ATR) systems due to its ability to produce large quantities of information (hundreds to thousands of spectral bands per pixel) that represent near-continuous measurements of spectral reflectance. This is akin to conducting chemical spectroscopy remotely, albeit challenging with the many noise sources that affect the measurement in practical applications.

In many hyperspectral classification applications, individual spectral bands are extracted as features for the identification of a target. When using statistical pattern recognition techniques, the large dimensionality of the feature space induces a requirement of a large amount of labeled training data, if the class distributions are to be accurately described. In practical scenarios, hyperspectral sensors usually results in a high dimensionality data sets with small numbers of labeled training data. The increase in spectral features along with small amount of labeled training data naturally causes hyperspectral ATR systems to suffer the "curse of dimensionality", resulting in lower classification accuracies [5]. This phenomenon reveals that the amount of training data is not sufficient to support the number of features produced by the sensor.

In the remote sensing community, the curse of dimensionality is often discussed in terms of the Hughe's phenomenon [6]. For a finite amount of training data, as the number of features increases the target detection accuracy increases. After a critical point, however, the target detection accuracy begins to decrease as the number of features increases. To account for the lack of labeled training data, i.e. ground truthed pixels, hyperspectral ATR systems typically reduce the high dimensional data via dimensionality reduction or feature extraction techniques such as Principal Component Analysis (PCA),

Linear Discriminant Analysis (LDA), Discriminant Analysis Feature Extraction (DAFE), and spectral band grouping before data is classified [1,5,7]. These commonly used techniques in the remote sensing community aim to reduce the dimensionality of the high dimensionality data while simultaneously retaining pertinent information that can be used to differentiate between ground cover classes. These methods usually employ a single classifier. Despite their popularity, problems arise when applying these methods to very high dimensional data. These problems arise during the learning or training stages of the statistical dimensionality reduction techniques, due to the use of higher order statistics such as covariance matrices. For example, LDA and DAFE require the computation of the inverse of the within-class covariance matrix. If there is not a sufficient amount of training data available, the covariance matrix will be sparse, and its inverse may not be computable. A variety of techniques exist to try to circumvent this issue, such as pseudo-inverses and whitening or regularization of covariance matrices [8, 9]. However, these approaches are not optimal solutions, as they merely reduce, rather than eliminate, the risk of errors in the ATR methods.

1.2 Motivation for Proposed Work

Many dimensionality reduction and feature extraction methods have been investigated for hyperspectral data [6-12]. In particular, spectral band grouping, combined with multiclassifiers and decision fusion (MCDF), has been shown recently to be a very promising solution [6, 9, 13, 14]. With this approach, the adjacent spectral bands are intelligently grouped in order to form lower dimensional subspaces. Then the spectral band groups are sent to a bank of classifiers, one classifier for each group. Next, the outputs of the classifiers are fused using decision fusion to produce one final

classification, e.g. target or non-target. The weights used in the decision fusion stage of the system typically take into account the reliability of each group/classifier combination to accurately classify a pixel. However, since the approach is based on localized spectral band grouping, it lacks the ability to extract large scale or global features from the hyperspectral data. The features which are extracted from the hyperspectral data only take in account the phenomenons which are presented in specific localized regions in the electromagnetic spectrum. Multiresolution wavelet analysis has the ability to extract local and global features which could be meaningful in many target recognition applications. In multiresolution wavelet analysis, local and global features are extracted by decomposing the high dimensional data by projecting it onto a scaled and translated version of a prototype function. This projection produces approximation and detail coefficients, which contain the local and global features of the hyperspectral data. Combining multiresolution wavelet analysis with the MCDF approach has the potential to provide significantly higher target detection and classification accuracies for hyperspectral systems as compared to the current state of the art ATR approaches, particularly when the amount of available training data is very limited as is the case in many practical applications.

1.3 Contributions of this work

This research seeks to design an ATR system that is capable of performing classification tasks on high dimensional data, such as remotely sensed hyperspectral data, when only a relatively small amount of training data is available. The recently developed MCDF approach is extended for use in a multiresolutional domain, such as a wavelet transform domain. The new approach is expected to outperform the conventional MCDF

approach, particularly in hyperspectral remote sensing ATR systems, due to its ability to exploit both local and global characteristics of the ATR systems input observations.

The primary contributions of this dissertation are listed below.

A. Design of multiresolutional MCDF ATR system

1. Design effective methods for grouping and down-selecting DWT coefficients for use within the framework of a MCDF ATR system.

2. Compare the efficacy of the newly developed DWT MCDF methods to existing ATR methods commonly used on hyperspectral remotely sensed data.

3. Determine the sensitivity of the DWT MCDF system to the selection of mother wavelet, DWT decomposition level, and coefficient grouping/down-selection methods.

4. Design effective methods for grouping and down-selecting redundant WPD coefficients for use within the framework of a MCDF ATR system, including supervised and unsupervised methods for pruning WPD filter trees resulting in options for either redundant or orthogonal decompositions.

5. Compare the efficacy of the newly developed WPD MCDF methods to existing ATR methods commonly used on hyperspectral remotely sensed data.

6. Determine the sensitivity of the WPD MCDF system to the selection of mother wavelet, WPD decomposition level, and WPD decomposition tree pruning method.

B. Application of multiresolutional MCDF ATR system to problem of remote sensing of agricultural food crop contaminations

7. Collect handheld spectroradiometer data and airborne hyperspectral imagery of food crop contaminations, either by airborne chemical application, specifically Glufosinate herbicide at varying concentrations applied to corn crops, or by biological infestation, specifically soybean rust disease in soybean crops.

8. Apply DWT MCDF and WPD MCDF ATR systems to said hyperspectral datasets and compare the newly developed methods to existing ATR systems' efficacies for detecting and classifying the varying levels of contamination.

9. Determine the WPD MCDF ATR system's sensitivity to (i) time delay between herbicide application and collection of remotely sensed data, (ii) abundance of ground truthed observations available for training the ATR system, and (iii) misalignment of training and testing data, i.e. scenarios where ground truthed (class labeled) training observations are collected at a vegetative growth stage that is different than the actual test imagery.

1.4 REFERENCES

[1] S. Theodoridis, K. Koutroumbas, *Pattern Classification*, Academic Press, 1999.

[2] F. Jelinek, *Statistical Methods for Speech Recognition*, MIT Press, 1998.

[3] J. P. Muller, Digital Image Processing in Remote Sensing, Taylor & Francis, 1998.

[4] M. Sameti, R. K. Ward, B. Palcic, J. M. Parkes, "Texture Feature Extraction for Tumor Detection in Mammographic Images," *IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing*, vol. 1, pp. 831 – 834, 1997.

[5] R. O. Duda, P. E. Hart, D. G. Strork, Pattern Classification, John Wiley & Sons, Inc., 2001.

[6] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers", IEEE Trans. Inform. Theory, vol. 14, pp. 55-63, 1968.

[7] A. Cheryadat, L. M. Bruce, A. Mathur "Decision Level Fusion with Best Bases for Hyperspectral Classification", *Proc. IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp. 399-406, October 2003.

[8] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, "Regularization studies on LDA for face recognition," *Proc. 2004 International Conference on Image Processing*, vol. 4, pp 63-66, October 2004.

[9] S. Prasad, L.M. Bruce, "Overcoming the Small Sample Size Problem in Hyperspectral Classification and Detection Tasks," *Proc. IEEE International Geoscience and Remote Sensing Symposium*,vol. 5, pp. 381-384, July 2007.

[10] S. Venkataraman, L. M. Bruce, "Hyperspectral Dimensionality Reduction via Localized Discriminant Bases," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Seoul, Korea, vol. 2, pp. 1245 -1248, July 25-29, 2005.

[11] T. West, L. M. Bruce, "Multiclassifiers and Decision Fusion in the Wavelet Domain for Exploitation of Hyperspectral Data", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, vol. 1, pp. 4850-4853, July 2007.

[12] 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. 29, no. 10, pp. 2217-2226, 2001.

[13] S. Prasad. L. M. Bruce, "Hyperspectral Feature Space Partitioning via Mutual Information for Data Fusion," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, vol. 1, pp. 4846-4849, 2007.

[14] S. Prasad, L. M. Bruce, "Decision Fusion with Confidence-Based Weight Assignment for Hyperspectral Target Recognition," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1448 – 1456, May 2008.

CHAPTER 2

CURRENT STATE OF KNOWLEDGE

2.1 Hyperspectral Imaging and Analysis

"Hyperspectral sensors (sometimes referred to as imaging spectrometers) are instruments that acquire images in many, very narrow, contiguous spectral bands throughout the visible, near-infrared (IR), mid-IR, and thermal IR portions of the spectrum" [1]. Hyperspectral sensors have the ability to produce several hundred to thousands of spectral bands per pixel. Figure 2.1 describes the method in which hyperspectral dataset is obtained. The charged couple device (CCD) array collects the reflected energy of light from the ground scene across the electromagnetic spectrum. Typically, the dataset collected is reported either as digital numbers (DN) or is atmospherically corrected and converted to reflectance. This collection of reflectance forms a hyperspectral cube. Hyperspectral signatures can be extracted per pixel from the hyperspectral cube. The cube is defined by pixels in which their positions can be determined by row and column. The *(i,j)* pixel with M bands forms what is referred to as a hyperspectral signature. From the signatures, pure endmember pixels can be identified or the abundance of multiple endmembers such as vegetation, soil, and water, which is shown in Figure 2.2.

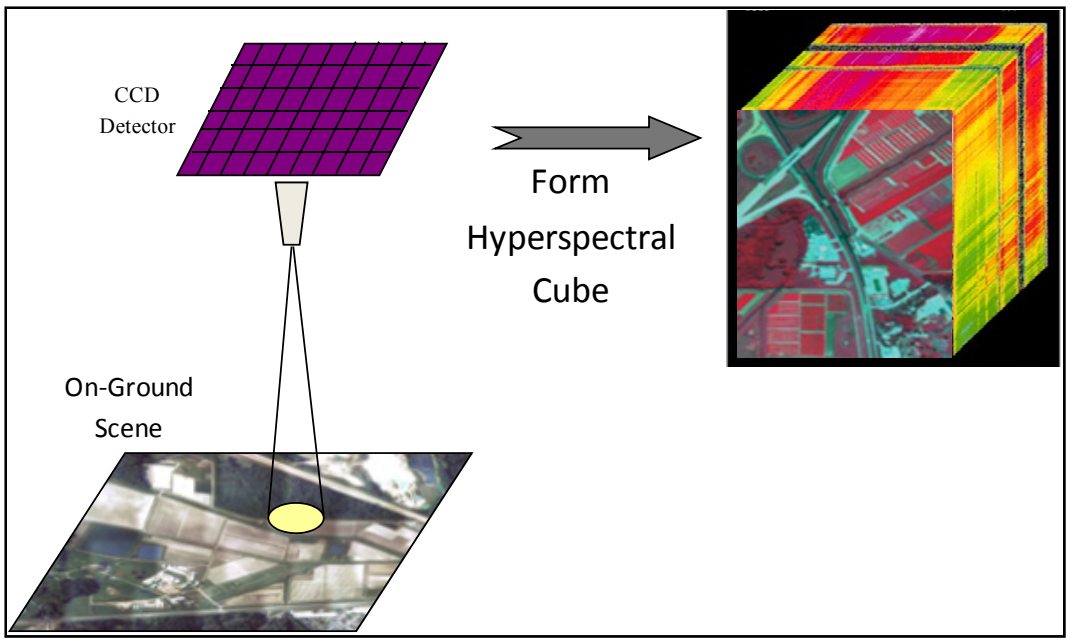


Figure 2.1 Representation of the collection of hyperspectral data

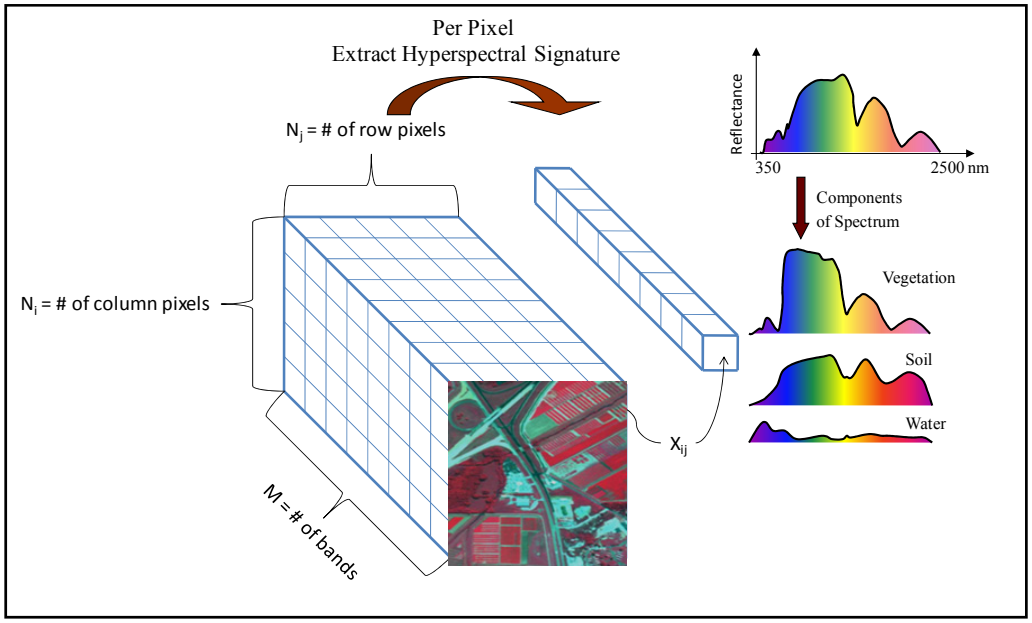


Figure 2.2 Representation of the extraction of hyperspectral signatures from hyperspectral cube

2.2 Current Hyperspectral Feature Extraction and Dimensionality Reduction

The space in which the hyperspectral data resides is mostly empty due to the fact that much of the data is redundant. The high spectral correlation between adjacent bands produces this redundancy. This redundancy allows the hyperspectral data to be projected on to a lower dimensional subspace, while simultaneously retaining pertinent information for classification and target recognition tasks.

Principal Component Analysis (PCA), Fisher's Linear Discriminant Analysis (LDA), Discriminant Analysis Feature Extraction (DAFE), spectral band grouping, multiresolution wavelet analysis, and multiclassifiers and decision fusion (MCDF) are some of the current methods utilized for hyperspectral dimensionality reduction and feature extraction in pattern classification applications in the remote sensing community [2-22].

2.2.1 Principal Component Analysis (PCA)

PCA is a commonly used method for dimensionality reduction in hyperspectral data analysis. PCA can be found in many commercial software packages for remote sensing such as ENVI^{TM} and IMAGINE^{TM}. PCA seeks to find a linear transformation which projects the data onto a subspace in which the features are mutually uncorrelated and the total variance of the data is maximized. The linear transformation involves applying eigen-analysis to the covariance matrix of the entire unlabeled data set [2, 3]. Thus PCA is an unsupervised method. For example, suppose there is an $\boldsymbol{h}$ – dimensional data set, and we compute the mean $\vec{\boldsymbol{\mu}}$ and the corresponding $\boldsymbol{h} \times \boldsymbol{h}$ covariance matrix $\mathbf{K}$ for the data set. The transformation is derived by obtaining the eigen-values and eigen-vectors from the total covariance matrix $\mathbf{K}$ [2, 3]. Next, the eigen-values and eigen-

vectors are sorted in descending order according to the eigen-value. Finally, the ***d*** eigen-vectors having the highest eigen-values are selected as the basis of the linear transformation. The number of eigen-vectors selected determines the dimensionality of the projected data set (i.e. the eigen-vectors are the rows of the transformation matrix). Thus, the dimensionality is reduced from ***h*** to ***d***. Although PCA is a commonly used method for dimensionality reduction in remote sensing, it has been shown that it is not an optimal feature extraction method [4]. It was shown by Cheriyadat and Bruce [5] and Prasad and Bruce [4] that PCA is not a sufficient method for dimensionality reduction (feature extraction) for classification and target recognition applications. This is primarily due to the fact that the method is based on the total covariance matrix, rather than class-specific covariance matrices. That is, class labels are not used, and the method is trained on unlabeled data.

2.2.2 Linear Discriminant Analysis

Fisher's LDA seeks to maximize the class separation between data by reducing the dimensionality through the projection of data onto a lower subspace. This separation is achieved my maximizing the between-class covariance matrix and minimizing the within-class covariance matrix [2]. Thus, LDA is a supervised method. The objective of LDA is to find a linear transformation matrix W such that $\vec{y} = W^T\vec{x}$, where $\vec{x} \in \mathcal{R}^d$ (original data), $\vec{y} \in \mathcal{R}^m$ (projected data), $m = c - 1$, (c is the number of classes), such that the between-class covariance matrix is maximized and the within-class covariance is minimized. This transformation matrix W^T can be obtained by maximizing the following criterion function,

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|} \quad (2.1)$$

which can be mathematically solved via a generalized eigen-value problem. This problem can be solved by

$$S_W^{-1} S_B = \lambda W \quad (2.2)$$

where λ is the eigen value, S_B is the between-class covariance matrix, S_W is the within-class covariance matrix which are derived by

$$S_B = \sum_{i=1}^{c} n_i (\overrightarrow{m_i} - \vec{m})(\overrightarrow{m_i} - \vec{m})^T \quad (2.3)$$

$$S_W = \sum_{i=1}^{c} \sum_{\vec{x} \in C_i} (\vec{x} - \vec{m})(\vec{x} - \vec{m})^T \quad (2.4)$$

$\overrightarrow{m_i}$ and $\vec{m}$ are the mean of the i^{th} class and the global mean, respectively [3]. Note that in calculating the transformation matrix W^T, the inverse of the within-class covariance matrix must be calculated. A problem arises in this calculation when there are too many features with too few training vectors which cause the S_W matrix to become sparse. The sparseness causes S_W to become ill-conditioned and can inhibit the calculation of its inverse. Thus, when the feature vector's dimensionality is relatively large compared to the number of training observations (which is typical with hyperspectral remote sensing) LDA can be intractable. In order to resolve this issue, researchers have investigated three approaches: (i) stepwise LDA [6,7], (ii) regularized LDA [8,9], and subspace LDA [4].

Stepwise LDA is an iterative implementation of LDA. The inputs to LDA, typically features, are sorted in descending order of class separation efficacy, using a performance metric, like class separation, e.g. Bhattacharyya Distance (BD). Next, a forward selection process is conducted to form (grow) a subset of features. This portion of the method is a bottom-up approach, where the top performing feature seeds the

subset. Features are added to the subset only if the BD of their LDA result is increasing. Next, a backward rejection process is conducted to form (shrink) a subset of features. This portion of the method is top-down approach, where the final subset of the forward selection seeds the subset. Features are removed from the subset only if the BD of the LDA of the reduced set is increasing. After the removal of all features in the subset has been considered, the result is finalized. LDA is applied to the final subset. Stepwise LDA is often referred to in the remote sensing community as DAFE, and is commonly employed in hyperspectral applications [6, 7].

Regularized LDA is a simple technique designed to stabilize LDA. A small amount of noise is added to the diagonal of the within-class scatter matrix, thus, ensuring the existence of its inverse [8]. In 2008, Prasad and Bruce applied regularlized LDA to hyperspectral feature extraction and reduction [9]. They found the method to work on par with PCA. Regularlized LDA is also referred to as "whitened LDA" in the remote sensing community.

Subspace LDA is a method that employs PCA, as a dimensionality reduction technique, prior to LDA. That is, subspace LDA is a two step linear transformation, where the first linear transformation is a PCA projection, which discards the null space of the overall scatter matrix (thereby, making the within-class scatter matrix full ranked.) The second linear transformation is a LDA projection from the PCA projected space[10]. In 2007, Prasad and Bruce applied subspace LDA to hyperspectral feature extraction and reduction [4]. They found the method to work on par with PCA.

2.2.3 Projections Pursuits

The principal objective of projection pursuits is to overcome the "curse of dimensionality" while at the same time retaining information within the hyperspectral signal that is pertinent to target detection and classification. The idea of performing orthogonal projections on hyperspectral data, such as projection pursuits, is not a new concept. However, it is not nearly as commonly used as methods like PCA and LDA. The method of projection pursuits has been applied to a few types of hyperspectral applications.

Lin and Bruce evaluated the use of projection pursuits for dimensionality reduction using hyperspectral data for applications involving agricultural target recognition [11]. The dataset was obtained by a handheld spectroradiometer which collected 2000 spectral bands of two types of vegetative species. The targets in their experiment were sicklepod and cocklebur. In their study, parallel parametric projection pursuits, projection pursuits best band selection, and sequential parametric projection pursuits (SPPP) methods were employed. The two projection indices used in their research were BD and the area under the receiver operating characteristic (ROC) curves [11]. The weights for the transformation matrix consisted of a vector that averaged the bands in a group, a vector that chose only one spectral band, and a vector that maximized the performance metric. The projection pursuits preprocessing methods employed in their study proved to have higher classification accuracies than data that were not preprocessed with the projection pursuits.

Another study in which projection pursuits was employed was performed by Ifarraguerri and Chang [12]. In their study, the hyperspectral imagery was collected by

the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor. The authors evaluated the use of projections pursuits in the analysis of hyperspectral data in an unsupervised method. The projection pursuit method was performed by applying PCA to the area of interest and the element which had the largest eigen-values were obtained and then placed in the transformation index [12]. The projection pursuit methods in this study proved that with the information divergence index the dimensionality of the hyperspectral image could be reduced while retaining the important characteristics of the image.

In 2006, West investigated the use of SPPP for the purpose of hyperspectral dimensionality reduction and applied the method to the problem of invasive species remote sensing [13]. The SPPP method was implemented in a top-down fashion, where hyperspectral bands were used to form an increasing number of smaller groups, with each group being projected onto a subspace of dimensionality one. Both supervised and unsupervised potential projections were investigated for their use in the SPPP method. Fisher's LDA was used as a potential supervised projection. Average, Gaussian-weighted average, and PCA were used as potential unsupervised projections. The BD was used as the SPPP performance index. The performance of the SPPP method was compared to two other currently used dimensionality reduction techniques, namely best spectral band selection (BSBS) and best wavelet coefficient selection (BWCS). The SPPP dimensionality reduction method was combined with a nearest mean classifier to form an ATR system. The ATR system was tested on two invasive species hyperspectral datasets: a terrestrial case study of cogongrass (*Imperata cylindrical*) versus johnsongrass (*surghm halopense*) and an aquatic case study of waterhyacinth versus American Lotus. For both

case studies, the SPPP approach either outperformed or performed on par with the BSBS and BWCS methods in terms of classification accuracy; however, the SPPP approach required significantly less computational time. For the cogongrass and waterhyacinth applications, the SPPP method resulted in overall classification accuracy in the mid to upper 90's.

2.2.4 Spectral Band Grouping

In spectral band grouping, adjacent groups of spectral bands are merged to reduce the data set's dimensionality. The spectral band grouping is achieved by either applying a fixed-size sliding window or by employing a bottom-up approach to the grouping of adjacent spectral bands. In the case of a fixed-sized window grouping, non-overlapping equally sized groups of bands are formed, (example case illustrated in Figure 2.3). In the bottom-up approach (example case illustrated in Figure 2.4), the system views each spectral band as a group in the initial stage. Adjacent groups are merged to form larger groups (i.e. the groups are then grown across the spectrum) until some predefined stopping criteria is met. The groups are merged as long as a pre-defined classification performance metric is increasing, such as class separation and/or classification accuracy. The merging of the groups is stopped when the metric is no longer adequately increasing or the group size becomes larger than what the training data can support. Once the groups are formed, the groups are reduced by projecting them onto a lower dimensional subspace. Typically, a linear transformation is used, such as the mean, LDA, or PCA, such that each group results in a small number of features.

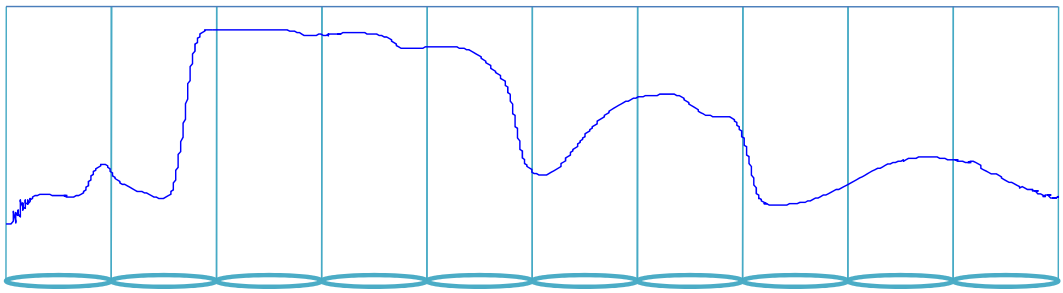


Figure 2.3 Example case of hyperspectral band grouping via a fixed-size sliding window

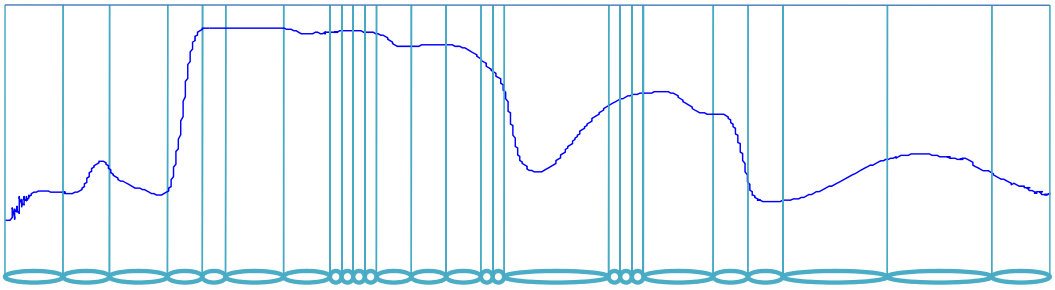


Figure 2.4 Example case of hyperspectral band grouping via bottom-up approach employing intelligent band grouping, resulting in unequally sized groups

Ball *et al.* utilized spectral band grouping for hyperspectral segmentation. They presented a supervised segmentation technique which involved the use of the level set segmentation, spectral information divergence, and best band analysis [14]. Best band analysis is performed by calculating the BD using the spectral information divergence for each class for different band sets. The highest BD features are then used to form a feature set which is used in the initial classification. The initial segmentation is performed by using the Euclidean distance classifier using the selected feature set The level set method is then applied to the initial segmentation employing a two-dimensional stopping map by treating the feature set for each pixel as a random variable and examining the feature set's cumulative distribution function [14]. This procedure was

applied to Compact Airborne Spectrographic Imager (CASI) hyperspectral image in an agricultural application.

Backer *et al.* developed a new band reduction technique which employs local continuous function weighting for grouping spectral bands [15]. In this technique, optical filters with two degrees of freedom are used to project the signal onto a lower dimensional subspace which allows for continuous optimization strategies for band selection approaches [15]. The two degrees of freedom are defined by the central wavelength and the width of the filter. The band settings are optimized, using the Bhattacharyya bound. This technique was applied to a CASI-2 image.

Du *et al.* proposed unsupervised band analysis techniques which use similarity measurements in order to group spectral bands. Multiple linear regression and the orthogonal subspace projection are the two methods which were used to select such spectral bands [16]. The multiple linear regression approach used the combination of different bands to form a new feature which is dissimilar to the combination of the original bands. These new features are then concatenated with the original bands [16]. The orthogonal subspace projection approach involves constructing a transformation matrix in which the columns are defined by the initial set of bands [16]. This transformation matrix is applied to the remaining set of bands and the band which yields the maximum orthogonal component is labeled the most dissimilar band. The band is then concatenated with the original bands. These approaches where applied to an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) lunar lake image.

Riedmann *et al.* presented a supervised band selection method which seeks to find a band subset which is optimized in band location, band number, and band width [7].

The Transformed Divergence (TD) criterion function was used along with a bottom up approach in the final selection of spectral bands [7]. The bandwidth sizes were determined by comparing adjacent bands using the criterion function and were allowed to grow until a mean threshold was met [7]. The band number was determined by dividing the TD value of a subset of bands by the TD value achieved with the entire band set. The optimal number of bands was set to be equal to the dimension of the smallest band set [7]. The approach was applied to a CASI data set

Spectral band grouping has also been investigated for use with MCDF systems. Cheriyadat *et al*. implemented a bottom-up approach of spectral band grouping, where the product of BD and correlation was the group performance metric [17, 18]. Resulting spectral band groups were applied to a bank of classifiers and decisions were fused using qualified majority voting. The results were very promising and led to the work of Prasad *et al.* in the area of MCDF systems. The later work utilized the product of BD and average mutual information, as well as the product of Jeffries-Matusita distance and average mutual information [19]. The advantage of using average mutual information was the production of smaller band groups. Thus, the method could be used in scenarios where even less hyperspectral dataset is available for training the MCDF.

2.2.5 Multiresolution Analysis

Multiresolution analysis or wavelet analysis has become a basis for many feature extraction methods in the last couple of decades in signal processing. In this approach, the signal is decomposed by projecting it onto a scaled and translated version of a prototype function known as the mother wavelet [20]. One of the most efficient methods of implementing this type of multiresolutional transformation is the Discrete Wavelet

Transform (DWT) and the Wavelet Packet Decomposition (WPD) via the dyadic filter tree [20]. In these methods, the wavelet approximation and detail coefficients are produced by low-pass and high-pass filters. Subsets of detail and/or approximation coefficients are viewed as features and are selected based on a cost function or performance metric. The result is often referred to as a best-basis. The resulting coefficients can be used as features in a feature vector.

In current research, the DWT and WPD have become leading methods in extracting local and global features in hyperspectral remotely sensed data. The technique has been employed in the classification and compression of hyperspectral data.

Hsu *et al.* presented a method which used the WPD and DWT for feature extraction and optimization in hyperspectral target recognition in an agricultural application [21]. For the DWT feature extraction method, the authors selected approximation and detail coefficients in a linear and nonlinear manner such that the dimensionality reduction was achieved. For the WPD feature extraction method, the authors formed a best-basis of the wavelet coefficients by using cost functions based on entropy. Both approaches were applied to an AVIRIS hyperspectral data set.

Bruce *et al.* investigated the use of the DWT in the dimensionality reduction of hyperspectral data via the Haar mother wavelet. Area under ROC curves were used to determine the best subset of wavelet coefficients for optimum class separation [22]. The selected wavelet coefficients were combined and reduced via Fisher's LDA. The resulting reduced feature vector was classified by a maximum-likelihood classifier [22]. This approach was applied to handheld spectroradiometer data for a precision agriculture application [23].

Zhang *et al.* developed a remote sensing soil classification system employing the DWT as a feature extraction method, where the goal was the classification of three different pure soil textures [24]. The DWT was applied to each soil texture, and at each decomposition level the coarsest scaling coefficients and the wavelet detail coefficients were used to form feature vectors [24]. LDA was applied to the feature vectors for optimization. This approach was applied to ASD hyperspectral soil data.

Hsu *et al.* investigated the use of artificial neural networks (ANN) and wavelet based feature extraction in the classification of hyperspectral data. In this method, the hyperspectral data is decomposed via the Morlet mother wavelet [25]. The DWT coefficients are then input to the ANN. The ANN weights were adjusted by minimizing the least-square error in the training stage [25]. The wavelet parameters were selected by reducing the error. The wavelet network method was applied to an agricultural application using AVIRIS hyperspectral data.

Kaewpjit *et al.* used the DWT and a size four Daubechies mother wavelet in the automatic dimensionality reduction of hyperspectral imagery were the goal was compression [26]. The DWT was applied to each hyperspectral signature in the image. Then at each level of decomposition, the signature was reconstructed and compared to the original signature via correlation. Based on the correlation, a decomposition level was selected for each pixel [26]. The selected levels for each pixel were then combined, and the lowest level needed for each pixel became the decomposition level in which the hyperspectral data was reduced. This approach was applied to two sets of airborne hyperspectral data including AVIRIS.

Agarwal *et al.* proposed the use of the wavelet decomposition and PCA for dimensionality reduction of Atmospheric Infrared Sounder (AIRS) hyperspectral data where the goal was compression [27]. Multiresolution wavelet analysis was applied to each one-dimensional hyperspectral signature. For each, the signature was then reconstructed using only the low-pass portion of the decomposition. The reconstructed signature is then compared to the original signature via the similarity function of correlation [27]. The similarity metric is then stored in a histogram and the optimum level of decomposition is selected based on a percentage-threshold.

2.2.6 Multiclassifiers and Decision Fusion

Multisource classification is a process in which classification is performed by using remotely sensed data and data from other multiple sources. Recently, multisource classification accompanied by different data fusion techniques has become an attractive method in classifying remotely sensed data.

Watanachaturaporn *et al.* fused different data types by employing support vector machines (SVM) for multisource classification [28]. In this study, the Indian Remote Sensing Satellite Linear Imaging Self-scanning Sensor III image, digital elevation model, and a Normalized Difference Vegetation Index (NDVI) image were used as inputs into the multisource fusion classification system which employed a bank of classifiers [28]. The authors showed that SVM classifiers have great potential in the classification of multisource data.

Benediktsson *et al.* presented a multisource classification method employing neural networks and statistical modeling [29]. In this work, each data source is modeled using different statistical methods described in [29] and were fused using weighted

selection schemes involving the Consensus Theory and a Consensus-based voting and rejection schemes. These methods were applied to Landsat Thematic Mapper (TM) images and European Remote-Sensing Satellite (ERS-1) Synthetic Aperture Radar (SAR) imagery.

Multiclassifiers and decision fusion more recently have become a popular method employed in overcoming Hugh's Phenomenon in hyperspectral applications. Cheriyadat *et al.* proposed a classification technique which involves the feature extraction and decision level fusion of low-dimensional subspaces of hyperspectral data [30]. The high-dimensional data was decomposed in to subspaces by using correlation and discrimination of the classes. For each subspace, statistical discriminating features were extracted using Fisher's LDA [30]. Then each feature subspace was sent to its own classifier, and a decision was assigned to each subspace. The decision of each subspace was then fused using Qualified Majority Voting (QMV) [30]. This approach was applied to hyperspectral data in a vegetation classification application.

Fauvel *et al.* proposed a decision fusion technique which involves fuzzy decision rules for the classification of urban remote sensing images [31]. In this study, each pixel is classified using a bank of fuzzy and neural classifiers, and for each class a membership degree is assigned. These membership values are then modeled as a fuzzy set [31]. The global accuracy is then defined for each class by aggregating the different fuzzy sets. This approach was applied to two IKONOS images.

Saurabh *et al.* presented a divide-and-conquer approach that employed decision fusion in the exploitation of hyperspectral data [32]. In this technique, the hyperspectral space was partitioned into contiguous subspaces via the use of higher order statistical

information such as correlation and mutual information [32]. The each subspace is classified and the decisions are fused based employing majority voting, linear and logarithmic opinion pools, and adaptive weight assignments [32]. This method was applied to hyperspectral data collected with a handheld spectroradiometer as well satellite hyperspectral (Hyperion) data.

2.3 REFERENCES

[1] T. M. Lillesand, R. W. Kiefer, J. W. Chipman, *Remote Sensing Image Interpretation*, John Wiley & Sons Inc., 2004

[2] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, John Wiley & Sons, Inc., 2001.

[3] S. Theodoridis, K. Koutroumbas, *Pattern Classification*, Academic Press, 1999.

[4] S. Prasad, L.M. Bruce, "Limitations of Subspace LDA in Hyperspectral Target Recognition Applications," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, pp. 625-629, July 2007.

[5] A. Cheriyadat, L. M. Bruce, "Why Principal Component Analysis is not an Appropriate Feature Extraction Method for Hyperspectral Data," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 6, pp. 3420 – 3422, 2003.

[6] T. West, L. M. Bruce, "Multiclassifiers and Decision Fusion in the Wavelet Domain for Exploitation of Hyperspectral Data", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, vol. 1, pp. 4850-4853, July 2007.

[7] M. Riedmann, E. J. Milton, "Supervised Band Selection for Optimal Use of Data from Airborne Hyperspectral Sensors", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3, pp. 1770 – 1772, July 2003.

[8] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, "Regularization studies on LDA for face recognition," *Proc. 2004 International Conference on Image Processing*, vol. 4, pp 63-66, October 2004.

[9] S. Prasad, L.M. Bruce, "Overcoming the Small Sample Size Problem in Hyperspectral Classification and Detection Tasks," *Proc. IEEE International Geoscience and Remote Sensing Symposium*, vol. 5, pp. 381-384, July 2007.

[10] J. Ye, Q. Li, "A two-stage linear discriminant analysis via QR-Decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 6, pp. 929-241, June 2005.

[11] L.M. Bruce, H. Lin, "Projection Pursuits for Dimensionality Reduction of Hyperspectral Signals in Target Recognition Applications," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2, pp. 960-963, 2004.

[12] A. Ifarraguerri, C. Chang, "Unsupervised Hyperspectral Image Analysis with Projection Purusit," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 38, pp. 2529 – 2538, 2000.

[13] T. West, "Hyperspectral Dimensionality Reduction via Sequential Parametric Projection Pursuits for Automated Invasive Species Target Recognition," Masters Thesis, Mississippi State University, 2006.

[14] J. E. Ball, T. R. West, S. Prasad, L. M. Bruce, " Level Set Hyperspectral Image Segmentation using Spectral Information Divergence- based Best Band Selection," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4053-4055, July 2007.

[15] S. D. Backer, P. Kempeneers, W. Debruyn, P. Scheunders, "A Band Selection Technique for Spectral Classification", *Geoscience and Remote Sensing Letters*, vol. 2, no. 3, July 2005.

[16] Q. Du, H. Yang, "Unsupervised Band Selection for Hyperspectral Image Analysis", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 282 -285, July 2007.

[17] A. Cheriyadat, L.M. Bruce, A. Mathur, "Decision level fusion with best-bases for hyperspectral classification," *Proc. IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp 399-406,October 2003.

[18] S. Venkataraman, "Exploiting Remotely Sensed Hyperspectral Data Via Spectral Band Grouping For Dimensionality Reduction And Multiclassifiers," Master Thesis, Mississippi State University, 2005.

[19] S. Prasad, "Multi-classifiers and decision fusion for robust statistical pattern recognition with applications to hyperspectral classification," Doctoral Dissertation, Mississippi State University, 2008.

[20] S. Burrus, R. Gopinath, and H. Guo, *Introduction of Wavelets and Wavelet Transforms: A Primer*, Englewood Cliffs, NJ: Prentice-Hall, 1998.

[21] P. Hsu, Y. Tseng, "Feature Extraction of Hyperspectral Data Using the Best Wavelet Packet Basis", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3, pp. 1667-1669, June 24-28, 2002.

[22] L. M. Bruce, C.H. Koger, J. Li, "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction," *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331- 2338, 2002.

[23] Analytical Spectral Devices Fieldspec Pro Spectroradiometer specifications. Available on line at http://asdi.com/products_specificationsFSP.asp

[24] X. Zhang, V. Vijayaraj, N. H. Younan, "Hyperspectral Soil Texture Classification", *Proc. IEEE workshop on Advances in Techniques for Analysis of Remote Sensing Data*, pp. 182-186, 2003.

[25] P. Hsu, H. Yang, "Hyperspectral Image Classification Using Wavelet Networks", *IEEE Trans. Geoscience and Remote Sensing*, pp. 1767 – 1770, 2007.

[26] S. Kaewpijit, J. L. Morigne, T. El-Ghazawi, " Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis", *IEEE Trans. Geoscience and Remote Sensing*, vol. 41, no. 4, pp. 863-871, 2003.

[27] Agarwal , J. LeMoigne, J. Joiner, T. El-Ghazawi, F. Cantonnet, " Wavelet Dimension Reduction of AIRS Infrared (IR) Hyperspectral Data", *IEEE Trans. Geoscience and Remote Sensing*, vol. 5, pp. 3249-3252, 2004.

[28] P. Watanachaturaporn, P. K. Varshney, M. K. Arora, "Multisource Fusion for Land Cover Classification using Support Vector Machines", *Proc. 8th International Conference on Information Fusion*, vol. 1, pp. 614 – 621, July 2005.

[29] J. A. Benediktsson, I. Kanellopoullos, "Classification of Multisource and Hyperspectral Data Based on Decision Fusion", *IEEE Trans. Geoscience and Remote Sensing*, vol. 37, vol. 3, May 1999.

[30] A. Cheryadat, L. M. Bruce, "Decision Level Fusion with Best Bases for Hyperspectral Classification", *Proc. IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, vol. 2, pp. 1245 -1248, October 2003.

[31] M. Fauvel, J. Chanussot, J. A. Benediktsson, "Decision Fusion for the Classification of Urban Remote Sensing Images", *IEEE Trans. Geoscience and Remote Sensing*, vol. 44, no. 1, pp. 2828-2838, Oct. 2006.

[32] S. Prasad, L. M. Bruce, "Decision Fusion with Confidence-Based Weight Assignment for Hyperspectral Target Recognition," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1448 – 1456, May 2008.

CHAPTER 3

COMBINING DISCRETE WAVELET TRANSFORM FEATURE EXTRACTION WITH MULTI-CLASSIFIERS AND DECISION FUSION FOR IMPROVED HYPERSPECTRAL CLASSIFICATION

3.1 Introduction

With their increasing affordability and potential for discriminating subtly different ground cover classes, hyperspectral sensors are becoming more attractive and more commonly used for a variety of remote sensing applications. In automated target recognition (ATR) systems, features are extracted from the hundreds to thousands of narrow, contiguous spectral bands. The increase in available spectral features can cause the ATR system to suffer the "curse of dimensionality" when amount of labeled training data (ground truthed pixels) is overly limited. That is, the number of features produced by the hyperspectral sensor cannot be supported by the amount of available training data.

Many different techniques have been investigated in the dimensionality reduction and feature extraction of hyperspectral data. Recently, spectral band grouping combined with multiclassifiers and decision fusion (MCDF) have become a very promising solution to the dilemma of the over-dimensionality of hyperspectral data [1-3]. An example block diagram of MCDF is shown in Figure 3.1. This approach involves the partitioning of the hyperspectral space into lower dimensional subspaces. Then the spectral band groups are sent to a bank of classifiers, one classifier for each group. Next, the classifications made

by the classifiers are fused using decision fusion to produce one final classification. The weights used in the system's decision fusion stage typically take into account the reliability of each group/classifier combination to accurately classify a pixel.

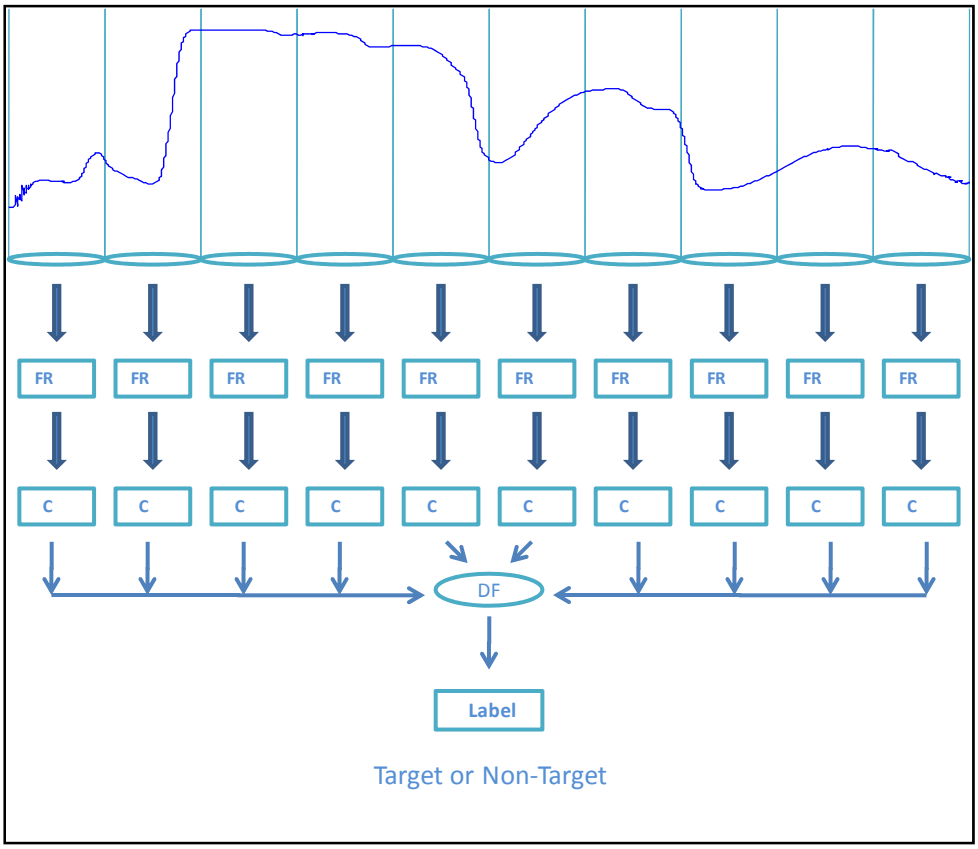


Figure 3.1 Block diagram representation of spectral band grouping, combined with multiclassifiers and decision fusion. FR, C, and DF notate feature reduction, classification, and decision fusion, respectively

One of the major potential drawbacks of the current MCDF approach is related to the band grouping method used in the initial stage. The spectral band grouping method has a limited ability to extract large scale or global features from the hyperspectral signatures. Typically, two approaches are usually employed during spectral band

grouping, either a fixed size sliding window approach or a bottom up approach. In the case of a fixed-sized window grouping, non-overlapping equally sized groups of bands are formed. In the bottom-up approach, the system views each spectral band as a group in the initial stage. Adjacent groups are merged to form larger groups (i.e. the groups are then grown across the spectrum) until some predefined stopping criteria is met. The merging of the groups is stopped when a pre-defined classification performance metric is no longer adequately increasing or the group size becomes larger than what the training data can support. Both approaches involve the grouping of local spectral bands, i.e. small scale or localized features in the hyperspectral signature. In previous hyperspectral research, multiresolution analysis (wavelet analysis) has been shown to extract both local and global spectral features successfully in target recognition [4-7].

In this work, the discrete wavelet transform (DWT) multiresolutional transformation is applied to the hyperspectral space. The DWT is implemented using the dyadic filter tree approach. Two-channel filter banks are used to obtain the approximation and detail wavelet coefficients via low-pass and high-pass filters. In this paper, different types of mother wavelets, including the Daubechies family of mother wavelets, will be investigated to study the approach's sensitivity to mother wavelets. It can be shown in [8] and [9] that the Haar wavelet, equivalent to Daubechies-1, is often one of the optimal mother wavelets when classification accuracy is the benchmark. Thus, it is anticipated that the Haar will result in pseudo optimum results. The maximum level of decomposition will be varied to investigate its effect on the overall classification accuracy.

Each set of wavelet detail coefficients, along with the final set of approximation coefficients, are considered as potential feature vectors. From the potential feature vector set, feature vectors are selected to aid in the classification based on various performance metrics, include supervised metrics (e.g. Bhattacharyya Distance (BD) and unsupervised metrics (e.g. entropy). The dimensionality of each selected feature vector is reduced via Fisher's linear discriminant analysis (LDA), and each reduced vector is input to an independent classifier in a MCDF scheme. Maximum-likelihood classifiers are used in this study. The classification outputs are fused using a standard decision fusion method known as qualified majority voting.

This chapter is organized as follows. In section 3.2, a brief description of DWT multiresolution transformation analysis is presented. In section 3.3, the details of the DWT MCDF proposed system employed in this work is explained which include the different feature extraction and feature selection techniques. Section 3.4, contains the specification of the handheld spectroradiometer and the description of the hyperspectral data sets investigated in this experiment. The experimental results of the proposed system are presented in section 3.5 and conclusions are drawn in section 3.6.

3.2 Discrete Wavelet Transform (DWT)

Wavelet analysis was established for the analysis of functions or signals which are non-stationary. Stationary signals or functions are periodic and can be predictable in most cases. These characteristics allow the signals or functions to be represented as combinations of sine or cosine waves with different frequencies which can be analyzed by methods such as Fourier analysis. Non-stationary signals and functions lack the characteristics of being periodic and in most cases cannot be predicted. These functions

or signals cannot be analyzed by common methods, such as Fourier analysis, because of their transient characteristics. The wavelet, also known as the "small wave", utilizes its characteristics of having its energy concentrated in time to analyze signals which are transient and non-stationary [10]. Wavelet analysis is performed by projecting a signal or a function onto a set of basis functions [11]. This set of basis functions, $\{\psi_{(a,b)}\}$, is generated by scaling and translating the prototype wavelet or what is called the mother wavelet, $\psi(x)$, in time which is described by the following,

$$\psi_{(a,b)}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \tag{3.1}$$

where $\boldsymbol{a} > 0$ and $\boldsymbol{b} \in \mathbb{R}$ and $\frac{1}{\sqrt{a}}$ is a normalizing factor. In equation 3.1, parameter $\boldsymbol{a}$ is the scaling parameter for the wavelet function, which describes the frequency information of the signal, and parameter $\boldsymbol{b}$ is the translation parameter, which relates the time information of wavelet analysis. This parameter defines the location of the wavelet function as it is applied to the signal. The scaling and translating of the mother wavelet, $\psi(x)$, generates the set of basis functions, $\{\psi_{(a,b)}\}$, in which the set have a similar shape of the mother wavelet. The global and local information of the signal is extracted by the scaling parameter of the mother wavelet. The scaling of the wavelet function either dilates the mother wavelet, and resulting analysis provides global information about the signal, or it compresses the mother wavelet, and resulting analysis provides local details about the signal. To classify a function as being a wavelet function, the function must oscillate, must have average value of zero, and must have finite support [10]. The wavelet functions for the DWT are represented by

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k) \tag{3.2}$$

and the DWT wavelet coefficients are obtained by

$$W_{jk} = \langle f(x), \psi_{jk}(x) \rangle \tag{3.3}$$

where $f(x)$ is the function, $\psi_{jk}(x)$ are the wavelet functions, and W_{jk} are the wavelet coefficients.

The property of multiresolution analysis (MRA) is an important property for any wavelet system. This property allows for the decomposition of a signal to be an iterative decomposition of resulting approximation signals [10]. If this property is met in a wavelet system, the DWT can be implemented using a tree-structured algorithm known as a dyadic filter bank, or dyadic filter tree, and is shown in Figure 3.2. This implementation is utilized in most wavelet system because it provides a computationally efficient method of obtaining the wavelet coefficients. This method decomposes a signal at each scale by applying a two-channel filter bank which are low-pass and high-pass filters. This filter bank decomposes the signal into approximation and detail coefficients.

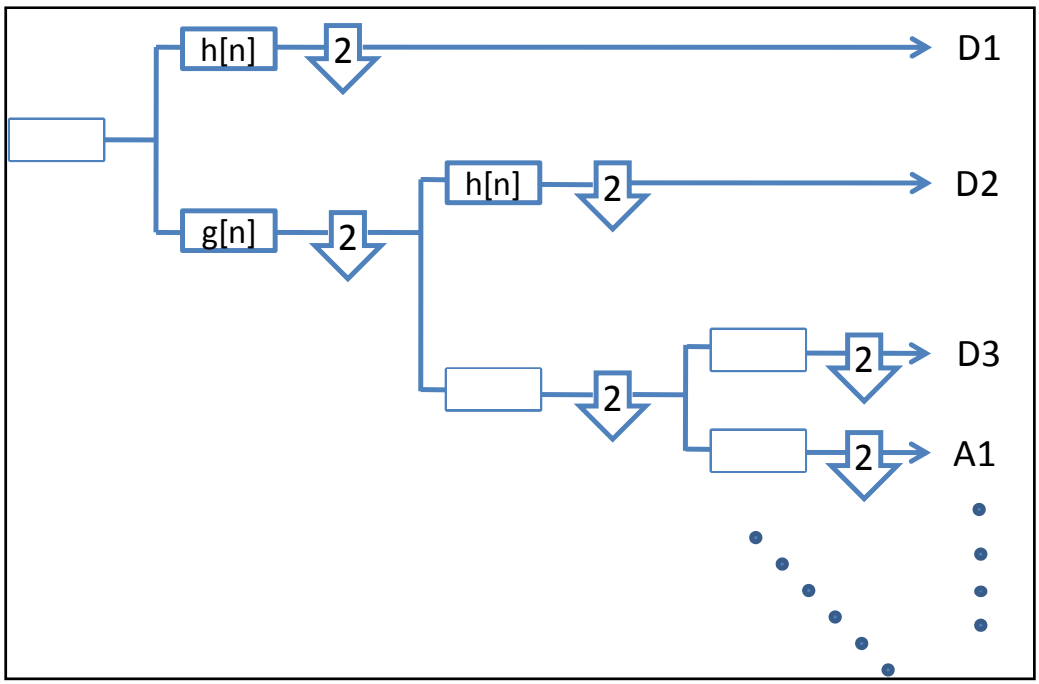


Figure 3.2 Dyadic filter tree (f[n] is the input signal, h[n] is the high-pass filter, g[n] is the low-pass filter D denotes detail coefficients and A denotes approximation coefficients).

Each level of the filter tree corresponds to a dyadic scale of the wavelet decomposition, producing approximation coefficients via the low-pass filter and detail coefficients via the high-pass filter. The approximation coefficients are decomposed again to form a new level of decomposition with new approximation and detail coefficients. This process is repeated until the maximum level of decomposition is met.

3.3 Discrete Wavelet Transform in Framework of Multiclassifiers and Decision Fusion

3.3.1 System Overview

A combined DWT MCDF scheme is proposed to circumvent the curse of dimensionality while extracting features which represent both the local and global characteristics of the hyperspectral signature. Figure 3.3 illustrates a block diagram of the proposed system. The proposed system applies the DWT to the hyperspectral signature in either a supervised or unsupervised manner depending on the utilized feature selection method. The DWT is computed via a dyadic filter tree. Then each set of wavelet detail and approximation coefficients are considered as potential feature vectors. Features are then extracted from the potential feature vectors to aid in the overall classification of the system based on either supervised or unsupervised metrics.

Individual feature vectors are then sent to independent classifiers. The classifier used in this work is the maximum-likelihood classifier. The objective of any classifier is to use the information from a feature set to correctly assign a class label to a sample. Maximum-likelihood classifiers are supervised classifiers which assign labels to samples based on a maximum probability [12]. The training samples for this classifier are assumed to have a normal distribution. The selection of this classifier was based on the distribution of the feature after feature selection and extraction. After the preprocessing stage of the system, the set of features have a normal distribution which meets the assumption for the training of the maximum-likelihood classifier. The maximum-likelihood decision rule for data samples having equal probabilities of occurring is defined by the following [12]:

$$M(w_i|\vec{x}) = -\frac{1}{2}\log_e|\Sigma_i| - \left[\frac{1}{2}(\vec{x} - \overrightarrow{\mu_i})^T\Sigma_i^{-1}(\vec{x} - \overrightarrow{\mu_i})\right] \quad (3.4)$$

where $\overrightarrow{\mu_i}$ is the mean vector for class i and Σ_i is the covariance matrix of class $\boldsymbol{i}$. The maximum likelihood decision rule for data samples having unequal probabilities of occurring is defined by the following [13]:

$$M(w_i|\vec{x}) = \log_e p(w_i) - \frac{1}{2}\log_e|\Sigma_i| - \left[\frac{1}{2}(\vec{x} - \overrightarrow{\mu_i})^T\Sigma_i^{-1}(\vec{x} - \overrightarrow{\mu_i})\right] \quad (3.5)$$

where $p(w_i)$ is the appropriate a priori probability for class $\boldsymbol{i}$.

After each feature set has been classified, the decision from each classifier is then fused into a single class label. The decision fusion method employed in this work is the simple majority vote scheme. This scheme assumes all classifiers have equal weight in the overall classification regardless of any *a priori* information. However, it should be noted that there are other decision fusion methods available which take in account *a priori* information such as linear and logarithmic opinion pools. In this work, the majority vote decision fusion scheme is employed because of its simplicity, allowing us to focus on the intial phases of the DWT MCDF approach (wavelet decomposition and coefficient grouping). The multiclassifier majority vote scheme is defined by the following:

$$w = \text{argmax}_{i\in\{1,2,\ldots,C\}}\, N(i) \quad (3.6)$$

where $\boldsymbol{w}$ is the class label from one of the C possible classes for the test pixel, and $N(i)$ is the number of times class i was detected in the bank of classifiers. The result is a single, final classification for the input hyperspectral signature.

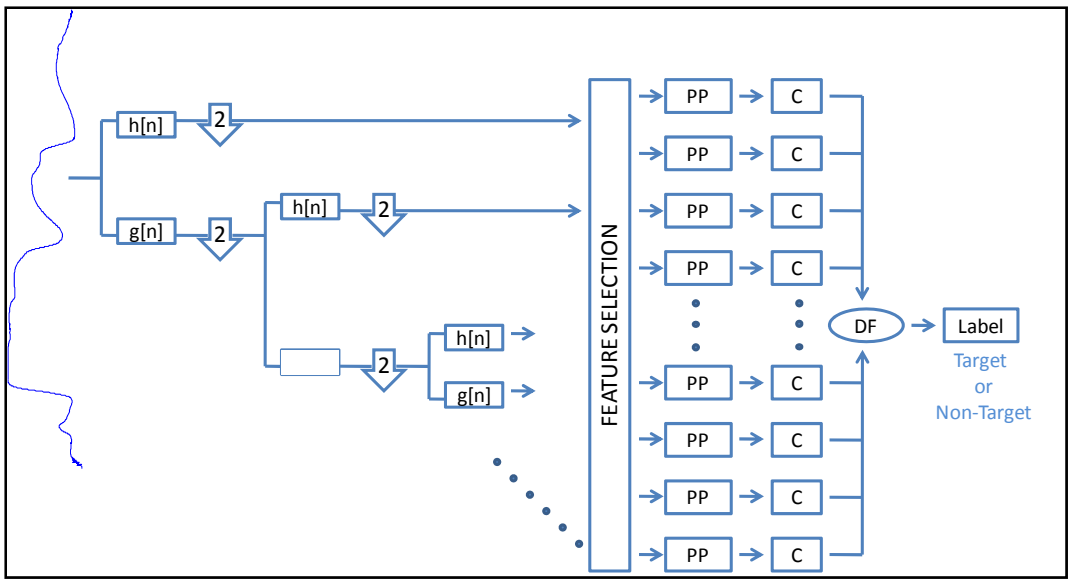


Figure 3.3 Block diagram representation of the DWT MCDF framework, where preprocessing (PP) is like LDA, C is the classifier and DF is the decision fusion scheme

3.3.2 Wavelet decomposition

Wavelet analysis has the ability to resolve the local and global information within a hyperspectral signature. In this work, the DWT is the wavelet analysis method employed for feature extraction. The DWT was implemented using the dyadic filter tree as described in section 3.2. The dyadic filter tree implementation was utilized due it computational efficiency, and is implemented by a low-pass and high-pass filter in a two channel scheme. The selection of the mother wavelet could play a major role in the overall performance of the proposed system. To better understand the effect of mother wavelet selection on the optimality of resulting, a mother wavelet sensitivity study was conducted. The mother wavelets utilized in this study are orthonormal, because of the dyadic filter tree requirements. The mother wavelets investigated in the study are the

Daubechies family of wavelets which ranged from Daubechies 1 to Daubechies 10. This family of wavelets was selected for analysis because the Haar mother wavelet, which is equivalent to Daubechies1, has been shown to be optimal or at least pseudo-optimal in other hyperspectral ATR research [8-9]. Also, the maximum level of decomposition for this system may play a significant role in the system's overall performance. The role of the decomposition level is noteworthy because each set of wavelet coefficients are considered as a feature vector. As the level of decomposition increases, the number of potential features increase as well, which may introduce the "curse of dimensionality". To ensure the level of decomposition is optimal, a sensitivity study is performed.

3.3.3 Wavelet Coefficient Feature Space Partitioning

The next stage in the proposed system involves the selection and grouping of wavelet features. In most classification applications, class separation is the governing benchmark in feature selection and extraction. In previous work involving feature selection and extraction in the wavelet domain, [13-14], metrics such as entropy, area under receiver operating characteristics (ROC) curves, and the Bhattacharyya distance (BD) have been investigated as performance metrics. In this study, both supervised and unsupervised techniques are investigated.

3.3.3.1 Coefficient concatenation with fixed-size Contiguous Partitioning

The coefficient concatenation with fixed-size contiguous partitioning (CONCAT) is a feature selection/grouping method similar to the common hyperspectral band grouping [1, 2, 15]. This method is investigated to determine if the combination of coefficients at any scale could aid in the overall classification accuracy of hyperspectral

data. After the hyperspectral data is decomposed into approximation and detail coefficients, these coefficients are concatenated into one large feature vector. These features are then partitioned based on a fixed-size, non-overlapping, sliding window. The window size governed by a 3-to-1 rule and by the amount of available training data, i.e the number of labeled hyperspectral signatures available for training. For instance if the number of available training signatures is $\boldsymbol{N}$ and the partitioning window size is $\boldsymbol{X}$, then the following criteria should be met,

$$X \leq 1/3\ N\ . \tag{3.7}$$

As with spectral band grouping, each partitioned feature space (set of wavelet coefficients) is then considered as a feature vector.

The dimensionality of each individual feature vector is then reduced by projecting the feature vector onto a lower dimensional subspace via LDA. Then each reduced subspace is treated as an independent feature vector and each subspace is the input to an independent classifier in the MCDF system.

3.3.3.2 Scalar Partitioning

The scalar (SCALAR) feature selection/grouping method involves utilizing the global and local feature extraction properties of the DWT to select optimum features. After the hyperspectral data set is decomposed into approximation and detail coefficients, each set of coefficients at each decomposition scale is considered as a feature vector (i.e. each set approximation and detail coefficients are considered feature vectors.). Then each feature vector is input to an independent classifier in the MCDF system.

This method takes advantage of the feature extraction properties of the DWT, but may introduce the "curse of dimensionality" due to the dimensionality of some sets of wavelet approximation or detail coefficients. That is, smaller scale detail and approximation coefficients (generated in earlier stages of the dyadic filter tree) have a larger number of coefficients, and if there is not enough training data to support the number of coefficients in these sets, the "curse of dimensionality" will still be present.

3.3.3.3 Scalar Subspace Partitioning

The scalar subspace (SUBSPACE) feature selection/grouping method is a combination of the SCALAR and CONCAT methods. The SCALAR method is first applied. Each set of wavelet detail and approximation coefficients is then evaluated for determine if CONCAT will be applied to that set. A 3-to-1 rule of thumb is used to determination if the number of coefficients in the set meets the 3-to-1 criteria (as defined in equation 3.7), then CONCAT is not applied. However, if the number of coefficients in the set does not meet the 3-to-1 criteria, CONCAT is applied to that particular set of wavelet coefficients. This approach takes advantage of the scalar subsetting of features via the DWT while also accounting for the possibility of introducing of the "curse of dimensionality" caused by high dimensional sets of wavelet coefficients.

3.3.3.4 Scalar Partitioning with Metric-Based Selection

The scalar partitioning with metric-based selection is a method that approaches the wavelet feature selection/grouping via performance metrics, where the metrics could be either supervised (BD) or unsupervised (ENTROPY).

The supervised (BD) approach selects/groups wavelet coefficients based on maximized class separation. Thus, the supervised approach requires labeled training data during its training phase. One of the commonly used supervised metrics in hyperspectral feature extraction is the Bhattacharyya Distance (BD) and is used in this study. The BD is a special form of another distance metric known as the Chernoff distance. These methods seek to find the upper bounds of the error of probability by finding the parameters that produce the maximum value for the distance $\mu(s)$ [16]. For a two class problem, the Chernoff distance is define as the following:

$$\mu(s) = \frac{s(1-s)}{s}(M_2 - M_1)^T[s\Sigma_1 + (1-s)\Sigma_2]^{-1}(M_2 - M_1) + \frac{1}{2}\ln\frac{|s\Sigma_1+(1-s)\Sigma_2|}{|\Sigma_1|^s|\Sigma_2|^{1-s}} \quad (3.8)$$

where M_i is the mean of class i and the Σ_i is the covariance for class i. The BD is the special case of the Chernoff distance where s=1/2. The BD for a 2 class problem is then defined as the following:

$$\mu\left(\frac{1}{2}\right) = \frac{1}{8}(M_2 - M_1)^T\left[\frac{\Sigma_1+\Sigma_2}{2}\right]^{-1}(M_2 - M_1) + \frac{1}{2}\ln\frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}} \quad (3.9)$$

The unsupervised (ENTROPY) approach selects/groups features without the use of labeled training data. Note, however, this approach still requires a training phase; it simply uses unlabeled training data. One of the most commonly used unsupervised metrics in wavelet applications is entropy and is used in this study. Entropy is a leading metric for the selection of wavelets scales and decomposition levels in compression and speech applications. Entropy measures the amount of uncertainty or information that a source contains [17]. This uncertainty is defined by the probability distribution of the sources. Suppose that an n-dimensional feature vector is represented by $\overrightarrow{Fv}$ and the

probabilities for $[\overrightarrow{Fv}]$ are $[q_1 \cdots q_n]$, then the entropy of the feature vector is then defined as the following:

$$H = -\sum_{i=1}^{n} q_i \log q_i \tag{3.10}$$

where the estimates of the probabilities $[q_1 \cdots q_n]$ of the features are obtained by the histogram of feature vector

In this work, metric-based feature groups, whether they are supervised (BD) or unsupervised (ENTROPY) metrics, are selected based on that group's metric and its relation to the mean metric of all groups. That is, let ***E*** be the collection of performance metric values for all groups of wavelet coefficients, where ***E*** is defined by $[E_1 \cdots E_n]$ and *n* is the number of scales in the decomposition. Let $\overrightarrow{Fv}$ be a feature vector representing a set of wavelet coefficients that is selected for input to the MCDF scheme, then the following criteria must be met:

$$E > \mu_e + n\sigma \quad E > \mu_e + n\sigma_e \tag{3.11}$$

where, μ_e and σ_e are the mean and standard deviation of ***E***, respectively. The parameter *n* may be set to any integer value. The higher (or lower) the value of *n*, the more (or less) restrictive the selection process, i.e. increasing (or decreasing) *n* decreases (or increases) the number of feature sets passed through to the MCDF scheme.

With the metric-based feature partitioning methods, there is a chance that the follow-on pre-processing stage of MCDF (namely LDA in our study) may not be appropriate, due to the fact that the dimensionality of the feature set is too low. LDA reduces the dimensionality of a feature vector by transforming the data on to a lower dimensional subspace that has a dimensionality of ***C – 1***, where ***C*** is the number of

classes. If the feature set's dimensionality is less $\boldsymbol{C}$, then LDA is not applied. If the feature set's dimensionality is greater than or equal to $\boldsymbol{C}$, then LDA is applied.

3.4 Experimental Case Study

3.4.1 Data

The proposed methods are applied to experimental hyperspectral data for an agricultural application, namely the early detection of a disease known as soybean rust (*Phakopsora pachyrhizi)* in soybean (*Glycine max*) crops [19]. Soybean rust is a windborne pathogen which can be transmitted over large areas in a matter of weeks causing widespread damage [20]. The ability to rapidly detect soybean rust onset is critical to the US economy, and agencies such as the U.S. Department of Agriculture (USDA) and U.S. Department of Homeland Security (DHS) are particularly interested in this very challenging remote sensing problem.

The hyperspectral dataset was collected using the Analytical Spectral Device (ASD^{TM}) Fieldspec Pro handheld spectroradiometer [21]. The ASD has a spectral range of 350 – 2500 nm, spectral resolution of 1-1.2 nm, and uses a single 512 element silicon photodiode array for sampling 350 - 1000 nm and two separate, graded index Indium-Gallium-Arsenide photodiodes for the 1000 - 2500 nm range.

For this study, two datasets of ASD readings of soybean, both control and diseased, were used. The first dataset was collected over a two week period in a green house outside the city of Encarnacion, Paraguay, in 2005 with the humidity at 100% and the temperature kept close to 80 - 85 F [19]. For this study, 678 hyperspectral signatures were used for

evaluation, 320 observations of the control soybean and 358 observations of the inoculated soybean. The second dataset was collected in January of 2008 in Stoneville, MS, during an actual outbreak of the disease in a commercial crop setting. Eighty-five observations were collected, 19 observations of the non-diseased soybean and 66 of the diseased soybean.

The parameters of the system (metric, mother wavelet, the level of decomposition, and grouping) were optimized by employing the 2005 data. Then the 2008 dataset was used to test the system. Thus the robustness of the system, i.e. the system's sensitivity to training data is tested

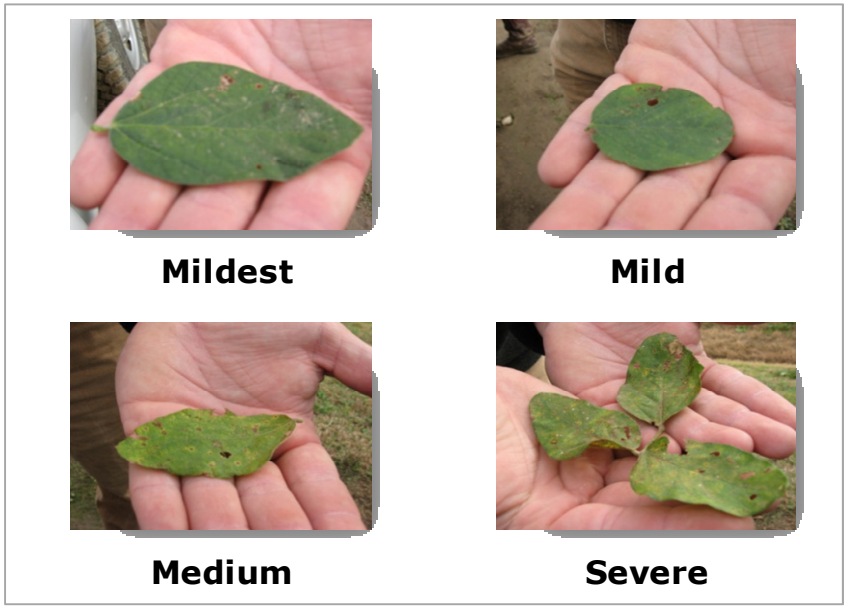
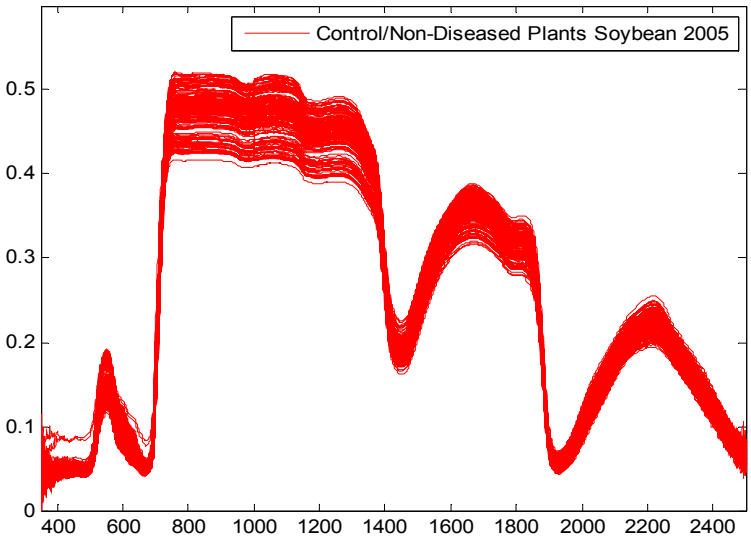
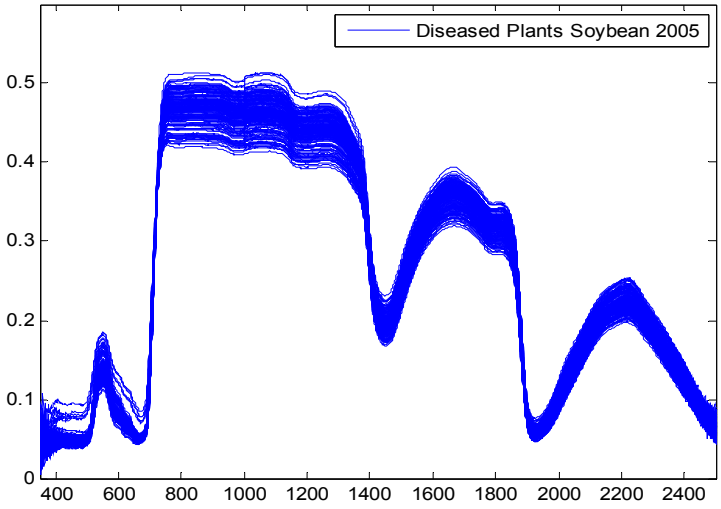


Figure 3.4 Examples of soybean plants, including control/non-diseased and diseased. Photos correspond to handheld hyperspectral data collected in January 2008 in Stoneville, MS

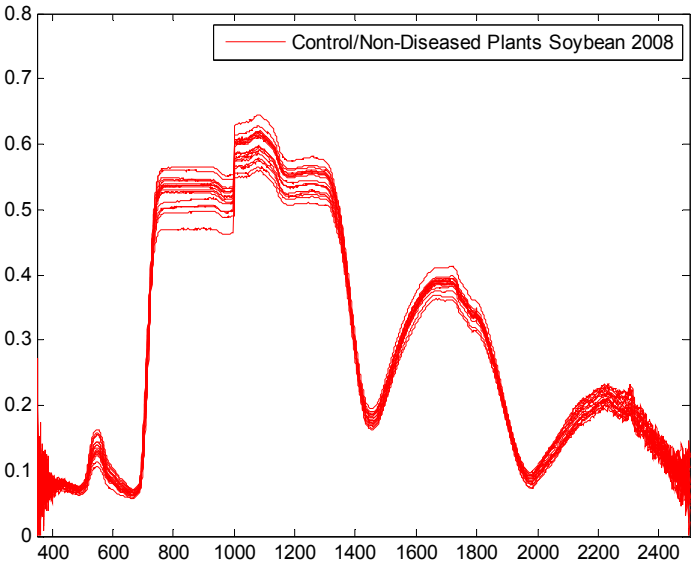


(a)

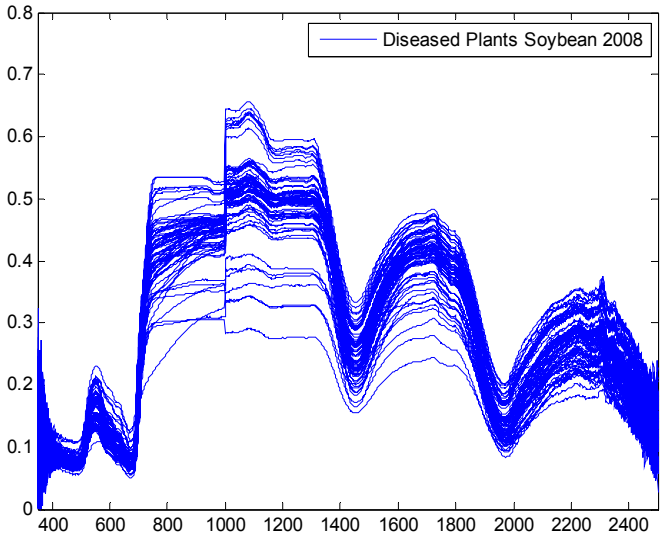


(b)

Figure 3.5 Example hyperspectral reflectance signatures of soybean plants, (a) control/non-diseased plants, 2005, (b) diseased plants, 2005, (c) control/non-diseased plants, 2008, (b) diseased plants, 2008.



(c)



(d)

Figure 3.5 (continued)

3.4.2 Testing Methods

The testing method employed in this work is the N-fold cross validation method [18]. When testing the supervised techniques, the available hyperspectral data is partitioned into labeled testing and labeled training data. When testing the unsupervised techniques, the available hyperspectral dataset is partitioned into labeled testing and unlabeled training data. In both scenarios, the training and testing data are sequestered, such that they are mutually exclusive to ensure unbiased results.

3.4.3 Experimental Results and Discussion

Figure 3.6 shows the results of the mother wavelet selection sensitivity analysis, utilizing the 2005 hyperspectral dataset. The results are reported in terms of overall classification accuracy and 95% confidence intervals vs. Daubechies family mother wavelet. The decomposition level for each mother wavelet was determined by the length of the hyperspectral signature and the type of mother. From these results, it is clear that the unsupervised metric-based DWT-coefficient grouping method consistently outperforms both the non-metric-based and the supervised metric-based approaches.

Figure 3.7 shows the results of the DWT decomposition level sensitivity analysis, utilizing the 2005 hyperspectral dataset. The results are reported in terms of overall classification accuracy and 95% confidence intervals vs. decomposition level for metric-based feature selection, including both BD and ENTROPY. These results are for the Haar mother wavelet. The results for the ENTROPY approach show that it is insensitive to DWT decomposition level. However, the results for the BD approach show that it is somewhat sensitive to DWT decomposition level, with the sensitivity being greater for center levels of decomposition.

Figure 3.8 shows the results of a comparison analysis of the proposed DWT-MCDF methods to conventional spectral-based and DWT-based single classifier approaches. The mother wavelet is Haar, and the DWT decomposition level is 11, i.e. the maximum level of decomposition for Haar mother wavelet. Results are reported in terms of overall classification accuracy and 95% confidence intervals. The conventional approaches used for comparison purposes include SLDA of the original spectra and SLDA in the DWT domain. The analysis was conducted utilizing the 2008 hyperspectral dataset. That is, the ATR systems were trained on 2005 hyperspectral data and tested on 2008 hyperspectral data. Both SLDA and DWT SLDA result in overall accuracies of around 40%. Thus, the single classifier approach does not perform well on this difficult dataset (very similar vegetation classes with relatively limited training data), regardless of whether the classification is conducted in the original spectral domain or the DWT domain. Three of the proposed methods (CONCAT with LDA, SUBSPACE, ENTROPY, and BD), however, perform quite well on this difficult dataset. These methods result in overall classification accuracies of 75-80%.

Figure 3.9 shows the results of a compassion analysis of the proposed DWT-MCDF methods to conventional spectral-based and DWT-based single classifier approaches. The mother wavelet is Daubechies-8, and the DWT decomposition level is 8, i.e. the maximum level of decomposition level for a Dauhechies-8 mother wavelet. Results are reported in terms of overall classification accuracy and 95 % confidence intervals. The conventional approaches used for comparison purposed include SLDA of the original spectra, SLDA in the DWT domain, and MCDF in the original spectra. The analysis was conducted utilizing the 2008 hyperspectral dataset. Both SLDA and DWT

result in overall accuracies of around 35 % - 44 %. Thus, the single classifier approach does not perform well on this difficult dataset (very similar vegetation classes with relatively limited training data), regardless of whether the classification is conducted in the original spectral domain or the DWT domain. Five of the proposed methods (CONCAT with LDA, SCALAR, SUBSPACE, ENTROPY, and BD), however, perform quite well on this difficult dataset. These methods result in overall classification accuracies of 70 % - 80 %.

Next we consider a more detailed analysis of the metric-based feature selection approaches. Figures 3.10 and 3.11 show the performance metric values vs. DWT decomposition level, with thresholds indicated for the metric mean and the metric mean plus one standard deviation, i.e. $\mu_e + n\sigma_e$ for n=[0,1]. Figures 3.10 and 3.11 shows the results for Haar and Daubechies-8 mother wavelets, respectively. For the Haar mother wavelet, the mid-range decomposition levels are producing the highest performance metrics. Using a threshold of with n as a negative integer results in virtually all of the wavelet coefficient groups being passed through to the MCDF system. If this is the case, it is advised that a more sophisticated decision fusion method, preferably one that utilizes *a priori* classification information, be used. In the other extreme, when $n > 1$, none of the wavelet coefficient groups are being passed through to the MCDF system; thus it is impractical. When n=0, approximately 3 to 6 wavelet coefficient groups are passed on to the MCDF system. When n=1, only 1 or 2 wavelet coefficient groups are passed on to the MCDF system, thus somewhat defeating the use of the multiclassifier approach in these cases.

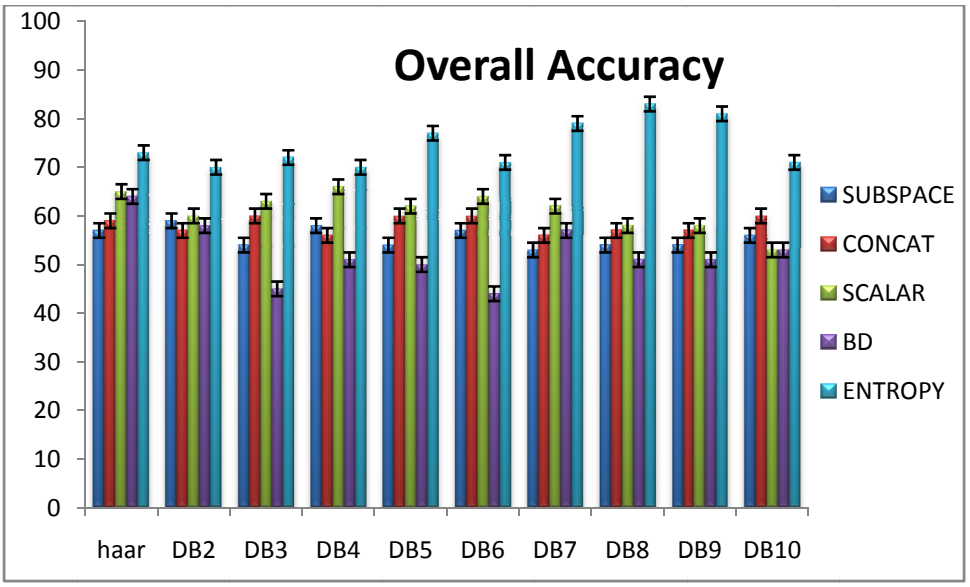


Figure 3.6 Results of mother wavelet selection sensitivity analysis, utilizing 2005 hyperspectral dataset. Results reported in terms of overall classification accuracy and 95% confidence intervals vs. Daubechies family mother wavelet

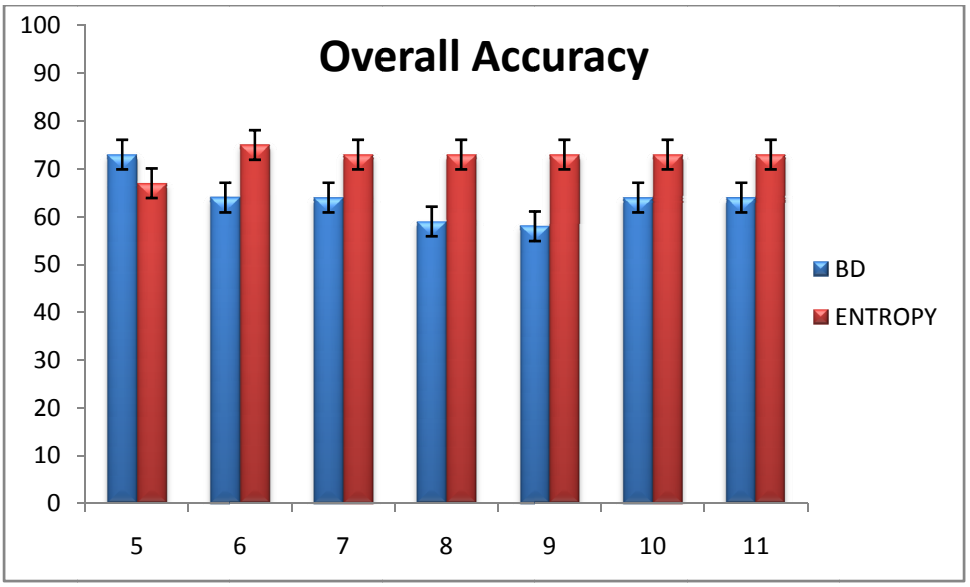


Figure 3.7 Results of DWT decomposition level sensitivity analysis, utilizing 2005 hyperspectral dataset. Results reported in terms of overall classification accuracy and 95% confidence intervals vs. decomposition level for metric-based feature selection: BD and ENTROPY

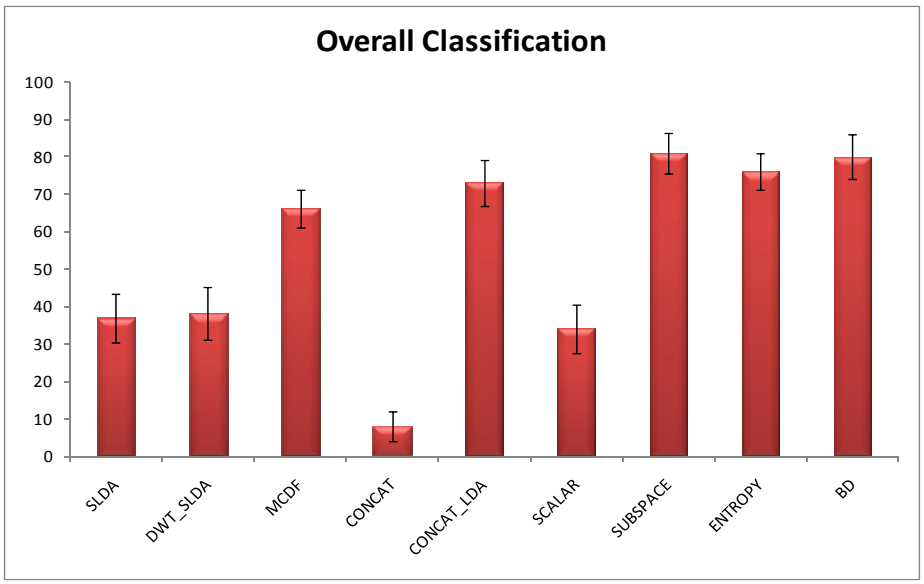


Figure 3.8 Comparison analysis of proposed DWT-MCDF methods to conventional spectral-based and DWT-based single classifier approaches, utilizing 2008 hyperspectral dataset. Mother wavelet is Haar and DWT decomposition is 11. Results reported in terms of overall classification accuracy and 95% confidence intervals

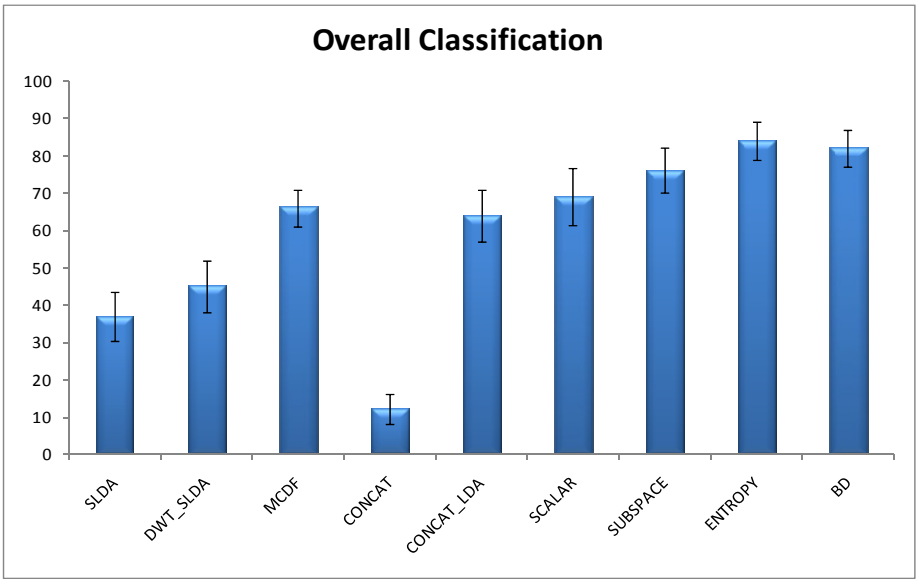


Figure 3.9 Comparison analysis of proposed DWT-MCDF methods to conventional spectral-based and DWT-based single classifier approaches, utilizing 2008 hyperspectral dataset. Mother wavelet is Daubechies-8 and DWT decomposition level is 8. Results reported in terms of overall classification accuracy and 95% confidence intervals.

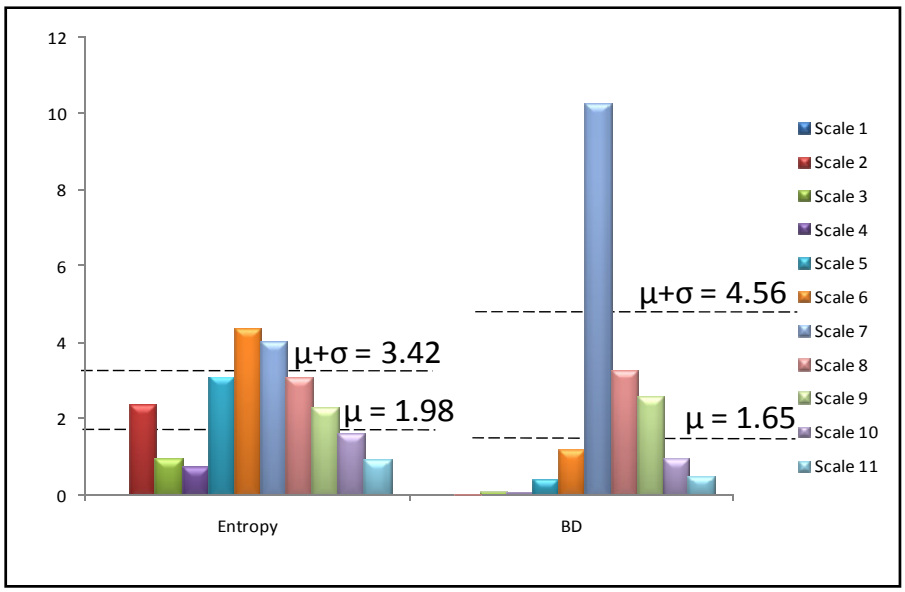


Figure 3.10 Performance metric values vs. Haar DWT decomposition level, with thresholds indicated for mean metric and mean plus one standard deviation. Results are show for 2005 hyperspectral dataset, i.e. training of the ATR systems.

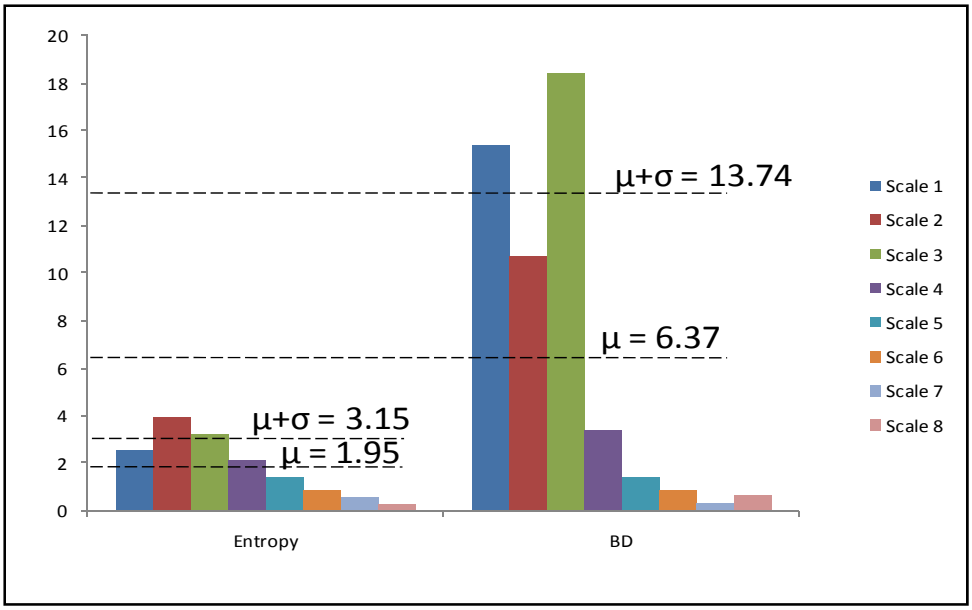


Figure 3.11 Performance metric values vs. Daubechies-8 DWT decomposition level, with thresholds indicated for mean metric and mean plus one standard deviation. Results are show for 2005 hyperspectral dataset, i.e. training of the ATR systems.

3.5 Conclusion

In this work, the DWT multiresolutional transformation is combined with the MCDF approach, as a means to overcome the shortcoming of current MCDF approaches. These shortcomings stem from the fact that MCDF approach used only localized groups of spectral bands, thus losing global or large scale features from the hyperspectral signature. To this end, the authors developed, implemented, and tested five DWT coefficient grouping/selection methods: CONCAT, SCALAR, SUBSPACE, ENTROPY, and BD.

Assessment of these newly developed approaches was conducted using experimental hyperspectral measurements for an agricultural application, where the ground cover classes were soybean with varying levels of soybean rust infestations. The parameters of the system (metric, mother wavelet, the level of decomposition, and grouping) were optimized by employing the 2005 data. Then the 2008 data is used to test the system. Thus, the robustness of the system, i.e. the system's sensitivity to training data is tested.

A sensitivity analysis of the five newly developed approaches was conducted, assessing the performance (in terms of classification accuracy) vs. the mother wavelet selection and DWT decomposition level. The BD method was more sensitive to mother wavelet selection, as compared to the other DWT coefficient grouping/selection methods. However, none of the five approaches demonstrated a significant sensitivity to mother wavelet selection. For the ENTROPY approach, the Daubechies-8 mother wavelet slightly outperforms the other mother wavelets investigated. However, it could be argued that the simplicity of the Haar mother wavelet could outweigh the slight increase in

performance gained by using a more complicated mother wavelet. The simplicity of the Haar mother wavelet could result in faster implementations, as well as potentially safe-guarding the ATR system against over-training. The BD method was more sensitive to the selection of decomposition level, as compared to the ENTROPY method. The ENTROPY approach mostly outperformed the BD method and was constant for any level of decomposition. One interesting point of this analysis is the performance of the unsupervised method ENTROPY, which does not take in account class labels during feature selection.

The performance of the newly developed DWT MCDF approaches was also compared to the performance of more conventional single classifier methods, namely SLDA in the spectral and DWT domains. In general, the multiclassifier approaches outperformed the single classifier approaches. It should be noted that the SLDA results for the spectral and DWT domains were generally equivalent. Thus, a projection of the data into the DWT domain for the single classifier approach did not improve results. Of the new methods, the most simplistic approach, CONCAT, performed quite poorly. Thus, simply combining DWT and MCDF without consideration of scalar grouping is not effective. Following CONCAT with LDA preprocessing did dramatically improve results. However, the results were still only on par with MCDF (without DWT preprocessing). SCALAR method was quite sensitive to the choice of mother wavelet, performing well for Daubechies-8 but poorly for Haar. SUBSPACE approach performed very well, resulting in approximately 80% accuracy for both types of mother wavelet. When comparing the metric-based approaches, both BD and ENTROPY perform very well, even when the methods are trained and tested on significantly different datasets

(training dataset was from a greenhouse study in 2005 and testing data was from a field campaign in 2008). One of the most interesting outcomes of the study was the high performance of the relatively simple ENTROPY method, which is unsupervised, and is more commonly used in DWT compression and denoising applications than in ATR applications.

3.6 REFERENCES

[1] S. Venkataraman, "Exploiting Remotely Sensed Hyperspectral Data Via Spectral Band Grouping For Dimensionality Reduction And Multiclassifiers," Master Thesis, Mississippi State University, 2005.

[2] A. Cheryadat, L. M. Bruce, "Decision Level Fusion with Best Bases for Hyperspectral Classification", *Proc. IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp 399-406, October 2003.

[3] S. Prasad, "Multi-classifiers and decision fusion for robust statistical pattern recognition with applications to hyperspectral classification," Doctoral Dissertation, Mississippi State University, 2008.

[4] L. M. Bruce, C.H. Koger, J. Li, "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction," *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331- 2338, 2002.

[5] P. Hsu, Y. Tseng, "Feature Extraction of Hyperspectral Data Using the Best Wavelet Packet Basis", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3, pp. 1667-1669 June 24-28, 2002.

[6] X. Zhang, N. H. Younan, C. G. O'Hara, "Wavelet domain statistical hyperspectral soil texture classification*," Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 43, no. 3, pp. 615 – 618, March 2005.

[7] T. West, L. M. Bruce, "Multiclassifiers and Decision Fusion in the Wavelet Domain for Exploitation of Hyperspectral Data," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, July 2007.

[8] P. Hsu, Y. Tseng, "Feature Extraction of Hyperspectral Data Using the Best Packet Basis", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, June 24-28-, 2002.

[9] L. M. Bruce, C. H. Koger, J. Li, "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction," *IEEE Trans. Geoscience and Remote Sensing*, vol. no. 10, pp. 2331-2338, 2002.

[10] S. Burrus, R. Gopinath, and H. Guo*, Introduction of Wavelets and Wavelet Transforms: A Primer*, Englewood Cliffs, NJ: Prentice-Hall, 1998.

[11] S. Jaffard, Y. Meyer, R. Ryan, *Wavelets Tools for Science & Technology*, Society for Industrial and Applied Mathematics, 2001.

[12] J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, Pearson Prentice Hall, 2005

[13] L. M. Bruce, C.H. Koger, J. Li, "Dimensionality Reduction of Hyperspectral Data Using Discrete Transform Feature Extraction," *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331 – 2338. 2002.

[14] T. West, S. Prasad, L. M. Bruce, "Wavelet Packet Tree Pruning Metrics for Hyperspectral Feature Extraction," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2, pp. 946 – 949, July 2008.

[15] S. Prasad, L. M. Bruce, "Hyperspectral Feature Space Partitioning via Mutual Information for Data Fusion." *IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, 2007.

[16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, California Academic Press Inc. 1990.

[17] R. W. Hamming, *Coding and Information Theory*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.

[18] T. M. Mitchell, *Machine Learning*, McGraw-Hill Inc., 1997.

[19] R. L. King, "A Comprehensive Threat Management Framework for a Crop Biosecurity National Architecture." Available online at http://www.lars.purdue.edu/seminar/.

[20] M. Livingston, R. Johansson, S. Daberkow, M. Roberts, M. Ash, and V. Breneman,"Economic and Policy Implications of Wind-Borne Entry of Asian Soybean Rust into the United States." Available online at http://usda.mannlib.cornell.edu/usda/ers/OCS/2000s/2004/OCS-04-27-2004_Special_Report.pdf.

[21] Analytical Spectral Devices Fieldspec Pro Spectroradiometer specifications. Available online at http://asdi.com/products_specifications-FSP.asp .

CHAPTER 4

UTILIZATION OF LOCAL AND GLOBAL HYPERSPECTRAL FEATURES VIA REDUCTION WAVELET PACKETS AND MULTICLASSIFIERS FOR ROBUST TARGET RECOGNITION

4.1 Introduction

The capabilities of hyperspectral sensors have proven attractive for applications requiring highly precise ground cover mapping. These sensors have the ability to produce hundreds to thousands of spectral bands per pixel. However, small amounts of labeled training data coupled with the large dimensionality of the spectral data often causes hyperspectral classification systems not to generalize well and thus perform poorly. Many dimensionality reduction and feature extraction techniques have been investigated to account for the "curse of dimensionality" in hyperspectral target recognition systems [1-4]. More recently, spectral band grouping coupled with multiclassifiers and decision fusion (MCDF) has been investigated to account for small amounts of label training data and to address the concerns of generalizability [1, 2, 5, 6].

Additionally, multiresolution analysis or wavelet analysis has become a basis for many feature extraction methods the last couple of decades in signal processing. Two of the most efficient methods for implementing multiresolution transformations are the discrete wavelet transform (DWT) and the redundant wavelet packet decomposition (WPD) via the dyadic filter tree [7]. The dyadic filter tree approach for both methods

involves the decomposing of a signal by projecting it onto scaled and translated versions of a prototype mother wavelet. The dyadic filter tree is implemented via a bank of low-pass and high-pass filters which produce the approximation and detail coefficients. In current research, the DWT and WPD have become leading methods in extracting local and global features in hyperspectral remotely sensed data. Hsu *et al.* used the WPD and DWT for feature extraction and optimization for hyperspectral target recognition in an agricultural application and found that the wavelet based features proved to have superior results than non-wavelet based features [8]. Bruce *et al*. investigated the use of the DWT in the dimensionality reduction of hyperspectral data and found that the local and global features were optimal in classification applications [4, 9]. Zhang *et al*. developed a remote sensing soil classification system employing the DWT as a feature extraction method, where the goal was the classification of three different pure soil textures [10].

In [3], the DWT was investigated as a preprocessing stage for a multiclassifier and decision fusion system for hyperspectral data. That is, the DWT was applied to the hyperspectral space and was implemented using the dyadic filter tree approach. Then each set of wavelet detail coefficients, along with the final set of approximation coefficients, were considered as potential feature vectors. From the potential feature vectors, a final set of feature vectors was selected based on classification-based performance metrics. Then each selected feature vector was sent to an individual classifier, and the classifications were fused to form a single output label for the hyperspectral signature. The DWT's contribution of local and global feature extraction was shown to improve classification accuracies as compared to the spectral-based MCDF approach.

In implementing the DWT, each level of the filter tree corresponds to a dyadic scale of the wavelet decomposition in which the high-pass filters (combined with 2-point decimation) produce the detail coefficients and the low-pass filters (combined with 2-point decimation) produce the approximation coefficients. To form the next level of detail coefficients (larger scale detail coefficients), the approximation coefficients are again subjected to, a 2-channel filter bank followed by 2-point decimation. However, resulting detail coefficients are never reanalyzed. In the corresponding WPD method, each set of detail coefficients is also decomposed into two parts using the same approach as in approximation coefficient splitting. This offers a richer analysis of the input signal, or hyperspectral signature in our case. The finer partitioning of the frequency space implies a better decorrelation of the signal than with the DWT. Thus, it is expected that that resulting feature vectors, and their input to a MCDF system, will result in improved classification potential.

In this study, a combination of the WPD and MCDF are investigated for a robust hyperspectral classification system. Specifically, a redundant WPD is used as the basis for multiresolution feature grouping and selection, forming groups of local and global spectral features, where each group is input to a classifier, resulting in local and global classifications. Then the decisions of the multiclassifiers are fused to form a final class label. This approach can be applied to a full WPD decomposition or to a "pruned" WPD tree [11]. In this work, a comparison of unsupervised and supervised cost functions for WPD tree pruning will be conducted. Also, the performance of the proposed WPD-MCDF method is compared with current state-of-the-art hyperspectral analysis techniques, such as stepwise-linear discriminant analysis (LDA) or discriminant analysis

feature extraction (DAFE) [12] and multiclassifiers and decision fusion (MCDF) in the original spectral domain [13,14]. The proposed and comparison methods are applied to hyperspectral data from an agricultural application, namely the detection or soybean rust infestations in soybean crops.

4.2 Wavelet Packet Decomposition (WPD)

The WPD is similar to the DWT in that both methods project the signal on to a scaled and translated version of the prototype mother wavelet and both can be implemented via a dyadic filter tree. The wavelet functions for the WPD are represented by

$$\psi_{jk}^{n}(x) = 2^{j/2}\psi^{n}\left(2^{j}x - k\right) n = 1,2, \ldots \quad (4.1)$$

where the wavelet packet function, ψ_{jk}^{n}, is defined by the parameters $n, j,$ and k in which n is the modulation, j is the scale, and k is translation parameter [12]. The WPD wavelets, ψ^{n}, are obtained by

$$\psi^{2n}(x) = \sqrt{2}\sum_{k=-\infty}^{\infty} h(k)\psi^{n}(2x - k) \quad (4.2)$$

$$\psi^{2n+1}(x) = \sqrt{2}\sum_{k=-\infty}^{\infty} g(k)\psi^{n}(2x - k) \quad (4.3)$$

The WPD wavelets are obtained recursively by (4.2) and (4.3). The quadrature mirror filters *h(k)* and *g(k)* are discrete filters which are related by the scaling and the mother wavelet functions [15]. In the WPD, the wavelet coefficients are obtained by a recursive high-pass and low-pass filtering (accompanied with a 2-point decimation) of both approximation and detail coefficients at each level. For example, the dyadic filter tree for a level 3 WPD is shown in Figure 4.1. Each level corresponds to a dyadic scale of the wavelet packet decomposition. The approximation and detail coefficients are

decomposed using a two-channel filter bank until either the desired decomposition level is achieved or the maximum allowable decomposition level is met, where maximum level is defined by the level at which the length of the filter impulse response is greater than or equal to the length of the filter's input signal. This maximum level is defined by the length of the original input signal and the mother wavelet utilized.

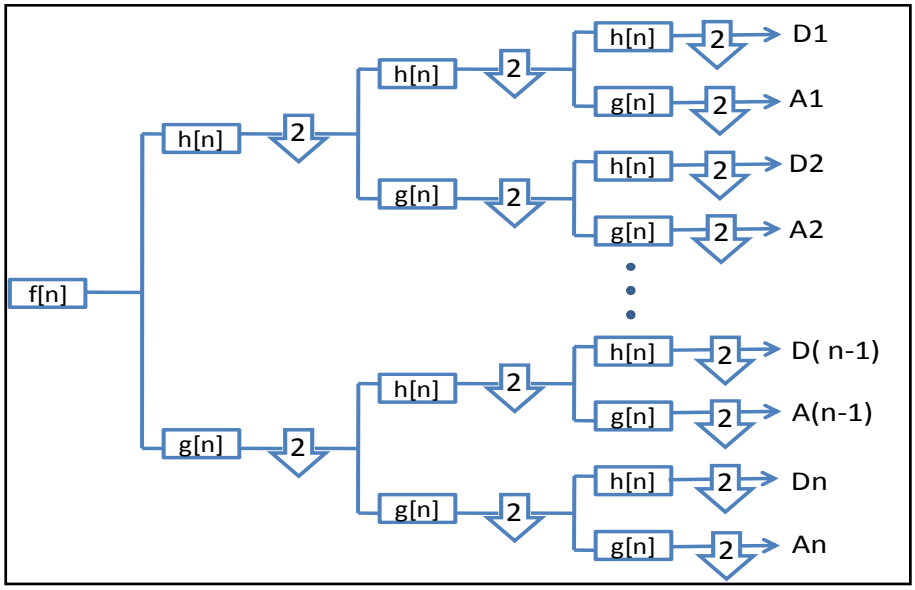


Figure 4.1 Block diagram of prototypical WPD dyadic filter tree, where f[n] is input signal, h[n] and g[n] are high-pass and low-pass filter impulse responses, respectively

4.3 Wavelet Packet Decomposition (WPD) in the frame work of Multiclassifiers and Decision Fusion

4.3.1 System Overview

A combination of the WPD and MCDF approach is proposed for a robust hyperspectral classification system. Figure 4.2 illustrates a block diagram of the

proposed system. A WPD is applied to the hyperspectral signature which produces the highest scale detail and approximation coefficients, also referred to as terminal leaves or nodes on the WPD tree. Each set of approximation and detail coefficients (leaves/nodes) are considered as a potential feature vector. The WPD tree may or may not be pruned. If it is left unpruned, all leaves are terminal nodes. If the tree is pruned, leaves may be non-terminal nodes, i.e. sets of approximation or detail coefficients from a lower scale. Regardless of whether WPD tree pruning is enacted, each WPD leaf is considered a feature vector. These feature vectors may be preprocessed for feature reduction/optimization and then passed to independent classifiers in a MCDF system [5, 6].

In this study, supervised and unsupervised methods of pruning are investigated. The preprocessing is a straightforward Fisher's LDA [13, 16]; the classifier is the commonly used maximum-likelihood classifier [16]; and the decision fusion is a simple majority vote [3, 12]. These components of the MCDF were intentionally chosen to be simple, well understood approaches, so that the focus of the study could be on the WPD-based feature grouping and selection.

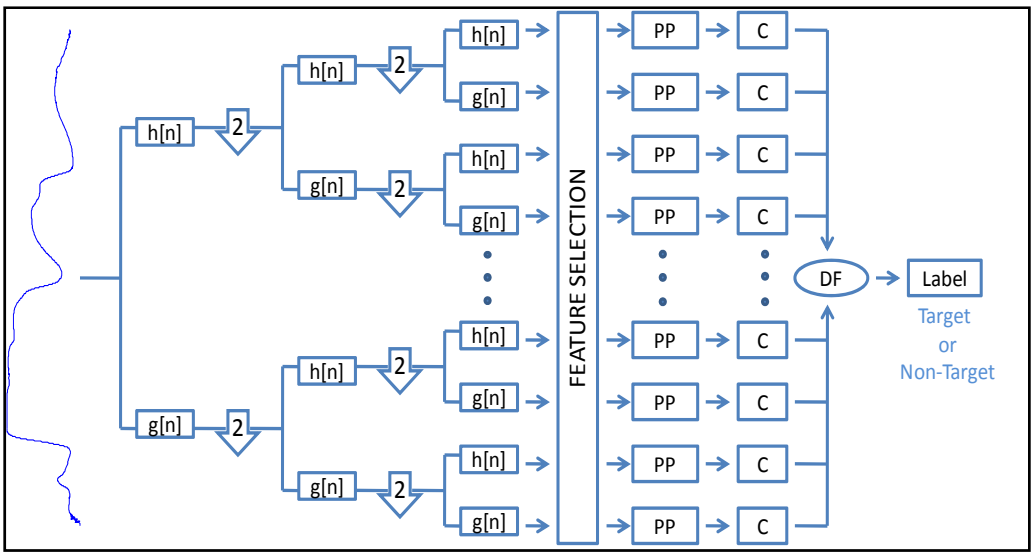


Figure 4.2 Block diagram representation of the proposed WPD MCDF framework, where PP is feature group preprocessing, C is a classifier, and DF is the decision fusion scheme

4.3.2 Wavelet Packet Decomposition

The selection of the mother wavelet has the potential to significantly impact the resulting hyperspectral features and their efficacy in discriminating ground cover classes. Thus, a mother wavelet sensitivity study was conducted. The mother wavelets utilized in this study are orthonormal, because of the dyadic filter tree requirements. The mother wavelets investigated in the study are the Daubechies family of wavelets which ranged from Daubechies 1 to Daubechies 10 (Including the Haar which is equivalent to Daubechies1).

The decomposition level can also play a major role in the overall performance of the target recognition system. The maximum allowed decomposition level directly affects the granularity of the frequency-space partitioning. Thus, a WPD decomposition level sensitivity study was performed, where the level is varied from 5 to 10. The lower

bound, 5, resulted from the fact that a decrease in decomposition level results in an increase in the number of coefficients in the terminal node. This relationship is governed by the length of the original input signal and the length of the WPD's low-pass/high-pass filter's impulse response, i.e. the choice of mother wavelet. In addition, to avoid the curse of dimensionality, the number of features in a given feature vector should be less than the number of labeled training data per class. In this case, the number of coefficients in a terminal node (feature vector) should be less than the number of ground-truthed hyperspectral signatures available for training the system. Considering the number of bands in our experimental hyperspectral data, the class of mother wavelets we investigated, and the number of class-labeled training signatures available in our dataset, the lower bound was set to 5. The upper bound is only affected by the length of the input hyperspectral signature (number of spectral bands) and the length of the WPD's low-pass/high-pass filter's impulse response, i.e. the choice of mother wavelet. Considering our experimental hyperspectral dataset and class of mother wavelets being investigated, the upper bound was set to 10.

4.3.3 Wavelet Packet Tree Pruning

In target recognition applications, the leaves/nodes on the WPD tree are pruned to form feature vectors, whereas in conventional WPD compression applications, the tree is pruned to form a basis for reconstruction. The leaves/nodes are pruned based on a cost function or a performance metric. The pruning of the decomposition tree usually occurs in a bottom-up approach, i.e. from the leaves (terminal nodes) to the root (original signal). Figures 4.3 and 4.4 give a visual example of a full WPD tree that has not been pruned and a WPD tree which has been pruned, respectively. For simplicity of these

figures, each line segment represents a high-pass or low-pass filter followed by 2-point decimation. For ground cover classification and target recognition applications in hyperspectral remote sensing, the cost function, or performance metric, should be an ATR-appropriate metric, such as class separation or target recognition accuracy. In WPD tree pruning, the selection of leaves/nodes for feature vectors must be optimum to ensure that the "curse of dimensionality" is not introduced. Therefore, the selection of cost functions or performance metrics must only extract the most optimum leaves/nodes.

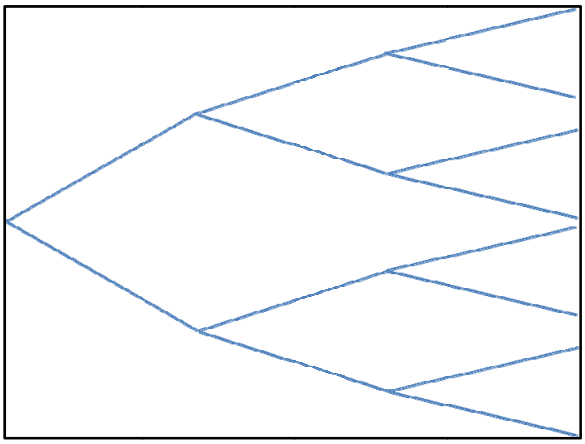


Figure 4.3 Full 3 level WPD tree, without pruning

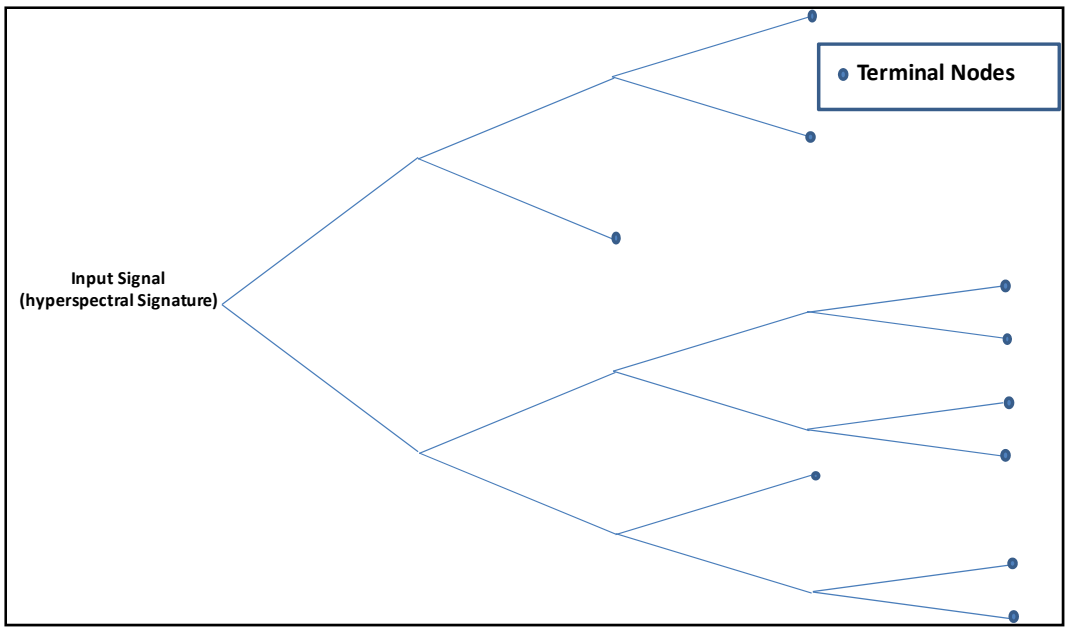


Figure 4.4 Example 4 level WPD tree that has been pruned

In this study, several metrics are investigated, supervised metrics based on target/non-target class separation and conventional unsupervised metric commonly used for WPD–based compression. The final feature set is selected based on the supervised and unsupervised performance metrics. In this work, the Bhattacharyya distance (BD) [16, 17] is the supervised metric employed, and entropy (ENT) [18] is the unsupervised metric employed in this study. Two types of pruning are investigated in this chapter. The first is a feature threshold selection and the second is a family tree pruning method.

4.3.3.1 Scalar Partition with Metric Based Pruning

In this work, metric-based pruning, whether they utilize supervised or unsupervised metrics, is based on that node's metric and its relation to the mean metric of all nodes. That is, let ***E*** be the collection of performance metric values for all nodes in the

full WPD, where $\boldsymbol{E}$ is defined by $[E_1 \cdots E_n]$ and n is the total number of nodes in the decomposition. Let $\overrightarrow{Fv}$ be a feature vector representing a set of wavelet coefficients (from one node) that is retained for input to the MCDF scheme. Then the following criteria must be met:

$$E[\overrightarrow{Fv}] > \mu_E + n\sigma_E \tag{4.4}$$

where, μ_E and σ_E are the mean and standard deviation of $\boldsymbol{E}$, respectively. The parameter n may be set to any integer value. The higher (or lower) the value of n, the more (or less) restrictive the pruning process, i.e. increasing (or decreasing) n decreases (or increases) the number of feature sets passed through to the MCDF scheme. This approach is referred to as "scalar partition pruning" or SPP.

With these pruning methods, there is a chance that the follow-on preprocessing stage of MCDF (namely LDA in our study) may not be appropriate, due to the fact that the dimensionality of the feature set is too low. LDA reduces the dimensionality of a feature vector by transforming the data on to a lower dimensional subspace that has a dimensionality of $\boldsymbol{C-1}$, where $\boldsymbol{C}$ is the number of classes. If the feature set's dimensionality is less $\boldsymbol{C}$, then LDA is not applied. If the feature set's dimensionality is greater than or equal to $\boldsymbol{C}$, then LDA is applied.

4.3.3.2 Metric Based Family Pruning

The metric based family pruning (FP) algorithm consists of the following steps:

1. Compute a full WPD of the training and testing hyperspectral signatures, i.e. a WPD to the maximum allowable level. Let j denote the j^{th} level in the WPD, i.e. j=1,2,…,J where j=J are the terminal nodes used to initialize the pruning. Let k denote the nodes

at a given level, i.e. k=1, 2, … 2^j Thus, $A^c_{j,k}$ denotes the set of WPD coefficients for the j^{th},k^{th} node for the class c.

2. Set the flag m=[1,2], where m=1 denotes the use of the performance metric BD, m=2 denotes the use of ENT.
3. Compute the appropriate performance metric for each node (j,k). For m=1, since BD is a supervised method, the metric is computed between $A^1_{j,k}$, $A^2_{j,k}$,.., $A^c_{j,k}$ for all j, k, and c. For m=2, since ENT is an unsupervised method, the metric is computed for each node not taking into account class labels. Thus, the WPD coefficients are combined across $A^1_{j,k}$, $A^2_{j,k}$,.., $A^c_{j,k}$ for a given j, k node. Denote the metric values for each node as $D_{j,k}$.
4. Sort the metric values in descending order and place in vector, $\vec{D}$, retaining node locations in an index vector $\vec{L}$.
5. Let i=1. Mark L_i as a selected feature vector.
6. Find all children and parents of L_i and remove these nodes from vector $\vec{L}$ and their corresponding metric values from $\vec{D}$.
7. Increment i, and repeat steps 5 and 6 until the entire list of metric values and associated nodes have been evaluated.

This pruning method removes the redundancy of the WPD. That is, if a node is selected for inclusion in the feature vectors, its children cannot be included. The result is a customized non-redundant dyadic decomposition. Figure 4.4 shows an example of a WPD tree that could result from the "family pruning" or FP method. The FP method could result in the classic dyadic discrete wavelet transform tree, i.e. all leaves corresponding to detail coefficients except for the final set of approximation coefficients. The FP approach has the potential benefit of non-redundancy in the resulting feature vectors, however, the method is more computationally expensive (in the training phase) than the scalar partition pruning approach.

4.4 Experimental Case Study

4.4.1 Data Collection

The proposed methods are applied to experimental hyperspectral data for an agricultural application, namely the early detection of a disease known as soybean rust (*Phakopsora pachyrhizi)* in soybean crops [19]. Soybean rust is a windborne pathogen which can be transmitted over large areas in a matter of weeks causing widespread damage [20]. In 2002/2003, Brazil suffered an estimated loss in soybean crop of 3.4 million tons and a $600 million estimated cost for fungicide sprays. The USDA estimates an economic loss of $640 million to $1.3 billion in the first year of a widespread soybean rust invasion in the United States [20]. The ability to rapidly detect soybean rust onset is critical to the US economy, and agencies such as the U.S. Department of Agriculture (USDA) and U.S. Department of Homeland Security (DHS) are particularly interested in this very challenging remote sensing problem.

The hyperspectral data was collected using the Analytical Spectral Device (ASD^{TM}) Fieldspec Pro handheld spectroradiometer [21]. The ASD has a spectral range of 350 – 2500 nm, spectral resolution of 1-1.2 nm, and uses a single 512 element silicon photodiode array for sampling 350 - 1000 nm and two separate, graded index Indium-Gallium-Arsenide photodiodes for the 1000 - 2500 nm range.

For this study, two datasets of ASD readings of soybean, both control and diseased, were used. The first dataset was collected over a two week period in a green house outside the city of Encarnacion, Paraguay, in 2005 with the humidity at 100% and the temperature kept close to 80 - 85 F [20]. For this study, 678 hyperspectral signatures

were used for evaluation, 320 observations of the control soybean and 358 observations of the inoculated soybean. The second dataset was collected in January of 2008 in Stoneville, MS, during an actual outbreak of the disease in a commercial crop setting. Eight-five observations were collected, 19 observations of the non-diseased soybean and 66 of the diseased soybean. Figure 4.5 shows photos of soybean plants with soybean run infestations of varying intensities. Figure 4.7 shows example signatures of 4 classes of soybean rust infestation plus the control. As one can see, the signatures in Figure 4.6 have considerable overlap and present a difficult detection case.

The parameters of the system (metric, mother wavelet, the level of decomposition, and pruning approach) were optimized by employing the 2005 data. Then the 2008 data was used to test the system. Thus the robustness of the system, i.e. the system's sensitivity to training data is tested.

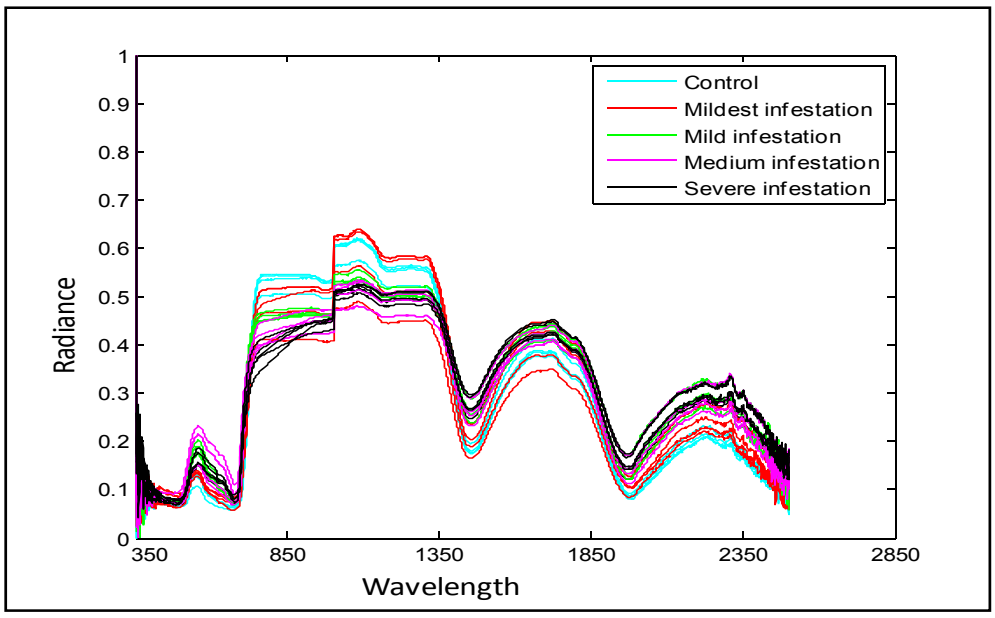


Figure 4.5 Example hyperspectral signatures for soybean vegetation, collected in January 2008 field campaign

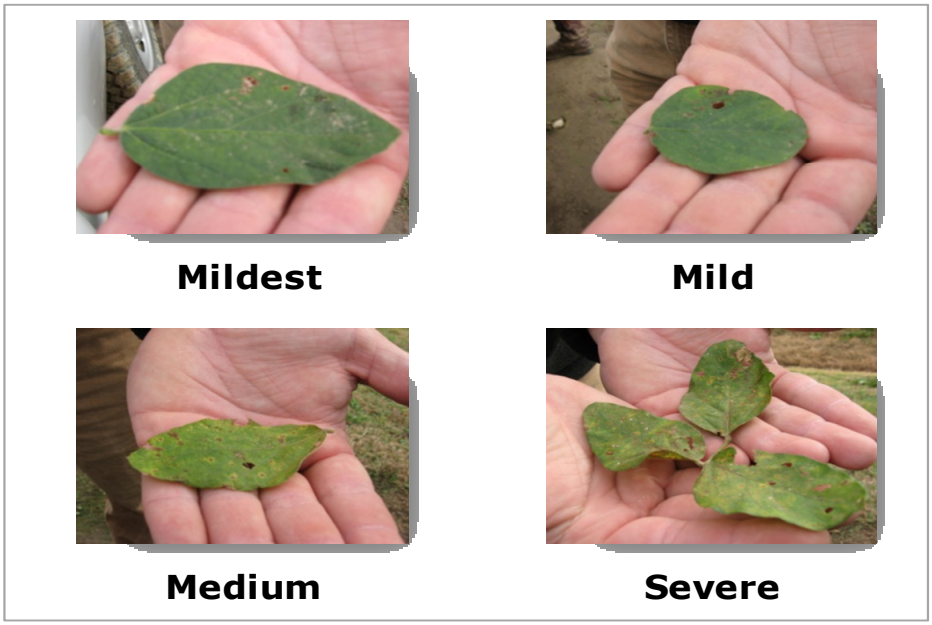


Figure 4.6 Photos of leaves collected from soybean vegetation at varying stages of soybean rust infestation, data collected in January 2008 field campaign

Table 4.1 Number of samples for each soybean rust infestation level collected during the January 2008 collection campaign

Level of Infestation	*Control*	*Mildest*	*Mild*	*Medium*	*Severe*
Number of samples	19	8	19	19	20

4.4.2 Experimental Results

Figure 4.7 shows the results of the mother wavelet selection sensitivity analysis, utilizing the 2005 hyperspectral dataset. The results are reported in terms of overall classification accuracy and 95% confidence intervals vs. Daubechies family mother wavelet. The decomposition level in each case was the maximum allowed based on the length of the input hyperspectral signature and the mother wavelet's corresponding low-pass/high-pass filter impulse response length. From these results, it is clear that the sensitivity to mother wavelet selection is quite high. The method "ENT" exhibits the least sensitivity to mother wavelet selection. Based upon overall classification accuracies, one could select Haar of DB10 as an effective mother wavelet, regardless of the pruning approach and metric. However, the Haar mother wavelet allows for simple implementations and fast computations, thus the Haar mother wavelet was selected for use in the follow-on experiments.

Figure 4.8 shows the results of the WPD decomposition level sensitivity analysis, utilizing the 2005 hyperspectral dataset. The results are reported in terms of overall classification accuracy and 95% confidence intervals vs. decomposition level. These

results are for the Haar mother wavelet. The SPP method with the BD metric exhibits significantly less sensitivity than the other methods. The FP method with the entropy metric exhibits the most sensitivity of the investigated methods.

Figure 4.9 shows the results of a comparison analysis of the proposed WPD-MCDF methods to conventional spectral-based approaches. The mother wavelet is Haar, and the WPD decomposition level is 5. Results are reported in terms of overall classification accuracy and 95% confidence intervals. The conventional approaches used for comparison purposes include SLDA of the original spectra and MCDF in the original spectral domain, i.e. without a wavelet decomposition preprocessing stage. The analysis was conducted utilizing the 2008 hyperspectral dataset. That is, the ATR systems were trained on 2005 hyperspectral data and tested on 2008 hyperspectral data. SLDA resulted in an overall accuracy of around 40%, demonstrating the level of difficulty of this particular application. The non-wavelet-based MCDF and the WPD-MCDF with FP and the entropy metric performed on par with one another, resulting in overall accuracies around 60-70%. The FP with BD metric and the SPP with either entropy or BD metric all performed quite well, resulting in overall accuracies ranging from 70% to around 80%.

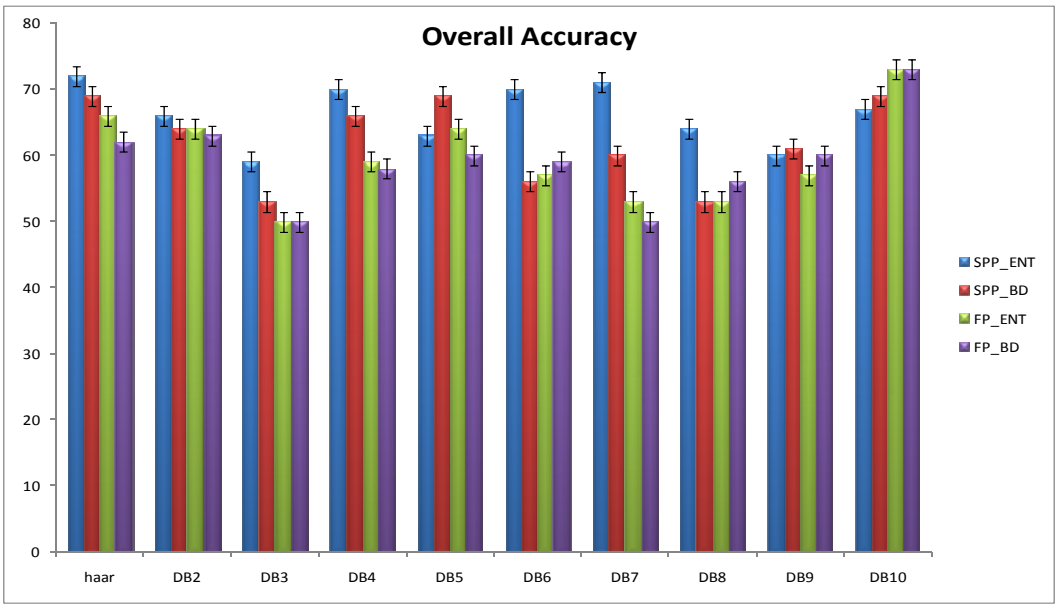


Figure 4.7 Results of mother wavelet selection sensitivity analysis, utilizing 2005 hyperspectral dataset. Results are reported in terms of overall classification accuracy and 95% confidence intervals vs. Daubechies family mother wavelet.

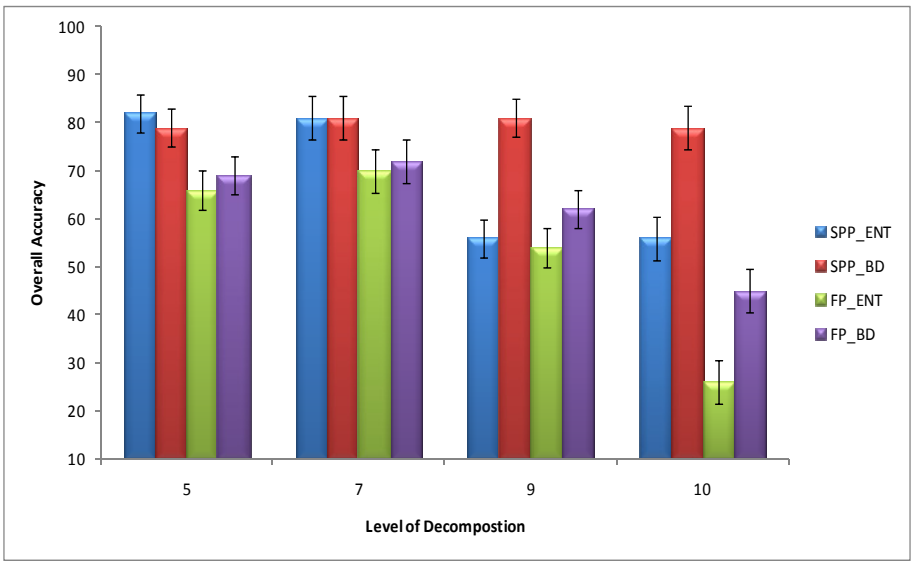


Figure 4.8 Results of WPD decomposition level sensitivity analysis, utilizing 2005 hyperspectral dataset and the Haar mother wavelet. Results are reported in terms of overall classification accuracy and 95% confidence intervals vs. decomposition

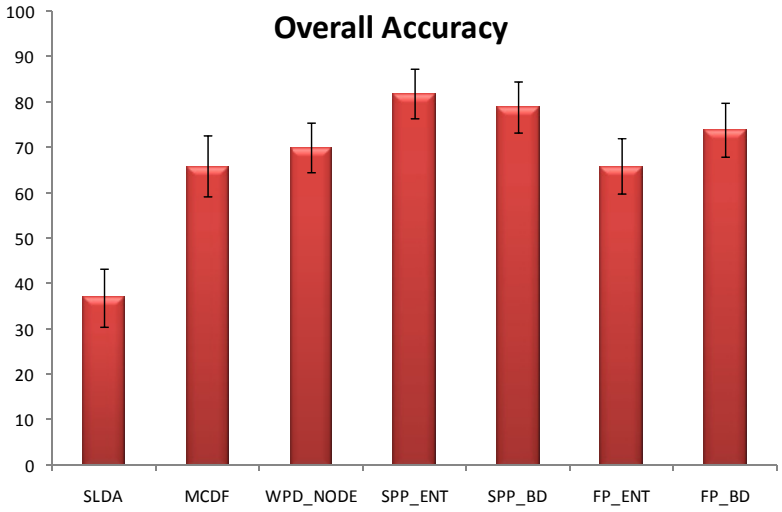


Figure 4.9 Comparison analysis of proposed WPD-MCDF methods to conventional spectral-based approaches. Note that WPD methods are designed using 2005 data, and all testing is conducted with 2008 dataset. Mother wavelet is Haar and WPD decomposition level is 5. Results reported in terms of overall classification accuracy and 95% confidence intervals.

4.5 Conclusions

For the given agricultural application, the ground cover classes were only subtly different and the amount of ground truth labeled data for training an ATR system was extremely limited. In some instances the number of spectral bands outnumbered the number of labeled pixels for training by a factor of 200-to-1. Thus, there is a critical need for dimensionality reduction methods which have capabilities of extracting pertinent class discriminatory information. Currently used methods like SLDA result in low multiclass classification accuracies. Incorporating the use of MCDF approaches in the spectral domain significantly improve classification performance but still only achieve accuracies of around 60-70%. The authors designed, implemented, and tested a new approach where wavelet packets are combined with MCDF. These approaches increased accuracies to greater than 80%.

The authors investigated the WPD MCDF method's sensitivity to mother wavelet selection and decomposition level. The authors also designed two WPD tree pruning methods to increase computational efficiency, and possibly improve classification accuracies simultaneously. The pruning approaches resulted in a set of WPD nodes/leaves, each containing a set of approximation or detail coefficients that were then used as a feature vector input to the MCDF scheme. Both pruning approaches were implemented using unsupervised performance metrics, namely entropy, and supervised metrics, such as BD. One pruning approach used a straightforward thresholding of metrics from all nodes/leaves to determine which nodes/leaves are selected as feature vectors. The second pruning approach used an intelligent approach that ensured a non-

redundant dyadic decomposition where the nodes with the highest performance metric were terminal nodes. Then all terminal nodes were selected as feature vectors.

The experimental results showed the WPD MCDF approaches to be significantly superior, in terms of overall accuracies, to the conventional SLDA approach. It was surprising that the experimental results showed the highest classification accuracies stemmed from the use of the simpler and less computationally expensive thresholding approach for pruning and unsupervised metric.

4.6 REFERENCES

[1] S. Venkataraman, L. M. Bruce, "Hyperspectral Dimensionality Reduction via Localized Discriminant Bases," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Seoul, Korea, vol. 2, pp. 1245-1248, July 25-29, 2005.

[2] A. Cheryadat, L. M. Bruce, "Decision Level Fusion with Best Bases for Hyperspectral Classification", *Proc. IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp. 399 – 406, October 2003.

[3] T. West, L. M. Bruce, "Multiclassifiers and Decision Fusion in the Wavelet Domain for Exploitation of Hyperspectral Data", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, pp. 4850-4853, July 2007.

[4] 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. 29, no. 10, pp. 2217-2226, 2001.

[5] S. Prasad. L. M. Bruce, "Hyperspectral Feature Space Partitioning via Mutual Information for Data Fusion," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, pp. 4846-4829, Spain, 2007.

[6] S. Prasad, L. M. Bruce, "Decision Fusion with Confidence-Based Weight Assignment for Hyperspectral Target Recognition," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1448 – 1456, May 2008.

[7] S. Burrus, R. Gopinath, and H. Guo*, Introduction of Wavelets and Wavelet Transforms: A Primer*, Englewood Cliffs, NJ: Prentice-Hall, 1998.

[8] P. Hsu, Y. Tseng, "Feature Extraction of Hyperspectral Data Using the Best Wavelet Packet Basis", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3, pp. 1667 – 1669, June 24-28, 2002.

[9] L. M. Bruce, C.H. Koger, J. Li, "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction," *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331- 2338, 2002.

[10] X. Zhang, V. Vijayaraj, N. H. Younan, "Hyperspectral Soil Texture Classification", *IEEE workshop on Advances in Techniques for Analysis of Remote Sensing Data*, pp. 182-186, 2003.

[11] S. Prasad, L.M. Bruce, "Limitations of Subspace LDA in Hyperspectral Target Recognition Applications," *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, pp. 625-629, July 2007.

[12] T. West, L. M. Bruce, S. Parsad,"Wavelet Packet Tree Pruning Metrics for Hyperspectral Feature Extraction", *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Boston, Massachusetts, pp. 946 – 949, July 2008.

[13] S. Venkataraman, "Exploiting Remotely Sensed Hyperspectral Data Via Spectral Band Grouping For Dimensionality Reduction And Multiclassifiers," Master Thesis, Mississippi State University, 2005.

[14] S. Prasad, "Multi-classifiers and decision fusion for robust statistical pattern recognition with applications to hyperspectral classification," Doctoral Dissertation, Mississippi State University, 2008.

[15] S. Jaffard, Y. Meyer, R. Ryan, Wavelets Tools for Science & Technology, Society for Industrial and Applied Mathematics, 2001.

[16] J. R. Jensen, Introductory Digital Image Processing: A Remote Sensing Perspective, Pearson Prentice Hall, 2005

[17] K. Fukunaga, Introduction to Statistical Pattern Recognition, California Academic Press Inc. 1990.

[18] R. W. Hamming, Coding and Information Theory, Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.

[19] King , R. L. "A Comprehensive Threat Management Framework for a Crop Biosecurity National Architecture," available online at *http://www.lars.purdue.edu/seminar/*

[20] M. Livingston, R. Johansson, S. Daberkow, M. Roberts, M. Ash, and V. Breneman,"Economic and Policy Implications of Wind-Borne Entry of Asian Soybean Rust into the United States," available online at http://usda.mannlib.cornell.edu/usda/ers/OCS/2000s/2004/OCS-04-27-2004_Special_Report.pdf

[21] Analytical Spectral Devices Fieldspec Pro Spectroradiometer specifications. Available http://asdi.com/products_specifications-FSP.asp

CHAPTER 5

RAPID DETECTION OF AGRICULTURAL FOOD CROP CONTAMINATION VIA HYPERSPECTRAL REMOTE SENSING

5.1 Introduction

Passive optical remote sensing techniques, including hyperspectral imaging, have been used in many different applications in agriculture, from detecting weeds to characterizing crop stresses to estimating crop yields. Many factors have been shown to affect the optical reflectance properties of crops, including water content, diseases, and soil nutrients. For example, MacNeil et al. used diffuse reflectance spectroscopy to differentiate between injury caused by the white apple leafhopper (*Typhlocyba pomaria*) and nitrogen deficiency on apple (*Malus sylvestris*) leaves [1]. Adcock et al. found that paraquat injury on soybeans (*Glycine max*) was detected using a radiometer at 800nm [2]. Mortimer et al. found that spectroradiometer readings correctly classified sublethal doses of glyphosate on non-transgenic cotton when using a linear discriminatory analysis, even when injury was not detected visually [3].

Thus, multispectral and hyperspectral imagers are powerful tools in remote sensing and provide great promise for rapid detection and characterization of agricultural food crop contaminants. Hyperspectral imagers have the potential to be useful in detecting when a contaminant has been introduced to an agricultural crop before the crop stresses are visible to the human eye, providing a valuable lead time in first response. In

some cases there is no visible indicator that the contaminant has been introduced to the vegetation; i.e. the optical reflectance is altered only in the non-visible regions of the optical spectrum. A hyperspectral image can provide densely sampled reflectance values across the visible and near infrared regions of the spectrum, resulting in hyperspectral signatures with 100's to 1000's of spectral bands. These signatures can then be analyzed with advanced mathematical algorithms, via automated target recognition (ATR) system, to determine if a particular target is present. In this application, the "target" would be a contaminated agricultural crop and the "nontarget" would be an agricultural crop under normal conditions. And even more challenging, the ATR system could be used to characterize the level of contamination, via a multiclass classification approach.

Subtle changes in vegetation, as a result of low levels of contamination, can prove quite difficult to recognize and thus require the use of more sophisticated spectral features, necessitating the use of hyperspectral sensors and advanced ATR schemes. However, the high dimensionality of hyperspectral data typically requires one to have a large number of training samples for designing and training the ATR system's algorithms. A common problem in many real-world applications is the lack of sufficient training data. The increase in spectral features coupled with the lack of available training data introduces the "curse of dimensionality". The need for larger amounts of training data stems from the fact that the number of training samples required is directly related to the dimensionality of the classifier [4]. In order to avoid this problem, the hyperspectral datasets must be preprocessed, thereby reducing the dimensionality to an acceptable level. Such preprocessing methods must reduce the dimensionality of the hyperspectral dataset while maintaining the pertinent information required for accurate classifications.

In previous work [5-10] and in the previous chapters, a variety of new methods have been explored for dimensionality reduction and classification of hyperspectral data, including spectral band grouping, wavelet coefficient feature extraction and selection, and multiclassifiers and decision fusion (MCDF) techniques. In these works, it was found that the combination of discrete wavelet transforms (DWT) or the wavelet packet decomposition (WPD) with MCDF schemes are quite powerful in exploiting hyperspectral data for classifying subtly different vegetative classes.

In this work, the WPD MCDF framework is tested on a practical classification task of detecting and characterizing chemical contaminations of corn and biological pathogens in soybean crops. The WPD framework is applied to both handheld spectroradiometer data and airborne hyperspectral imagery and is compared to ATR methods currently commonly used in the remote sensing community, including those based on principal component analysis (PCA), multiclassifiers and decision fusion (MCDF) in the spectral domain, discriminant analysis feature extraction (DAFE) which is also known as step-wise Fisher's linear discriminant analysis (SLDA), and single maximum-likelihood classifiers [4]. The results from this work will demonstrate that the WPD MCDF framework can be effectively applied to airborne hyperspectral imagery for accurate detection and classification of crop contaminations, even when the amount of training data is very limited.

5.2 Need for Accurate Detection and Characterization of Crop Contaminations

Chemical contamination of the agricultural food supply could cause irreparable economic damage to the U.S., where one in eight jobs depends on food production. The economic losses would be particularly damaging to states whose economies are primarily

based on agriculture, such as Mississippi. Approximately 37 % of Mississippi's 30 million acres is designated as farmland. Of that, 11 million acres of farmland, approximately 4.1 million acres are designated as harvested cropland. Thus, approximately 14% of Mississippi's total land area is designated as harvested cropland. And approximately 20% of all Mississippi jobs are farm or farm-related jobs [11]. A widespread chemical or biological contamination on the state of Mississippi would obviously cause significant economic damage to Mississippi's economy. Table 5.1 below lists Mississippi's top five agricultural exports for FY2005 and lists each export's rank amongst the nation's states [11]. From these statistics, it is clear that a disruption in Mississippi's agricultural production would significantly impact the nation's access to agricultural commodities and the nation's food supply. Whether the contamination is a deliberate matter, act of terrorism, or is spread by a natural disaster, the chemical or biological contamination could have an impact that could not only affect the current crops but also could have long residuals that would affect crops in subsequent years.

In the case of a deliberate act of chemical or biological contamination, general use herbicides and pesticides could be used as the chemical contamination agent. For example, glyphosate, the active ingredient in Roundup, has a very high LD_{50}, is classified as a general use pesticide, and is highly injurious to most crops that are not genetically modified to withstand the herbicide. glyphosate and pyrithiobac applied at $1/64^{th}$ of the use rate on 6-leaf corn have resulted in significant yield reductions [12, 13].

In the case of an unintentional act of biological contamination, such as a natural disaster, a prime example would be a widespread infestation of soybean rust (SBR) across key agricultural regions in the US due to unusually widespread wet/humid weather

patterns, which could be caused by an active hurricane season. Two fungal species, *Phakopsora pachyrhizi* (also known as the Asian species) and *P. meibomiae*, cause SBR and are spread primarily by windborne spores that can be transported over long distances. Figure 5.1 illustrates regions of the country known to have climatic conditions to support an unintentional, natural SBR outbreak along with USDA reported soybean production.

In the first year, of natural SBR infestation, assuming that U.S. producers were able to treat with fungicides upon SBR detection, the expected value of losses across all U.S. agricultural producers and consumers would range from $640 million to $ 1.3 billion, depending on the severity of infestation [14]. There exists a strong need for a means to rapidly and accurately detect such an event. Thus, the contaminated crops could be treated more quickly and effectively, reducing the spread of the infestation and minimizing losses.

Table 5.1 Mississippi's Top 5 Agriculture Exports, estimates, FY 2005

	Rank among states	*Value (millions $)*
1. Cotton and linters	4	336.2
2. Poultry and products	5	217.4
3. Soybeans and products	14	168.6
4. Rice	4	92.9
5. Feed grains and products	25	92.9
	Total Value	917.8

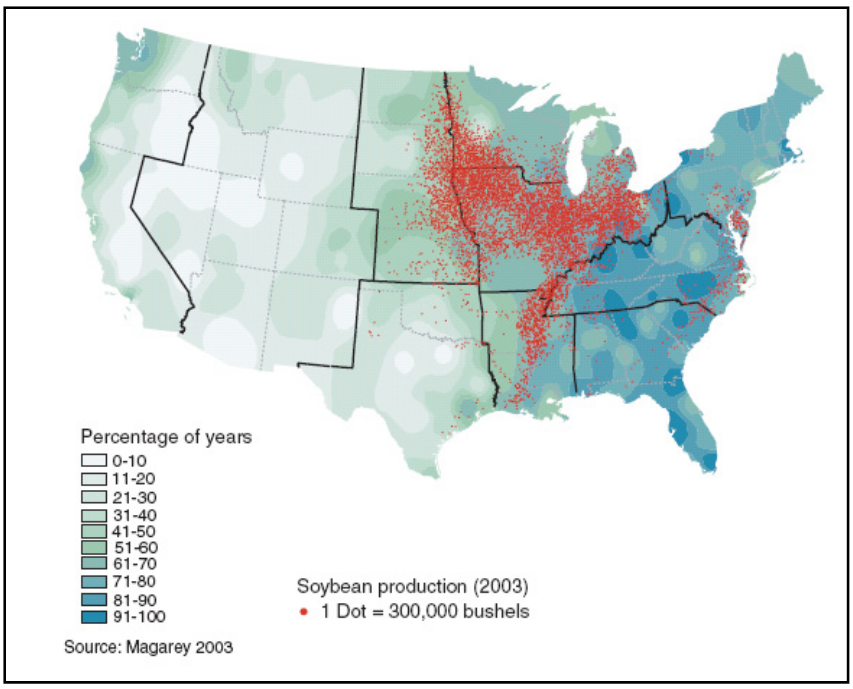


Figure 5.1 Percentage of years that climatic conditions exist that support Soybean Rust outbreaks. [14]

5.3 Experimental Case Study and Field Campaigns

5.3.1 Data Collection Methods

The authors conducted an extensive 2-year field campaign, consisting of field-level experiments of corn to highly controlled, varying levels of chemical contaminations. Both handheld and airborne hyperspectral data were collected multiple times throughout the two growing seasons. The experiments were designed to mimic a agricultural herbicide drift event. In addition, handheld spectroradiometer dataset was collected for soybean under normal conditions and under varying levels of SBR infestation.

The handheld hyperspectral data was collected using the Analytical Spectral Devices (ASD™) Fieldspec Pro handheld spectroradiometer [15] and the SpecTIR™ airborne hyperspectral imager [16]. The ASD data has 2151 spectral bands from a spectral range of 350 – 2500 nm, spectral resolution of 3 nm @ 700 nm and 10 nm @ 1400/2100 nm [15]. The airborne SpecTIR sensor has 128 bands, which range from 400 nm to 994 nm, with a spectral resolution of 10nm and a spatial resolution of 1m [16]. A 25° instantaneous field of view (IFOV) foreoptic was used, and the sensor was held nadir at approximately 2 feet above the vegetation canopy. Reflectance values in the regions 1350nm - 1430nm and 1800nm – 1980nm were interpolated using a cubic spline method to remove the atmospheric water absorption effects. Figure 5.2 displays an aerial view of one of the data collection sites, showing a false-color display of the airborne hyperspectral imagery.

5.3.2 Chemical Contamination of Corn

The hyperspectral dataset collected in this study was acquired at the Plant Science Research Center and the Black Belt Branch Experiment Station in Brooksville, Mississippi over a two year time period. The corn was planted in 96.5 cm rows in 3.86 m by 12.2 m plots at a seeding rate of 108,000 seed/ha. The fields were sprayed with Glufosinate herbicide which was diluted with water to form 8 different concentrations. All treatments were applied at the 6 – to – 8 leaf growth stage with a tractor-mounted compressed air sprayer. The corn had 8 concentrations of herbicide, and the solutions were 2, 1, 0.5, 0.25, 0.125, 0.0625, 0.032125, and 0 (control) where the value corresponds to the fraction of the label-recommended dose (r-g ae/ha), e.g. class “0.25” corresponds to a spray rate of one-fourth the label recommended rate. In this study, each

level of concentration is considered as a specific ground-cover class, which makes this a difficult target classification study, since there are multiple classes that are very similar. To ensure an unbiased data set, the concentrations were sprayed in a randomized spray pattern across the field which is shown in Figures 5.2 and 5.3.

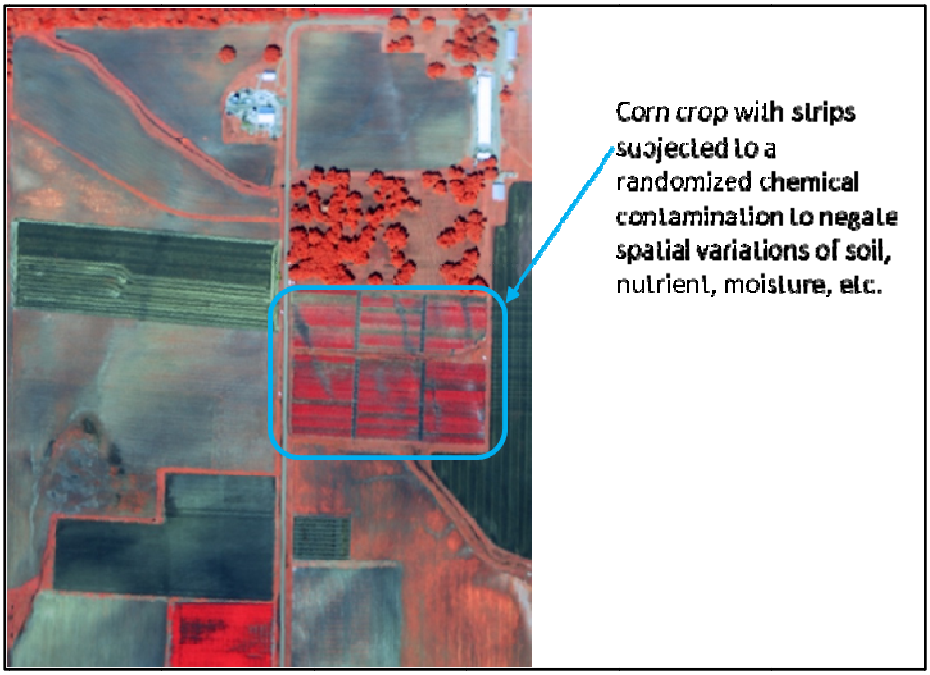


Figure 5.2 Aerial view of data collection site for corn crop subjected to varying levels of chemical contamination - Plant Science Research Center and the Black Belt Branch Experiment Station in Brooksville, Mississippi

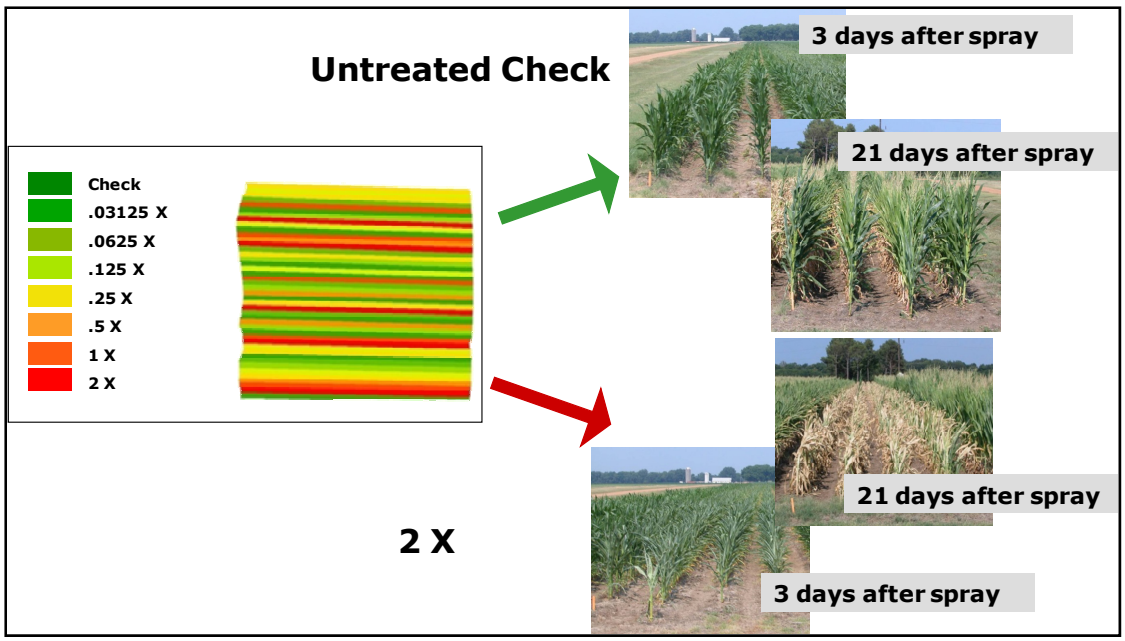


Figure 5.3 Diagram of randomized spray pattern for application of chemicals with varying concentrations

The 2008 corn dataset in this experiment was collected over a 14 day period for each crop, with data collections on 1, 4, 8, and 14 days after herbicide application. The 2009 corn dataset in this experiment was collected over a 22 day period for each crop, with data collections on 1, 13, 14, and 22 days after application of the herbicides. The airborne imagery was collected on June 6, 2008, 6 days after application of the herbicides.

Ground truth for the handheld collected data and the airborne imagery was recorded with the use of a mobile Global Positioning System (GPS) unit known as the Real Time Kinematic (RTK) system. The experimental setup for collecting the handheld hyperspectral data was a systematic method where the ASD and a GPS unit were used simultaneously. Both the ASD and GPS units were attached to a platform which was

transported by a tractor, as shown in Figure 5.4. The ASD instrument was set to collect an average of 10 hyperspectral signatures each second as the tractor moved across the field at 3 miles per hour. Thus, each hyperspectral signature represented an average of approximately 1 meter along-track and approximately 0.5 meter across-track. At every hyperspectral data collection point a corresponding GPS point was acquired for validation. Figure 5.5 illustrate the GPS locations (blue dots) where dataset was systematically collected across the field; note that only 1/10 of the actual locations are displayed in this figure in order to facilitate visualization of the data. This approach of semi-automated approach to data collection with the handheld spectroradiometer resulted in relatively large quantities of ground-truthed hyperspectral signatures. For example, for a given date, this data collection approach resulted in approximately 5000 samples collected for the 8-class problem.

Ground truth for the airborne imagery was obtained using a mobile GPS unit to measure the outlines of the randomized herbicide spray maps, producing shape files that could be overlaid on the imagery. Using this method, the authors were able to obtain an approximate 5000 ground truthed pixels for a given date.

Figure 5.4 Experimental setup for semi-automated handheld spectroradiometer data collection; photo shows white referencing of ASD unit

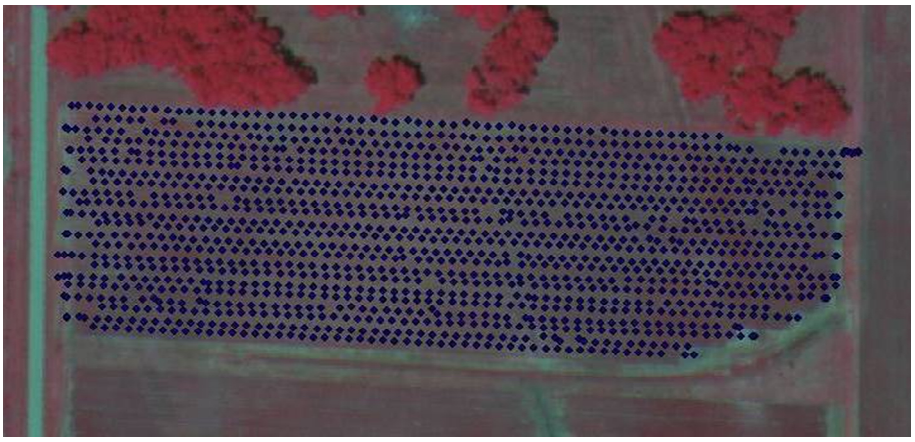


Figure 5.5 GPS locations (blue dots) where data was systematically collected across the experimental test sight; note that only 1/10 of the actual locations are displayed in this figure in order to facilitate visualization of the data

5.3.3 Soybean Rust Biological Infestation

The data in this experiment were collected in the Mississippi Delta region in January of 2008. The signatures were collected with a handheld spectroradiometer (ASD). The data set consists of 5 classes ranging from control to severe infestation level of soybean rust. Classes were “Control”, “Very Mild”, “Mild”, Moderate”, and “Severe”. During this data collection, the authors were accompanied by plant science experts from the Mississippi Bureau of Plant Industry [17] to determine the severity of soybean rust infestation for each observation. Figure 5.6 shows photographs of soybean plants in this study.

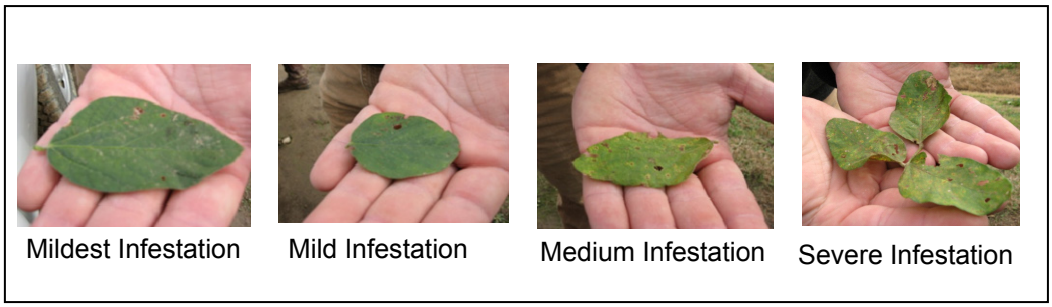


Figure 5.6 Examples of soybean plants, including control and diseased. Photos correspond to handheld hyperspectral data collected in January 2008 in Stoneville, MS

5.4 Hyperspectral Analysis Methods

In many hyperspectral classification applications, individual spectral bands are extracted as features for the identification of a target. When using statistical pattern recognition techniques, the large dimensionality of the feature space induces a requirement of a large amount of labeled training data, if the class distributions are to be accurately described. The increase in spectral features along with small amounts of

labeled training data naturally causes hyperspectral ATR systems to suffer the "curse of dimensionality", resulting in lower classification accuracies [4]. To account for the lack of labeled training data, i.e. ground truthed pixels, hyperspectral ATR systems typically reduce the high dimensional data via dimensionality reduction or feature extraction techniques. The newly developed WPD MCDF system is applied to both the corn and soybean hyperspectral data sets. In addition, for comparison purposes, various other commonly used and/or state-of-the-art methods will also be tested, such as PCA, LDA, SLDA, and MCDF.

PCA is a commonly used method for dimensionality reduction in hyperspectral data analysis. PCA can be found in many commercial software packages for remote sensing such as ENVI™ and IMAGINE™. PCA seeks to find a linear transformation which projects the data onto a subspace in which the features are mutually uncorrelated and the total variance of the data is maximized. The linear transformation involves applying eigen-analysis to the covariance matrix of the entire unlabeled data set [4].

Fisher's LDA seeks to maximize the class separation between data by reducing the dimensionality through linear projections of the data onto a lower subspace. This separation is achieved by maximizing the between-class covariance matrix ($\boldsymbol{S_B}$) and minimizing the within-class covariance matrix ($\boldsymbol{S_W}$) [4]. Thus, LDA is a supervised method. Simply stated, the objective of LDA is to find a linear transformation matrix $\boldsymbol{W}$ such that $\vec{y} = \boldsymbol{W}^T\vec{x}$, where $\vec{x} \in \mathcal{R}^d$ (original data), $\vec{y} \in \mathcal{R}^m$ (projected data), $m = c - 1$, (c is the number of classes), such that the following criterion is maximized: $J(\boldsymbol{W})=.(\boldsymbol{W}^T\boldsymbol{S_B}\boldsymbol{W})/(\boldsymbol{W}^T\boldsymbol{S_W}\boldsymbol{W})$.

Stepwise LDA is an iterative implementation of LDA. The inputs to LDA, typically features, are sorted in descending order of class separation efficacy, using a performance metric, like class separation, e.g. Bhattacharyya Distance (BD). Next, a forward selection process is conducted to form (grow) a subset of features. This portion of the method is a bottom-up approach, where the top performing feature seeds the subset. Features are added to the subset only if the performance metric of their LDA result is increasing. Next, a backward rejection process is conducted to form (shrink) a subset of features. This portion of the method is top-down approach, where the final subset of the forward selection seeds the subset. Features are removed from the subset only if the performance metric of the LDA of the reduced set is increasing. After the removal of all features in the subset has been considered, the result is finalized. LDA is applied to the final subset.

Spectral band grouping, combined with multiclassifiers and decision fusion (MCDF), has been shown recently to be a very promising solution [5, 6]. With this approach, the adjacent spectral bands are grouped in order to form lower dimensional subspaces. The grouping can be as simple as a non-overlapping sliding window of fixed size or more sophisticated methods like those that maximize a performance metric such as the product of average mutual information and BD [5, 6, and 18]. Then the spectral band groups are sent to a bank of classifiers, one classifier for each group. Next, the outputs of the classifiers are fused using decision fusion to produce one final classification, e.g. target or non-target. The weights used in the decision fusion stage of the system typically take into account the reliability of each group/classifier combination to accurately classify a pixel. In this comparison method, the adjacent spectral bands are

grouped by using a fixed window approach and after each group is formed the dimensionality of each group is reduced using the LDA, as this simple approach has been demonstrated to often perform on par with more sophisticated band grouping methods [18]. The maximum likelihood classifier is used in all comparison methods in this study and thus is used in the multiclassifier bank of the MCDF approach. The decision fusion scheme is a majority vote method; again, this simple approach has been shown to often perform on par with more sophisticated decision fusion schemes [18]. For this study, the mother wavelet is Haar, and the WPD decomposition level is 5, where both parameters were selected via sensitivity studies described in Chapter 4.

To aid in the extraction and selection of pertinent hyperspectral features, a WPD and MCDF approach is used for a robust hyperspectral classification system. A brief description of the WPD MCDF approach is given here; extensive details can be found in chapter 4. In previous chapters of this dissertation, the newly proposed system applies the WPD to the hyperspectral data set which produces a set of leaves/nodes in the WPD decomposition "tree". The WPD is applied by the implementation of the dyadic filter tree, which consist of a bank of high-pass and low-pass filters, resulting in wavelet detail and approximation coefficients respectively. Each set of approximation and detail coefficients (residing at leaves/nodes) are considered as a potential feature vector. However, this results in a large number of potential feature vectors, i.e. $2^N \cdot 2^{N-1} \cdot \cdots \cdot 2^1 + 2^{N-1} + \cdots + 2^1$ for an *N*-level decomposition. Thus, the WPD filter tree is pruned to reduce the number of leaves/nodes used as feature vectors. Both supervised and unsupervised methods are investigated. The remaining leaves/nodes after pruning are preprocessed and each set of features are passed to individual classifiers in a MCDF

system. The classifiers used in this method are maximum likelihood classifiers, and the decision fusion scheme is a majority vote method.

In this study, two approaches to pruning are investigated. The first is based on a thresholding of leaves/nodes' performance metrics. The threshold is set in an automated fashion. The mean and standard deviation of all leaves/nodes' performance metrics are computed, and the threshold is set to the mean plus one standard deviation. For an unsupervised approach, ENTROPY was used as a performance metric, and for a supervised approach, BD was used as a performance metric. These methods were described in detail in chapter 4.

The second approach to pruning was based on a more computationally expensive bottom-up approach, and the approach is referred to as "family pruning" or FP. Terminal nodes, or leaves, are pruned based on the BD (supervised) and ENTROPY (unsupervised) performance metrics. Since this approach results in a pruned decomposition tree where only terminal nodes are allowed as feature vectors, the redundancy of the wavelet packet decomposition is removed. And the resulting tree is an optimized tree for that particular application.

Since both the handheld and airborne data sets for the chemical contamination of corn experiments were quite large (approximately 600 hyperspectral signatures per class per date), a two-fold cross-validation method was used for training and testing all of the analysis methods. For the soybean rust experiments, 84 samples were used for evaluation of a 5-class problem. Since the amount of data was very limited (as few as 10 observations for some classes), the leave-one-out cross-validation (or N-fold cross-validation) testing method was employed. The labeled observations-to-feature-ratio per

class in this study exemplifies the situation of high data dimensionality with extremely limited training data, i.e. the phenomenon known as the "curse of dimensionality".

5.5 Experimental Results and Discussion

Figure 5.7 shows the results of a comparison analysis of the newly developed WPD MCDF methods to conventional spectral-based single classifier approaches and spectral-based multiclassifier approaches. Results are reported in terms of overall classification accuracy and 95% confidence intervals. The conventional approaches used for comparison purposes include SLDA of the original signal and MCDF fixed windowing of the original signal. The analysis was conducted utilizing the 2008 Soybean hyperspectral dataset. SLDA resulted in an overall accuracy of around 40%. Thus, the single classifier approach does not perform well on this difficult dataset (very similar vegetation classes with relatively limited training data). Both the MCDF and FP_ENTROPY result in overall accuracies or around 65%. Three of the proposed methods (ENTROPY, BD, and FP_BD), however, performed well on this difficult dataset. These methods result in overall classification accuracies of 75 – 80 %.

Figures 5.8 and 5.9 show the results of a comparison analysis of the newly developed WPD MCDF methods to conventional LDA, PCA, SLDA, and MCDF methods for the detection and classification of varying levels of chemical contamination of corn for the 2008 and 2009 handheld spectroradiometer datasets, respectively. Results are reported in terms of overall classification accuracy and 95% confidence intervals. It is clear from both Figure 5.8 and Figure 5.9, the WPD MCDF methods (ENTROPY and BD) outperformed LDA, PCA, SLDA, and MCDF feature extraction and reduction methods for all dates. The efficacy of the WPD MCDF methods increase as time progresses (days

after the chemical application). For the 2008 data set, the overall accuracy (average accuracy across all 8 spray rates) for the WPD MCDF system increase from 40 to 75 % as time progresses from 1 to 14 days after chemical application. For the 2009 data set, the overall accuracy (average accuracy across all 8 spray rates) for the WPD MCDF system increases from 30 % to 70 % as time progresses from 1 to 22 days after application. However, on the later collection dates, both MCDF and LDA efficacies increases as time elapses. These results indicate the WPD MCDF approaches have the potential to discriminate between different levels of contamination rates, particularly in the critical, low-concentration rates of the contamination. However, the accuracies are considerably lower in the early stages of the contamination and, thus, indicate a limited potential for reliable early detection.

Figure 5.10 shows the results of an analysis method comparison for a scenario in which the training dataset is temporally misaligned with the test data. Results are reported in terms of overall classification accuracy and 95% confidence intervals. The analysis was conducted utilizing the 2008 corn hyperspectral dataset. In practical situations, there might not be training data (ground truthed pixels from a hyperspectral image or hyperspectral signatures collected with a handheld system) that is perfectly aligned temporally with the test imagery. That is, the end user might have training data from growth stage $V(n_0)$ (n_0th leaf) but needs to test an image collected at growth stage $V(n_1)$, where $n_0 \neq n_1$. Or, the end user might have training data from d_0 days after chemical application but needs to test an image collected at d_1 days after chemical application, where $d_0 \neq d_1$. For the temporal misalignment analysis, the 2008 corn hyperspectral dataset is analyzed for 3 periods of time. For the first period, the ATR

systems are trained on data collected d_0 days after chemical application, and the ATR systems are tested on data from d_0±4 days after chemical application. For the second and third misalignment periods, the ATR systems are trained on data collected d_0 days after chemical application, and the ATR systems are tested on data from d_0±8 and d_0±14 days after chemical application, respectively. The newly developed WPD MCDF (BD) method significantly outperformed PCA, SLDA, and MCDF feature extraction and reduction methods for all three misalignment periods. Thus, the WPD MCDF approach appears to be much less sensitive to temporal misalignments. However, overall classification accuracy for the WPD MCDF (BD) method decreased as the misalignment periods increased.

Figure 5.11 shows the results of an analysis method comparison for a scenario in which the amount of training data is varied. In practical operating conditions, the amount of ground truthed observations available for training hyperspectral ATR systems is typically very limited. Even though hyperspectral sensors offer the potential for rich spectral feature vectors, many ATR systems cannot utilize these large numbers of spectral features if the amount of training data is too small, i.e. the ratio of training samples to spectral bands is too low. This can severely limit the practicality of operational use of hyperspectral ATR systems. Thus to determine the practicality of the analysis techniques (PCA, SLDA, MCDF, and WPD MCDF (BD)), they were studied to determine their sensitivity to the number of ground truthed observations available to train the ATR system. Sensitivity was tested in a series of 10 experiments where the number of hyperspectral signatures used for each class in the training data was varied from10X, 9X, …, 1X (i.e. number of training samples is 10 times the number of spectral bands , 9

times the number of spectral bands, etc). The analysis was conducted utilizing the SpecTIR imagery for the 2008 corn experiments. As shown in Figure 5.11, the conventional method of PCA was less sensitive to training data abundance as compared to some of the more recently developed methods; however, PCA produced very low overall classification accuracies regardless of the amount of available training data. SLDA provided the highest classification accuracies when the amount of training data was very high (10X, 9X, and 8X), but when the training data abundance was severely limited (2X to 1X), SLDA's accuracies dramatically decreased. MCDF is less sensitive the amount of training data, but its performance also decreases when training data abundances are very limited (2X to 1X). The WPD MCDF approach (BD) demonstrates a very low sensitivity to training data abundance, maintaining relative steady classification accuracy over all training ratios and achieving an overall 8-class classification accuracy of around 55-60% even when the training data abundance was limited to just 1X..

Figure 5.12 shows example classification maps resulting from the PCA, SLDA, and the WPD MCDF (BD) ATR systems applied to the 2008 airborne imagery with training ratios of 10X and 1X. The 8-class overall accuracy of the 10X training ratio maps were 23% for PCA, 64% for SLDA, 51% for MCDF, and 63% for WPD MCDF (BD). The overall accuracy of the 1X training ratio maps were 23% for PCA, 33% for SLDA, 41% for MCDF, and 55% for WPD MCDF (BD). One can see from the maps that the WPD MCDF approach retains the structure of the randomized herbicide concentration spray rates within the field, even when the training data abundance is very limited (1X). Thus, if vicinal information, such as spatial features, were also

incorporated into the ATR system, the WPD MCDF approach has the potential to achieve quite high detection and classification accuracies in operational use.

The field experiments for the chemical contamination of corn consisted of 8 ground cover classes, i.e. 8 varying rates of herbicide concentration. The classes are only subtly different, and as a result the 8-class classification accuracies are low, even for the newly developed WPD MCDF approach. In a practical operating ATR system, it may not be necessary to discriminate the classes to such a fine granularity. Thus, the confusion matrices were analyzed and accuracies were determined for the WPD MCDF (BD) approach for 3 scenarios: 8-class, 4-class (herbicide concentration classes aggregated into classes of control, mild (0.03125X, 0.0625X, 0.125X), moderate (0.25X, 0.5X), and severe (1X, 2X)), and 2-class (herbicide concentration classes aggregated into classes of control and contamination (0.03125X, 0.0625X, 0.125X, 0.25X, 0.5X, 1X, 2X)). Tables 5.2, 5.3, and 5.4 show the confusion matrices for the for the 8-class, 4-class, and 2-class scenarios, respectively. These results are for the WPD MCDF (BD) ATR system applied to the 2008 handheld data collected 14 days after the herbicide application. From these tables, we can see that the accuracies are increased if the classification resolution is decreased. And for the 2-class problem (simply detecting the presence of chemical contamination and not classifying the amount of contamination), the overall accuracy is increased to 92%.

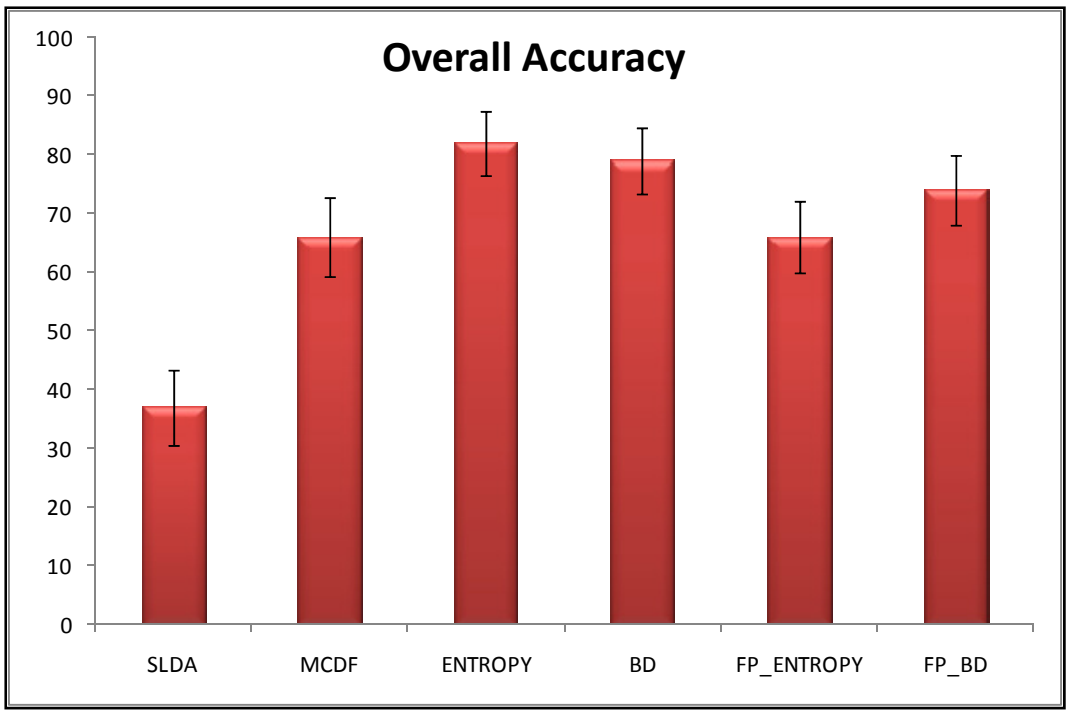


Figure 5.7 Overall classification accuracies resulting from analysis method comparison study, for a 5-class varying level of soybean rust infestation application.

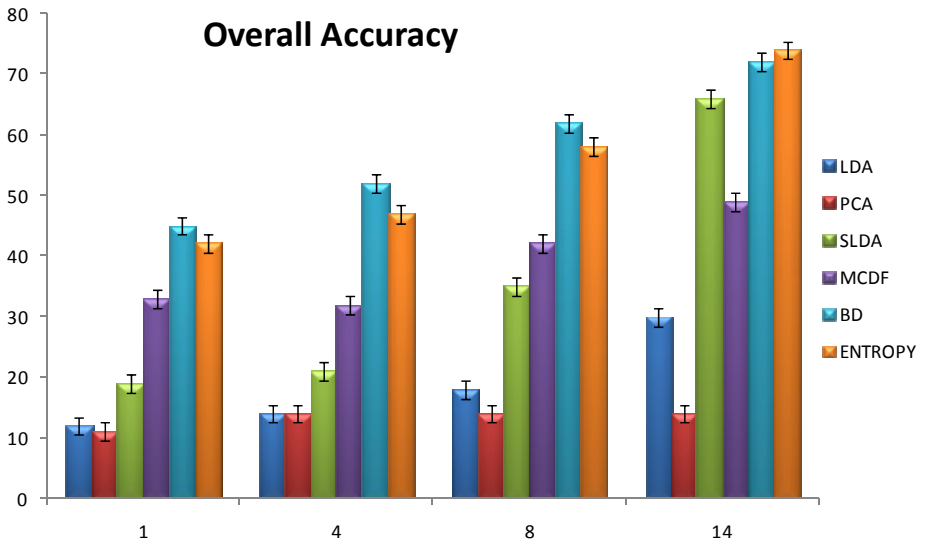


Figure 5.8 Overall classification accuracies resulting from analysis method comparison study, for an 8-class varying level of chemical contamination of corn application (2008 handheld spectroradiometer data set).

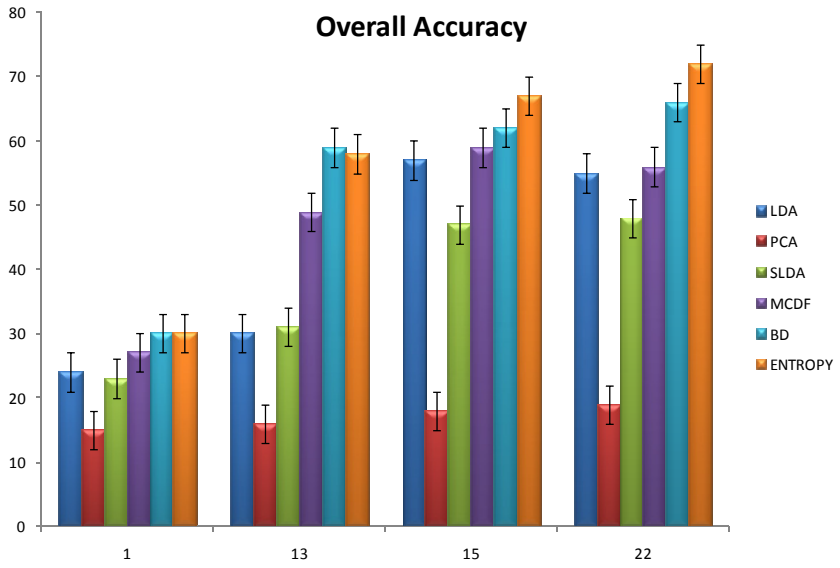


Figure 5.9 Overall classification accuracies resulting from analysis method comparison study, for an 8 class varying level of chemical contamination of corn application (2009 handheld spectroradiometer data set)

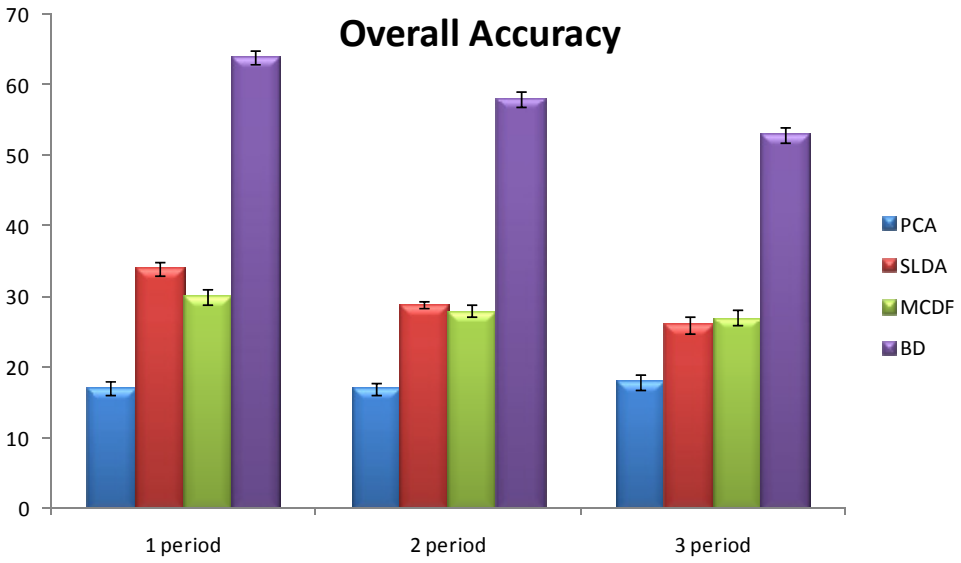


Figure 5.10 Overall classification accuracies for the WPD MCDF (BD) approach versus PCA, SLDA, and MCDF methods when training and test data are temporally misaligned: 1 period = ± 4 days, 2 period = ± 8 days, 3 period = ±14 days (2008 corn data set).

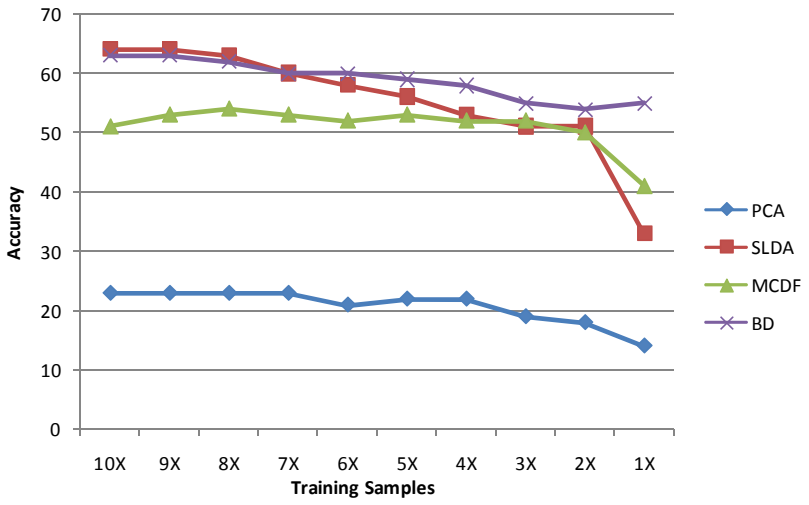
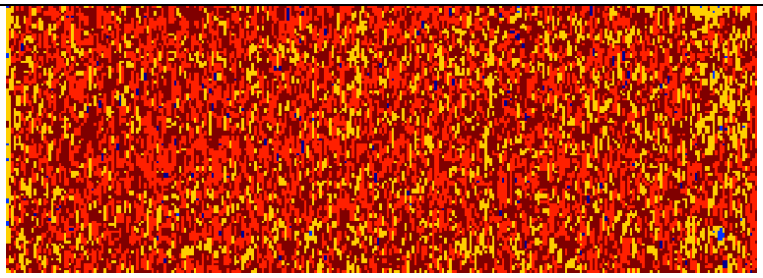
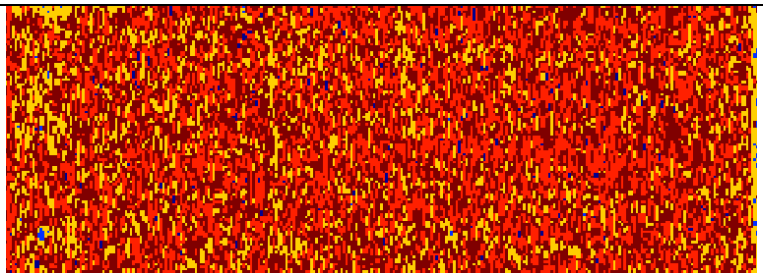


Figure 5.11 Training abundance sensitivity results: overall accuracy of classification algorithms vs. ratio of number of training samples to hyperspectral dimensionality. (2008 corn experiment, SpecTIR hyperspectral imagery)

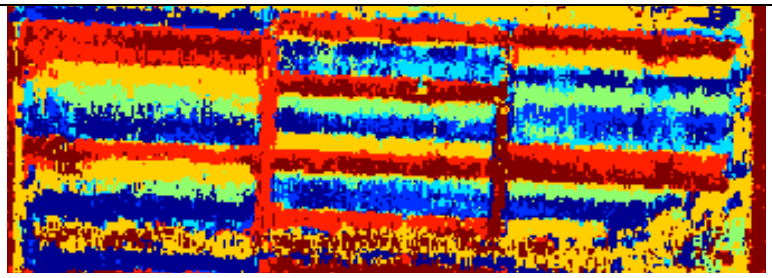
(a) SpecTIR Image (Natural color RGB)



(b) PCA 10X

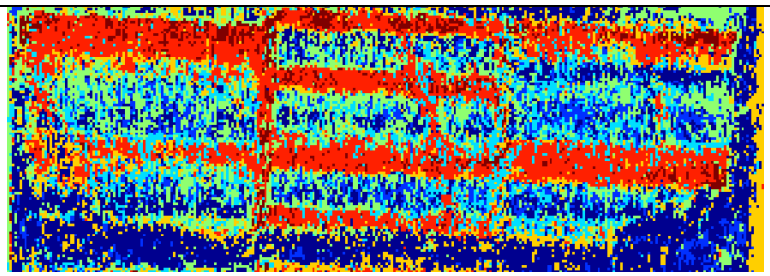


(c) PCA 1X

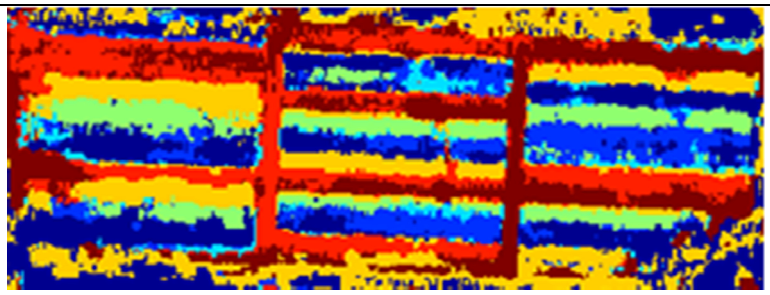


(d) SLDA 10X

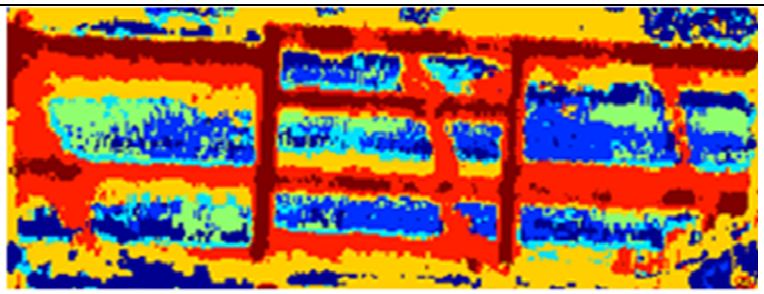
Figure 5.12 Example of herbicide concentration classification maps for agricultural field where experimental tests were carried out. (SpecTIR hyperspectral imagery, 2008 corn experiment)



(e) SLDA 1X



(f) WPD MCDF 10X



(g) WPD MCDF 1X

0 1/32 1/16 1/8 1/4 1/2 1

Herbicide Concentration Rates (Ratio of labeled rates)

Figure 5.12 (continued)

Table 5.2 Confusion matrix for WPD MCDF (BD) ATR system, 14 days after herbicide application (8-class scenario)

	2X	1X	.5X	.25X	.125X	.0625X	.03125X	Control	Producer's Accuracy
2X	84	15	2	1	0	1	0	3	79%
1X	11	82	7	1	0	0	0	16	70%
.5X	4	11	71	8	1	7	1	2	68%
.25X	3	3	3	72	11	1	4	6	70%
.125X	0	0	0	7	66	5	13	7	67%
.0625X	0	0	2	1	9	48	9	0	70%
.03125X	0	0	0	0	15	3	73	0	80%
Control	4	1	2	2	12	4	6	69	69%
User's Accuarcy	79%	73%	82%	78%	58%	70%	69%	67%	72%

Table 5.3 Confusion matrix for WPD MCDF (BD) ATR system 14 days after herbicide application (4-class scenario) [mild (0.03125X, 0.0625X, 0.125X), moderate (0.25X, 0.5X), and severe (1X, 2X)]

	Servere	Moderate	Mild	Control	Producer's Accuracy
Servere	192	11	1	0	94%
Moderate	21	154	25	13	72%
Mild	0	10	241	21	89%
Control	5	4	22	69	69%
User's Accuarcy	88%	86%	83%	67%	83%

Table 5.4 Confusion matrix for WPD MCDF (BD) ATR system 14 days after herbicide application (2-class scenario: control and all spray rates combined into one class)

	Target	NonTarget	Producer's Accuracy
Target	655	34	95%
NonTarget	31	69	69%
User's Accuracy	95%	67%	92%

5.6 Conclusion

In this study, the authors designed, developed, and applied various ATR systems to remotely sensed hyperspectral data for the detection and classification of crop contaminations, both biological, namely soybean rust, and chemical, namely herbicide applications to corn. Conventional and newly developed hyperspectral ATR methods, including PCA, LDA, SLDA, MCDF, and WPD MCDF, were tested for their efficacy with handheld spectroradiometer and airborne hyperspectral imagery. For the soybean rust experiments, the WPD MCDF approach performed significantly better than SLDA and MCDF methods. For the 5-class problem (not only detecting soybean rust but resolving the level of infestation to 4 classes of severity), the WPD MCDF techniques resulted in overall classification accuracies of 75-85%, where SLDA resulted in accuracies of 30-40%. The dramatic improvement in detection/classification accuracies stem from the fact that the WPD MCDF techniques are designed to take advantage of the rich spectral data while accounting for very limited amounts of available ground truthed training observations.

The herbicide contamination of corn experiments were repeated over two growing seasons, and similar results were obtained for both experiments. Again, the WPD MCDF approaches outperformed the conventional and current state-of-the-art analysis techniques, including PCA, LDA, SLDA. and MCDF methods, regardless of the amount of time elapsed between chemical application and collection of remotely sensed data, amount of available training data, or the quality of the available training data. The results showed that early detection of chemical applications, i.e. within a few days of application, is very difficult, especially when attempting to resolve the level of

contamination to a very fine granularity. For example, for an 8-class problem (i.e. control data and 7 levels of chemical concentration ranging from 0.03125X to 2X), the maximum accuracies achieved for 1, 4, 8, and 14 days after chemical application were approximately 40%, 50%, 60%, and 70%, respectively. However, if the classes of chemical contamination are aggregated to a lower specificity, the classification accuracies are much improved. For example, if the 8-class problem is aggregated to a 4-class problem, where the classes are control, mild, moderate, and severe contamination, the overall accuracy is increased to more than 80%. And if the 8-class problem is aggregated to a 2-class problem, where the classes are control and any level of contamination (i.e. a simple detection system), the overall accuracies are increased to more than 90%. The WPD MCDF ATR system was also shown to have relatively low sensitivity to quantity and quality of training data. When the amount of training dataset was very limited, i.e. numbers of observations are on the same order as the number of hyperspectral bands, the WPD MCDF ATR system reported overall accuracies within approximately 10% of those reported for very high abundances of training data. In practical situations, the training data might not only be limited by its abundance but also by its similarity to the test data. One example of this misalignment of training and testing data is the case where ground truthed (class labeled) training observations are collected at a vegetative growth stage that is different than the actual test imagery. From the experiments conducted in this study, the WPD MCDF ATR system produced relatively high accuracies even when this type of temporal misalignment was as severe as ±14days.

The experimental results from this study demonstrate the high potential for use of hyperspectral remote sensing for detecting and classifying various levels of biological

and/or chemical stressors in agricultural food crops. Future work should include the employment of vicinal information, such as spatial features, extracted from the hyperspectral imagery. This study utilized only spectral features (per pixel analysis), and the classification accuracies could be dramatically improved by combining the spectral features with spatial information. Also, green house studies should be performed to determine the hyperspectral ATR system's ability to discriminate between various sources of vegetative stress, such as airborne chemical, soil nutrient, and/or moisture. In this study, the chemical spray rates were randomized across the field to negate effects of soil nutrient and moisture stress. However, this should be studied in more detail.

5.7 REFERENCES

[1] J. D. Macneil, M. Hikichi, R. S. Downing, “An Investigation of the Effects of Seasonal Changes, Leaf Maturity, Nitrogen Deficiency and Leafhopper Injury on the Chlorophyll Content and Diffuse Reflectance Spectroscopic Properties of Orchard Leaves”, *International Journal of Environmental Analytical Chemistry*, vol. 31, no. 1, Nov. 1987, pp 55 – 62.

[2] T. E. Adcock, F. W. Nutter, P. A. Banks, “Measuring herbicide injury to soybeans (Glycine max) using a radiometer”, *Weed Science* , vol. 38, no. 6, Nov. 1990, pp. 625 – 627.

[3] M. L. Mortimer, J. H. Massey, D. R. Shaw, J. R. Steil, J. A. Ballweber, and M.C. Smith, “Employing Remote Sensing to Evaluate Changes in Land Use and Estimate Probable Pesticide Runoff to Surface Waters” , *Proceedings of Southern Weed Science Society*, vol. 54, 2001.

[4] R.O. Duda, P.E. Stark, D.G. Stork, *Pattern Classification*, Wiley Inter-science, October 2000.

[5] S. Prasad, L.M. Bruce, "Decision Fusion with Confidence based Weight Assignment for Hyperspectral Target Recognition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, May 2008.

[6] S. Prasad, L.M. Bruce, "Multiple Kernel Discriminant Analysis and Decision Fusion for Robust Sub-pixel Hyperspectral Target Recognition," *Proceedings of IEEE Geoscience and Remote Sensing Symposium*, Boston, MA, July 2008.

[7] T. West, L. M. Bruce, “Multiclassifiers and Decision Fusion in the Wavelet Domain for Exploitation of Hyperspectral Data”, *Proceedings of IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, July 2007.

[8] T. West, L. M. Bruce, S. Prasad, “Wavelet Packet Tree Pruning Metrics for Hyperspectral Feature Extraction”, *Proceedings of IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Boston, Massachusetts, July 2008.

[9] T. West, S. Prasad, L. M. Bruce, D. Reynolds, “Utilizations of Local and Global Hyperspectral Features via Wavelet Packets and Multiclassifiers for Robust Target

Recognition", *Proceedings of IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Cape Town, South Africa, July 2009.

[10] T. West, S. Prasad, L. M. Bruce, D. Reynolds, T. Irby, " Rapid Detection of Agricultural Food Crop Contamination via Hyperspectral Remote Sensing", *Proceedings of IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Cape Town, South Africa, July 2009.

[11] USDA Economic Research Service, "Mississippi Fact Sheet", 2006, available online at http://www.ers.usda.gov/StateFacts/MS.htm.

[12] Ellis, J.M., J.L Griffin, P.R. Vidrine and J.L. Godley, "Corn response to simulated drift of roundup ultra and liberty and utility of drift agents," *Proc. South. Weed Sci. Soc*., p. 21., 1998.

[13] Rowland, C.D., *Crop tolerance to non-target and labeled herbicide applications*, M.S. Thesis. Mississippi State University, Mississippi State, MS., 2000.

[14] M. Livingston, R. Johansson, S. Daberkow, M. Roberts, M. Ash, and V. Breneman, "Potential Economic and Policy Implications of Wind-Borne Entry of Asian Soybean Rust into the United States", 2004, available online at http://www.ers.usda.gov/AmberWaves/september04/pdf/findings_resources&environmentsept2004.pdf.

[15] Analytical Spectral Devices Fieldspec Pro Spectroradiometer specifications. Available on line at http://asdi.com/products_specificationsFSP.asp

[16] SpecTIR Airborne Hyperspectral Remotely Sensed Imagery specifications. Available on line at http://www.spectir.com/srs.htm

[17] Mississippi Department of Agriculture and Commerce. Available on line at http://www.mdac.state.ms.us/n_library/departments/bpi/index_bpi.html

[18] S. Prasad, "Multi-classifiers and decision fusion for robust statistical pattern recognition with applications to hyperspectral classification," Doctoral Dissertation, Mississippi State University, 2008.

CHAPTER 6

CONCLUSIONS

The ongoing development and increased affordability of hyperspectral sensors are increasing their utilization in a variety of applications, such as agricultural monitoring and decision making. Hyperspectral ATR systems typically rely heavily on dimensionality reduction methods, and particularly intelligent reduction methods referred to as feature extraction techniques. This dissertation reports on the development, implementation, and testing of new hyperspectral analysis techniques for ATR systems, including their use in agricultural applications where ground truthed observations available for training the ATR systems are typically very limited.

In recent years, MCDF approaches for hyperspectral ATR systems have been developed and shown to be quite effective for scenarios where training data is very limited. However, these approaches have the limitation that they utilize only narrow groups of contiguous spectral bands. That is, the current MCDF systems only utilize localized subsets of the spectral signature and do not take advantage of global characteristics of the signature. This dissertation combines multiresolutional analysis, namely DWT and WPD, with MCDF to overcome this limitation. New methods are developed and tested for grouping and selecting wavelet coefficients so they can be input and effectively used by a MCDF scheme. The new wavelet-based MCDF systems are

tested for their sensitivity to choice of mother wavelet, level of decomposition, DWT coefficient partitioning technique, wavelet packet filter tree pruning technique, quantity of available training data, and quality of available training data (in terms of temporal alignment between training and testing data). In all ATR systems investigated in this dissertation, MCDF approaches used LDA preprocessing, maximum likelihood classifiers, and majority vote decision fusion schemes. All single classifier methods that were used for comparison purposes utilized LDA for feature optimization along with maximum likelihood classifiers. This consistency across ATR systems allowed for a fair comparison of methods.

The newly developed methods, as well as commonly used current state-of-the-art methods for comparison purposes, were applied to hyperspectral data from an agricultural application. The methods were tested on handheld spectroradiometer data and airborne hyperspectral imagery, both of which were collected over two growing seasons for this dissertation. The application was the detection and classification of food crop contamination, either by an airborne chemical application, specifically Glufosinate herbicide at varying concentrations applied to corn crops, or by biological infestation, specifically soybean rust disease in soybean crops.

6.1 Conclusions from DWT MCDF approach

The author developed, implemented, and tested five DWT coefficient grouping/selection methods: CONCAT, SCALAR, SUBSPACE, ENTROPY, and BD. In general, two approaches (ENTROPY and BD) outperformed the others. The two methods performed quite well, even when the methods were trained and tested on significantly different datasets (training data was from a greenhouse study in 2005 and

testing data was from a field campaign in 2008). Each of the two approaches applies a performance metric to each terminal node of the DWT decomposition. The nodes are down selected based on a thresholding of the performance metrics, where the threshold is based on the performance metric statistics. The DWT coefficients of each selected node are then used as a feature vector input to an individual classifier in the follow-on MCDF system. ENTROPY and BD approaches use unsupervised (entropy) and supervised (Bhattacharyya distance) performance metrics. The BD method was more sensitive to mother wavelet selection, as compared to the other DWT coefficient grouping/selection methods. However, none of the five approaches demonstrated a significant sensitivity to mother wavelet selection. For the ENTROPY approach, the Daubechies-8 mother wavelet slightly outperformed the other mother wavelets investigated. However, it could be argued that the simplicity of the Haar mother wavelet could outweigh the slight increase in performance gained by using a more complicated mother wavelet. The simplicity of the Haar mother wavelet could result in faster implementations, as well as potentially safe-guarding the ATR system against over-training. The BD method was more sensitive to the selection of decomposition level, as compared to the ENTROPY method. The ENTROPY approach mostly outperformed the BD method and was constant for any level of decomposition. One of the most interesting outcomes of the study was the high performance of the relatively simple ENTROPY method, which is unsupervised, and is more commonly used in DWT compression and denoising applications than in ATR applications. BD was expected to significantly outperform ENTROPY, yet they typically performed on par with one another.

6.2 Conclusions from WPD MCDF approach

The author developed, implemented, and tested several variations of a WPD MCDF ATR system, where the variations mainly stemmed from the method of pruning used to optimize the WPD tree for ATR purposes. The author designed and tested two types of WPD tree pruning methods. These WPD tree pruning methods were designed to increase computational efficiency and possibly improve classification accuracies simultaneously. The pruning approaches resulted in a set of WPD nodes/leaves, each containing a set of approximation or detail coefficients that were then used as a feature vector input to the MCDF scheme. Both pruning approaches were implemented using unsupervised performance metrics, namely entropy, and supervised metrics, such as BD. One pruning approach used a straightforward thresholding of metrics from all nodes/leaves to determine which nodes/leaves were selected as feature vectors. The second pruning approach used an intelligent approach that ensured a non-redundant dyadic decomposition where the WPD nodes with the highest performance metric were terminal nodes. Then all terminal nodes were selected as feature vectors for input to a follow-on MCDF system.

The experimental results showed the WPD MCDF approaches to be significantly superior, in terms of overall accuracies, to the conventional LDA and SLDA approach. For example, for the soybean rust detection/classification application, SLDA resulted in overall classification accuracies around 30-40%; spectral domain MCDF approaches resulted in overall accuracies around 65%; and WPD MCDF approaches achieved overall accuracies as high as 80%. It was surprising that the experimental results showed the

highest classification accuracies stemmed from the use of the simpler and less computationally expensive thresholding approach for pruning and unsupervised metric, namely entropy. The WPD MCDF approaches were tested for their sensitivity to mother wavelet selection, across the family of Daubechies wavelets. Interestingly, most of the WPD MCDF techniques worked well with a simple Haar mother wavelet. The Haar and Daubechies-10 mother wavelets both performed quite well regardless of WPD tree pruning method. Again however, as with the DWT MCDF approach, it could be argued that the simplicity of the Haar mother wavelet could outweigh the slight increase in performance gained by using a more complicated mother wavelet. The simplicity of the Haar mother wavelet could result in faster implementations, as well as potentially safe-guarding the ATR system against over-training.

6.3 Conclusions from agricultural application

The author applied the newly developed DWT MCDF and WPD MCDF approaches to both the soybean rust and the corn herbicide hyperspectral datasets, using both handheld and airborne data. In general, the optimum system design was determined to be a WPD MCDF approach using threshold pruning with the BD performance metric, Haar mother wavelet, and 5 levels of decomposition. An in depth study of this WPD MCDF ATR system was conducted to determine its potential for use in hyperspectral remote sensing of crop contaminants. The WPD MCDF ATR system was compared with more conventional methods, including PCA, LDA, SLDA, MCDF, and WPD MCDF, to determine their comparative efficacies with handheld spectroradiometer and airborne hyperspectral imagery.

For the soybean rust experiments, the WPD MCDF approach performed significantly better than SLDA and MCDF methods. For the 5-class problem (not only detecting soybean rust but resolving the level of infestation to 4 classes of severity), the WPD MCDF techniques resulted in overall classification accuracies of 75-85%, where SLDA resulted in accuracies of 30-40%. The dramatic improvement in detection/classification accuracies stem from the fact that the WPD MCDF techniques are designed to take advantage of the rich spectral data while accounting for very limited amounts of available ground truthed training observations.

The herbicide contamination of corn experiments were repeated over two growing seasons, and similar results were obtained for both experiments. Again, the WPD MCDF approaches outperformed the conventional and current state-of-the-art analysis techniques, including PCA, LDA, SLDA. and MCDF methods, regardless of the amount of time elapsed between chemical applications and collection of remotely sensed data, amount of available training data, or the quality of the available training data. The results showed that early detection of chemical applications, i.e. within a few days of application, is very difficult, especially when attempting to resolve the level of contamination to a very fine granularity. For example, for an 8-class problem (i.e. control data and 7 levels of chemical concentration ranging from 0.03125X to 2X), the maximum accuracies achieved for 1, 4, 8, and 14 days after chemical application were approximately 40%, 50%, 60%, and 70%, respectively. However, if the classes of chemical contamination are aggregated to a lower specificity, the classification accuracies are much improved. For example, if the 8-class problem is aggregated to a 4-class problem, where the classes are control, mild, moderate, and severe contamination,

the overall accuracy is increased to more than 80%. And if the 8-class problem is aggregated to a 2-class problem, where the classes are control and any level of contamination (i.e. a simple detection system), the overall accuracies are increased to more than 90%. The WPD MCDF ATR system was also shown to have relatively low sensitivity to quantity and quality of training data. When the amount of training data was very limited, i.e. number of observations are on the same order as the number of hyperspectral bands, the WPD MCDF ATR system reported overall accuracies within approximately 10% of those reported for very high abundances of training data. In practical situations, the training data might not only be limited by its abundance but also by its similarity to the test data. One example of this misalignment of training and testing data is the case where ground truthed (class labeled) training observations are collected at a vegetative growth stage that is different than the actual test imagery. From the experiments conducted in this study, the WPD MCDF ATR system produced relatively high accuracies even when this type of temporal misalignment was as as severe as ±14days.

The experimental results from this study demonstrate the high potential for use of hyperspectral remote sensing for detecting and classifying various levels of biological and/or chemical stressors in agricultural food crops. Additional future work should include the employment of vicinal information, such as spatial features, extracted from the hyperspectral imagery. This study utilized only spectral features (per pixel analysis), and the classification accuracies could be dramatically improved by combining the spectral features with spatial information. Also, green house studies should be performed to determine the hyperspectral ATR system's ability to discriminate between various

sources of vegetative stress, such as airborne chemical, soil nutrient, and/or moisture. In this study, the chemical spray rates were randomized across the field to negate effects of soil nutrient and moisture stress. However, this should be studied in more detail.

6.4 Suggested Future Work

The methods developed and tested in this dissertation could be extended in a variety of ways, both in terms of their implementation and their application. For example, this dissertation investigated only one family of mother wavelets, namely Daubechies. The newly developed methods could be tested for other families of mother wavelets or with adaptive mother wavelets for a more generalized solution.

Also, the current implementation of the WPD MCDF system uses a simplistic method to determine whether or not to utilize a feature reduction/optimization preprocessing method (e.g. LDA). If the preprocessing method is inappropriate for any single WPD node, it is not applied to any and all nodes. For example, LDA cannot be applied to a feature vector whose dimension is less than the total number of classes in the application. In the current implementation of the WPD MCDF system, each node's dimension is assessed. If any single node's dimension is too small for use of LDA, then LDA is not applied to any node. These determinations of preprocessing for WPD nodes could be more intelligent and adaptive.

This dissertation reports on only two performance metrics, namely entropy and BD. Preliminary investigations were conducted on other performance metrics, such as product of BD and correlation, product of BD and average mutual information, etc. While these particular performance metrics did not result in accuracies higher than BD or

entropy, it could be worthwhile to investigate different performance metrics in more detail.

This MCDF scheme used in this dissertation utilized a very simple decision fusion method, specifically the majority vote. The proposed methods could be improved if a more sophisticated decision fusion method were used, such as qualified majority vote, linear opinion pooling, and logarithmic opinion pooling.

This dissertation only investigated the DWT and WPD methods of multiresolution analysis. The proposed methods could be extended to cases where other multiresolution decomposition methods were used, such as curvelets, ridgelets and non-dyadic wavelet trees.

The experimental results from this dissertation demonstrate the high potential for use of hyperspectral remote sensing (when utilizing a WPD MCDF approach) for detecting and classifying various levels of biological and/or chemical stressors in agricultural food crops. Additional future work should include the employment of vicinal information, such as spatial features, extracted from the hyperspectral imagery. This study utilized only spectral features (per pixel analysis), and the classification accuracies could be dramatically improved by combining the spectral features with spatial information. Also, green house studies should be performed to determine the hyperspectral ATR system's ability to discriminate between various sources of vegetative stress, such as airborne chemical, soil nutrient, and/or moisture. In this study, the chemical spray rates were randomized across the field to negate effects of soil nutrient and moisture stress. However, this should be studied in more detail.

Finally, the proposed DWT MCDF and WPD MCDF approaches could be applied to other ATR applications where the observations/measurements have a dimensionality much higher than amount of available training data, such as speech processing, face recognition, medical imaging applications, etc.