12-11-2009

# Fusion of Spectral Reflectance and Derivative Information for Robust Hyperspectral Land Cover Classification

Hemanth Reddy Kalluri

# FUSION OF SPECTRAL REFLECTANCE AND DERIVATIVE INFORMATION FOR ROBUST HYPERSPECTRAL LAND COVER CLASSIFICATION

By

Hemanth Reddy Kalluri

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

Mississippi State, Mississippi

December 2009

FUSION OF SPECTRAL REFLECTANCE AND DERIVATIVE INFORMATION FOR ROBUST HYPERSPECTRAL LAND COVER CLASSIFICATION

By

Hemanth Reddy Kalluri

Approved:

| | |
|---|---|
| Lori M. Bruce<br>Associate Dean<br>Bagley College of Engineering<br>Professor<br>Department of Electrical and Computer Engineering (Major Advisor) | Saurabh Prasad<br>Assistant Research Professor<br>GeoSystems Research Institute<br>Adjunct Professor<br>Department of Electrical and Computer Engineering (Co-Major Advisor) |
| James E. Fowler<br>Professor and Graduate Program Director<br>Department of Electrical and Computer Engineering<br>(Committee Member) | James E. Fowler<br>Professor and Graduate Program Director<br>Department of Electrical and Computer Engineering<br>(Graduate Coordinator) |

Sarah A. Rajala
Dean of the Bagley College of Engineering

Name: Hemanth Reddy Kalluri

Date of Degree: December 11, 2009

Institution: Mississippi State University

Major Field: Electrical Engineering

Major Professor:

Title of Study: FUSION OF SPECTRAL REFLECTANCE AND DERIVATIVE INFORMATION FOR ROBUST HYPERSPECTRAL LAND COVER CLASSIFICATION

Pages in Study: 79

Candidate for Degree of Master of Science

Developments in sensor technology have made high resolution hyperspectral remote sensing data available to the remote sensing analyst for ground cover classification and target recognition tasks. Further, with limited ground-truth data in many real-life operating scenarios, such hyperspectral classification systems often employ dimensionality reduction algorithms. In this thesis, the efficacy of spectral derivative features for hyperspectral analysis is studied. These studies are conducted within the context of both single and multiple classifier systems. Finally, a modification of existing classification techniques is proposed and tested on spectral reflectance and derivative features that adapts the classification systems to the characteristics of the dataset under consideration. Experimental results are reported with handheld, airborne and spaceborne hyperspectral data. Efficacy of the proposed approaches (using spectral derivatives and single or multiple classifiers) as quantified by the overall classification accuracy (expressed in percentage), is significantly greater than that of these systems when exploiting only reflectance information.

# DEDICATION

I would like to dedicate this thesis to my parents, brothers, sisters, and friends who have been my continual support.

ACKNOWLEDGEMENTS

I would like to thank my major advisor Dr. Lori M. Bruce for her advice and support while working on this research. I would like also to thank my co-major advisor Dr. Saurabh Prasad for his supervision, trust, support and guidance from the early stages of this research. Finally, I would like to acknowledge my committee member, Dr. James E. Fowler for his support throughout the research.

I gratefully acknowledge the financial support provided by the Geosystems Research Institute (GRI) at MSU for this research. I would also like to acknowledge Jeff Brantley, Jacob Bowen, Terrance West, Matthew Lee, Dr. Lori Bruce, Dr. Saurabh Prasad and Dr. Daniel Reynolds for collaborating in the hyperspectral data collection campaigns. I would like to thank fellow graduate students and faculty members of the Geosystems Research Institute and Department of Electrical and Computer Engineering for their support and advice. I would like to thank my parents, brothers, sisters, and friends for their continuous love and support.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

Development of accurate and robust image classification algorithms has been a major area of research in the field of remotely sensed data analysis. With the advancements in imaging sensor technology, remote sensing has been made possible even for inaccessible and dangerous areas. Insufficient amounts of ground truth (labeled) data and the redundancy present in many current state-of-the-art remotely sensed data affect the performance of image classification algorithms. Although significant accomplishments have been made for signal processing and exploitation of remotely sensed data, there is typically a tradeoff between the performance of an algorithm and its complexity. A new approach of exploiting spectral derivatives for improved classification of hyperspectral imagery is presented in this thesis. The proposed approach simultaneously exploits information in the reflectance signatures and higher order derivatives in an efficient manner – a multi-classifier decision fusion framework is employed to efficiently utilize the high dimensionality of the resulting feature space.

## 1.1 Remote sensing and applications

Remote sensing involves the acquisition of information about an object or a scene using a sensing device that makes no physical contact with that object. Data is typically collected by sensors on-board aircrafts (airborne imagery), satellites (spaceborne imagery), ships etc., which also enables acquisition over otherwise inaccessible areas

such as forests, valleys and glaciers. Remote sensing can be passive or active. A passive sensor records radiation reflected by the object with sun being the source of radiation while an active sensor emits its own radiation and detects the reflected (or backscattered) radiation from the object under observation. Examples of passive sensors include optical sensors and radiometers, and those of active sensors include Radio Detection and Ranging (RADAR) and Light Detection and Ranging (LIDAR) sensors. Such sensing techniques have a wide variety of applications, such as monitoring forest fires and deforestation, monitoring floods, making topographic maps, mineral mapping, soil moisture estimation, land-cover classification and target recognition.

## 1.2 Multispectral and hyperspectral data

In the earlier days of remote sensing, aerial photography and camera recordings were used for topographic mapping and radiometric analysis. For image analysts, these images provided good differentiation in recognizing classes that were distinctly separate such as vegetation and soil, soil and water etc. Recognition of classes that are closely related (e.g. separating deciduous tree species in a forest, different soil conditions) requires more information from a wide range of wavelengths in the electromagnetic spectrum, at a finer spectral resolution. Developments in optical sensor technology have made this possible by capturing the data in hundreds of bands over a broad range of wavelengths. Sensors capable of recording data in multiple bands are categorized into two types - multispectral and hyperspectral.

Multispectral images contains data collected in a few spectral bands that are optimally chosen and are typically not contiguous, while hyperspectral sensors collect data in hundreds to thousands of contiguous bands. Examples of multispectral and

hyperspectral sensors include National Aeronautics and Space Administration (NASA) Landsat Multispectral Scanners (MSS) aboard Landsat satellites1-5, Sea-Viewing Wide Field-of-view Sensor (SeaWiFS), Indian Remote Sensing System (IRS), Linear Imaging Self-scanning Sensor (LISS-III and LISS IV), NASA Terra Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), DigitalGlobe, Inc. (QuickBird), Space Imaging, Inc. (IKONOS), Leica Geosystems, Inc. Airborne Digital Sensor System (ADS-40) and Analytical Spectral Devices, Inc. (ASD) handheld Spectroradiometer [1]. Applications of multispectral and hyperspectral imaging include agriculture crop management, mineral extraction and air surveillances. Figure 1.1 depicts the overview of a typical remote sensing system and Figure 1.2 shows a remotely sensed image of Mississippi State fields at Brooksville, MS using an airborne hyperspectral sensor named Pro-SpecTIR-VNIR acquired by SpecTIR$^{TM}$ for Mississippi State University.

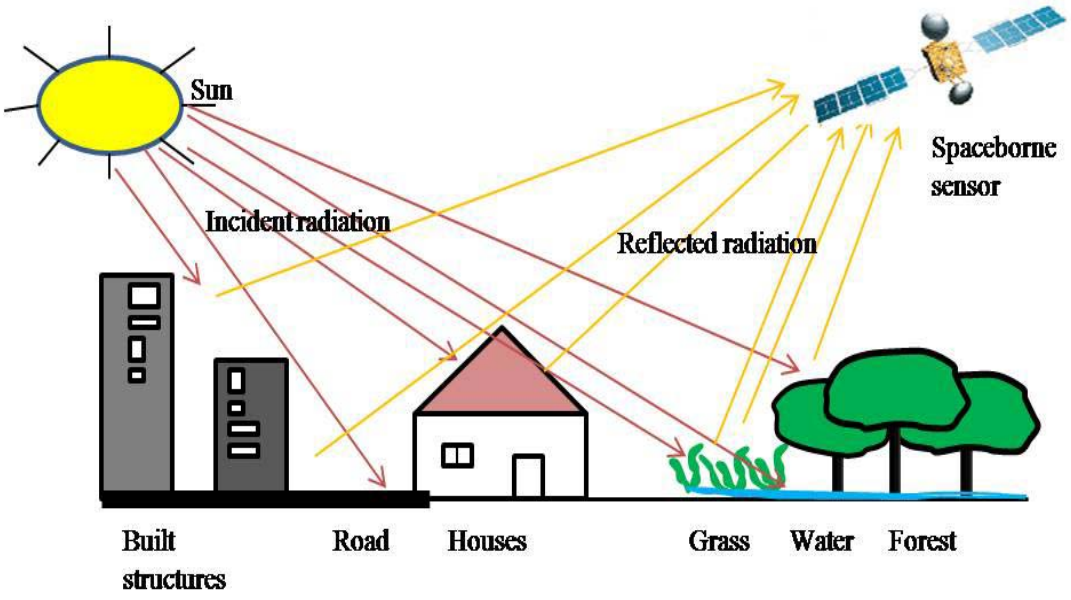


Figure 1.1 Example optical passive remote sensing procedure using a space borne sensor.

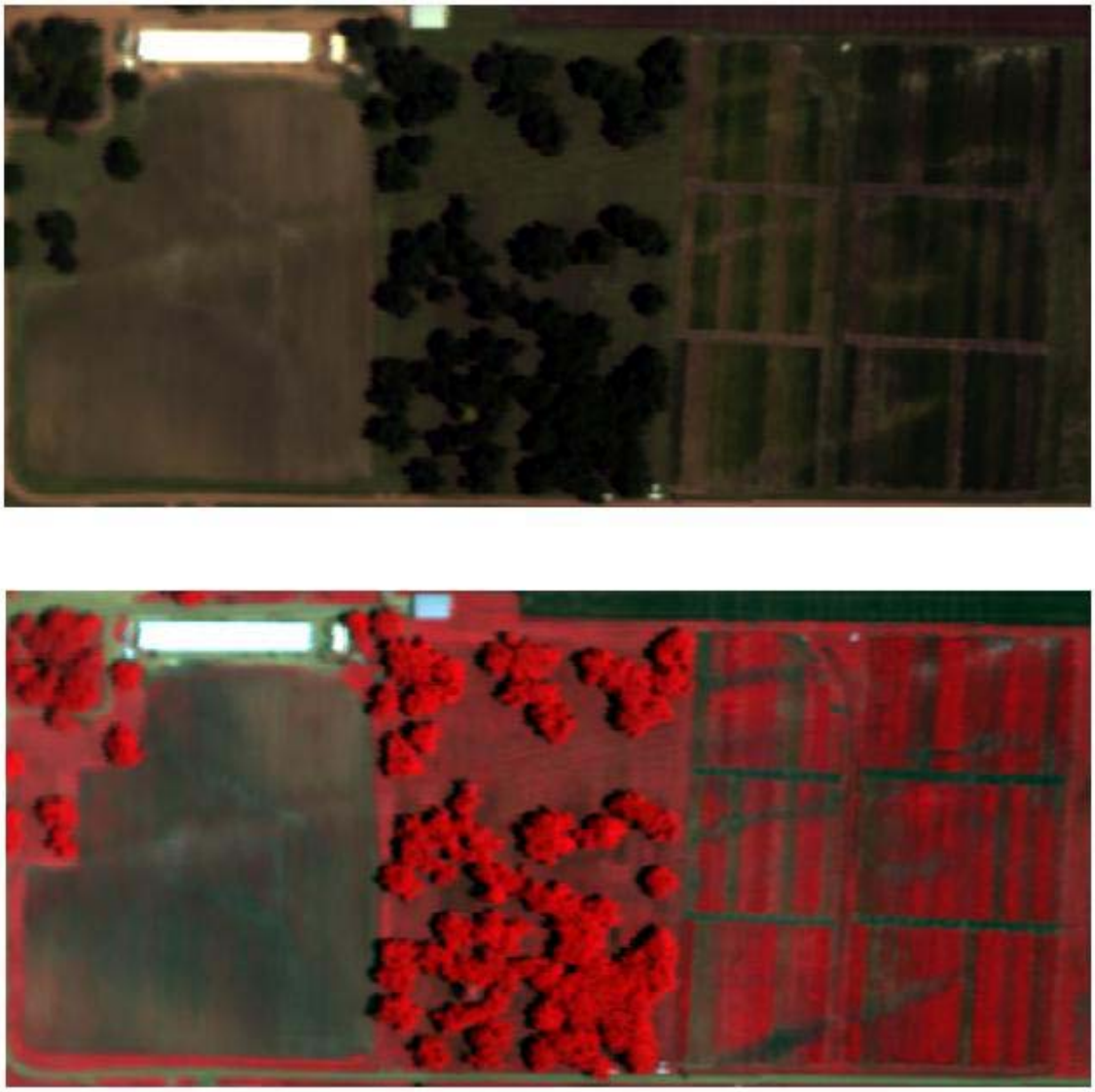

Figure 1.2 Optical remote sensing – Top: RGB true color composite with Blue-488nm, Green-533nm and Red-602nm. Top: RGB false color composite with Blue-533nm, Green-602nm and Red-753nm. Both show hyperspectral image of Mississippi State fields at Brooksville, MS, acquired using an airborne SpecTIR$^{TM}$ inc. sensor.

## 1.3 Pattern recognition and methodologies

Pattern recognition involves “labeling” input data with different category labels (classes). Pattern recognition problems are of two types – supervised and unsupervised. In supervised classification, the classifier parameters are learned from available labeled sample data (training data), which are then used to label unlabelled data samples. Unsupervised classification is similar to a data organization problem where the user is provided with only unlabeled data which is to be classified into different types. An example of an unsupervised classification technique is clustering. In this study, we restrict ourselves to supervised pattern recognition techniques. Figure 1.2 shows the block diagram of a typical pattern recognition system.

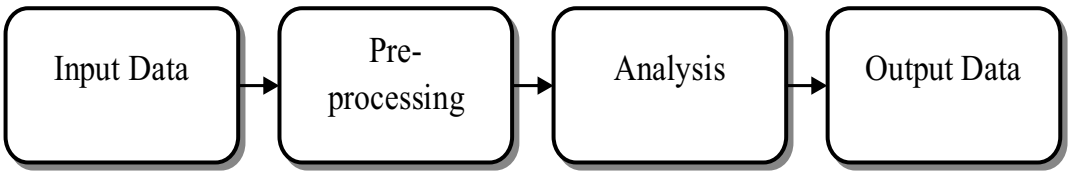


Figure 1.3 Block diagram of a typical pattern recognition system.

Input data refers to data which requires further analysis to be understood or for deriving some useful information from it. Some examples of input data in a pattern classification setting are speech data, medical imagery, and remotely sensed data. After data acquisition, the data is typically preprocessed before proceeding with any analysis. Preprocessing includes steps such as noise removal, registration of images, atmospheric corrections and calibration in the case of remotely sensed optical data. Analysis involves employing a pattern classification technique to label or cluster the processed input data.

For remote sensing classification tasks, output data from such analyses could be a classification map, abundance estimates and soil characteristics, soil moisture maps, etc.

## 1.4 Limitations of hyperspectral data analysis

Despite the abundance of information, hyperspectral datasets present some key challenges to data analysts. With increased dimensionality, one would expect increased target recognition and classification accuracies with multispectral and hyperspectral data, but accuracy is often traded off for complexity (large data computations) in the algorithms. Insufficient amounts of training data (small sample size) with high dimensionality when used to learn a classifier always result in over fitting (Hughes's phenomenon) and misclassification. Thus the use of dimensionality reduction and feature extraction schemes has become an important part of hyperspectral image analysis systems. Currently, most research in hyperspectral image classification is concentrated on developing algorithms that provide near-optimal dimensionality reduction and feature extraction. Figure 1.3 shows a general hyperspectral image classification system.

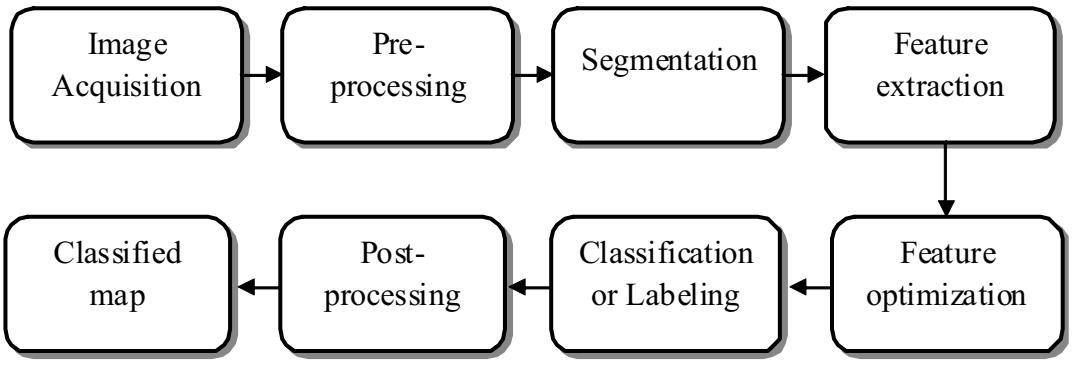


Figure 1.4 Block diagram of a typical hyperspectral image analysis system.

## 1.5 Contributions of this work

This thesis investigates the benefits of spectral derivatives as "features" for improving hyperspectral image classification. Current state-of-the-art hyperspectral feature extraction algorithms exploit only reflectance information from the spectral signatures for classification, ignoring the available slope or derivative information. Figure 1.4 shows a simple block diagram of the hyperspectral classification system used in this study. In this study, we explore the benefits of considering higher order derivatives as features to improve the classification performance of hyperspectral data. In particular, we study the benefits of combining spectral reflectance information with derivative information for classification.

The benefits of spectral derivatives for classification are studied in the context of two different types of pattern recognition systems – a traditional single classifier system, and a recently proposed multi-classifier system. The single classifier system is based on employing Stepwise-Linear Discriminant Analysis (S-LDA) for feature reduction and optimization and a single Maximum Likelihood (ML) classifier for class labeling (also called SLDA-ML in this work). Combining reflectance features with spectral derivatives further increases the dimensionality of the feature space, thereby exacerbating the problem of over dimensionality on typical traditional classification system. To overcome this, a recently developed classification framework, the Multi-Classifier Decision Fusion (MCDF) is employed and its' benefits are studied [2].

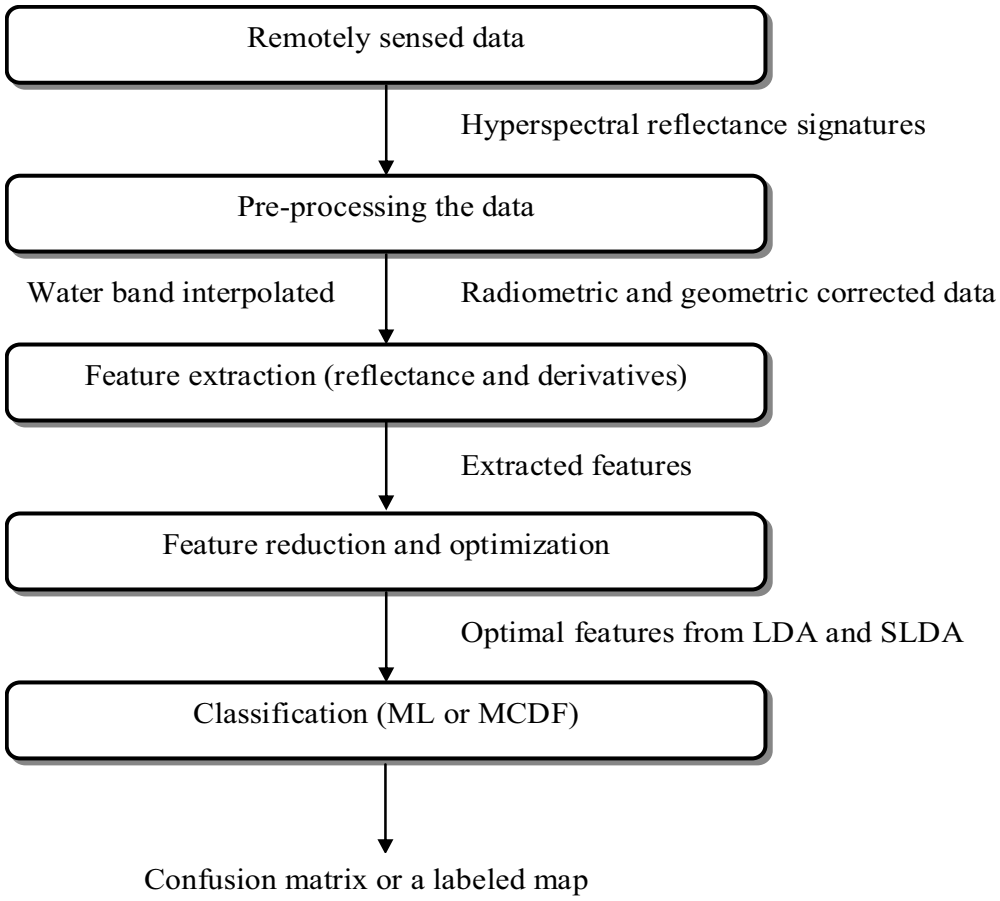


Figure 1.5 Block diagram of the hyperspectral image analysis system used in this study.

The outline of this thesis is as follows. Chapter 2 presents a literature survey of current state-of-the-art methods employing dimensionality reduction, single classifier, multi-classifier and decision fusion techniques in the raw reflectance domain and the spectral derivative domain for hyperspectral classification and target recognition. Chapter 3 provides a description of the proposed approach and the algorithms employed in this work for incorporating spectral derivatives effectively for classification. Chapter 4 provides a description of the experimental setup and the hyperspectral datasets employed for quantifying the efficacy of the proposed algorithms, and provides a summary of

classification and target recognition results. Chapter 5 concludes this thesis with a summary of results and potential future directions

# CHAPTER 2

# LITERATURE REVIEW

The main aim of hyperspectral image analysis in most applications is to classify or label different kinds of patterns present in an image. For the past two decades, researchers have been successful in developing different kinds of classification systems, which were either application specific or generic. Figure 1.3 (page 6) shows a typical hyperspectral image classification system. Optimization of each block in this figure is an ongoing research area. This study concentrates on optimizing feature extraction and classification for improved hyperspectral land cover classification.

## 2.1 Feature reduction (optimization)

State-of-art pattern recognition methods developed previously for multispectral and gray-scale imagery were based on datasets with small dimensionality. These techniques are not optimal when working with high dimensional spaces, such as those resulting from hyperspectral imagery. Although such datasets can potentially provide valuable classification information, they can exacerbate problems such as data redundancy and over-fitting (Hughes's phenomenon [3]). In such circumstances, one would prefer to use feature extraction and classification techniques that classify data using a smaller dimensional subset of the original feature space while keeping most of the relevant information intact. Most feature reduction and optimization techniques attempt

to select features and perform dimensionality reduction projections along directions that best separate the classes under consideration in these projected spaces.

Principal Component Analysis (PCA) [4] is a basic mathematical dimensionality reduction technique, which transforms a highly correlated vector space into an uncorrelated space, conserving most of the variability present in the input data. It is a very popular data compression scheme and is also often used for dimensionality reduction. PCA performs an Eigen-analysis on the second order covariance matrix developed from the training data. Eigenvectors corresponding to larger eigenvalues are referred to as the principal components, which tend to retain most of the variability in the original data. Dimensionality reduction of the feature space is achieved by ignoring the components along which the data scatter is the least. Farrell *et al.* [5] demonstrated how some principal components of PCA can be used for dimensionality reduction and feature optimization in data classification. The authors were also successful in achieving greater classification accuracies with PCA when compared to that obtained using some other popular dimensionality reduction schemes of their time. Although it is often used for dimensionality reduction, it is not necessarily optimal for classification tasks. Prasad *et al* [6] demonstrated mathematically and experimentally that PCA can potentially discard "useful" directions, and is not optimal for classification tasks.

Fisher's Linear Discriminant Analysis (FLDA) [7] projects the input data onto a $c$-1 dimensional hyper plane ($c$ is the number of classes in the classification task) and then finds an optimal direction for the hyper plane that best separates the data projections. FLDA finds the optimal direction by using both the within class and between class scatter matrices that best separate the classes under consideration. However in some situations,

such as when the input data is multi-modal in distribution FLDA projections fail to separate the classes adequately. Prasad *et al* [8] studied the limitations of FLDA in such data conditions, and proposed alternate nonlinear approaches to overcome this limitation [9] [10]. Lee *et al.* [11] proposed a novel approach for extracting the optimal features based on decision boundaries. With this approach the authors were able to find a minimum number of features that yield same classification accuracy, which is achieved by considering the original space for any given classification problem. The feature optimization technique was successful in removing data redundancy, thereby reducing the computational costs associated with processing high dimensional data. A general case where the above feature optimization technique fails to function properly is when the data is multimodal. Stefan *et al.* [12] developed a feature extraction algorithm for hyperspectral image analysis that takes some real world scenarios into consideration, such as pixel mixing. Peter *et al.* [13] introduced the concept of ranking the bands based on some metrics like entropy, contrast measure and correlation measure. After ranking the bands, the best ranked ones are selected as the optimal features for further classification.

Apart from their individual limitations, a common problem with all the above methods is that the performance of these algorithms depends on the amount of training data available, which is used to learn these projections. In situations where the available training data is much less relative to the dimensionality of the data (the 10N rule [14]), statistical estimates required to learn the projections are likely to be ill-conditioned, thereby yielding sub-optimal features.

Considering the limitations of FLDA and other dimensionality reduction techniques, some researchers developed algorithms that employ these dimensionality reduction techniques on a reduced subset of available features. One such attempt was made in [15] [16], where features were selected in a stepwise format based on the forward-selection and backward-rejection of individual features from a high dimensional feature space into a small subset of features, based on the class separation that the feature provides. Class separation provided by the feature is obtained by using a metric that relies on means and covariance-matrices of the individual classes for calculating the inter class distance. Next, FLDA is employed on this subset of features for dimensionality reduction. The authors were successful in reducing the high dimensional feature space into a small subset of optimal feature space, which provided good target recognition accuracies when compared with that achieved by considering the complete feature space in small-sample-size conditions. One limitation of this algorithm is that it ignores the remaining features once a set of features are selected, and hence can discard potentially useful information. The authors limited the application of the above algorithm to a two class problem, which is extended to multi class problems in this study.

In general, one common drawback of these algorithms is that they tend to throw away valuable information present in the narrowly spaced hyperspectral bands. This suggests the need for a hyperspectral image classification framework that better exploits the available high-dimensional features without discarding features away.

## 2.2 Multi-Classifier Decision Fusion framework (MCDF)

Recently, there has been an increased research in data fusion techniques for remotely sensed data, where data from different sensors is fused for various recognition

and identification tasks, for example, Memarsadeghi *et al*. [17] who studied fusion of data from two different sensors for invasive species forecasting. Although the multi-sensor data fusion was carried out due to a lack of sufficient information (spatial and spectral resolution) from one single sensor, this fusion technique can also be employed for a high dimensional single-sensor classification tasks – where we partition the high dimensional feature space into many smaller dimensional subspaces, treating each subspace as having come from a different sensor, and then fusing results from each subspace for a combined classification result per pixel. With this approach, all features in the hyperspectral data are likely to be used effectively in the classification process.

A similar idea was successfully studied and implemented by Fauvel *et al.* [18] where the authors used the idea for classification of urban images. Fuzzy fusion techniques were used by Chanussot *et al.* [19] for detecting linear features in synthetic aperture radar images with application to road network extraction. The authors used the fusion techniques for combining the results from multi-temporal data. Recently, Prasad *et al.* [20] [21] and [22] developed a multi-classifier decision fusion framework, where the feature space from high dimensional hyperspectral data is partitioned into "optimal" subsets that are treated independently for classification and these "local" classification results are then fused. The authors used a divide-and-conquer approach to overcome the small sample size problem by dividing the high dimensional data into subgroups, where feature optimization and classification in each subgroup is carried out separately and then classification results from each subgroup were combined using a decision fusion mechanism. The authors developed efficient algorithms for optimal feature partitioning

and techniques for effective data fusion. An attempt to improve the performance of this classifier system to multi-class data is made by using adaptive learning techniques.

## 2.3 Spectral derivatives and benefits

The concept of derivative analysis has its roots in analytical chemistry, where it was successfully used in spectroscopy for many years. The same theory is applied to remote sensing applications by researchers for improving classification performance of the classification systems for remote sensing data [23]. Goodin *et al.* [24] used first and second order numerical derivatives of the reflectance spectrum for discriminating chlorophyll signals from those of suspended solid particles present in the water. William D. Philpot [25] was successful in avoiding the atmospheric effects from airborne remotely sensed data using derivatives and band ratios. The author was able to discriminate two classes (vegetation and water) using a derivative ratio algorithm from the distorted data. Fuan Tsai *et al.* [26] also used higher order spectral derivatives in a land cover based classification. The authors used a PCA based dimensionality reduction scheme to reduce the dimensionality of features from the reflectance spectrum and its derivatives, which are then fused and sent to the classifier system. Derivative features entering the classifier are restricted using PCA reduction, which, as previously discussed is not an optimal way of selecting features. More recently, Demir *et al.* [27] studied the fusion of first and second order spectral derivatives with spectral reflectance in an attempt to improve the classification performance. Their usage of spectral derivatives was also confined to lower order derivatives, and they also used conventional techniques for feature extraction and classification.

# CHAPTER 3

# METHODOLOGIES

## 3.1 Motivation

Most of the conventional pattern recognition systems for hyperspectral image analysis discussed in chapter 2 confined their feature extraction procedure to the available reflectance values present in the narrowly spaced spectral bands. These methods did not exploit the slope or derivative information present in the reflectance signatures. The key motivation behind this work was to study the benefits of spectral derivative information for effective hyperspectral classification. The outline of this chapter is as follows. Section 3.2 describes the concept of estimating spectral derivatives from hyperspectral data. Sections 3.3 – 3.6 describe the various dimensionality reduction and classification techniques employed in this work.

## 3.2 Spectral derivatives

Derivatives quantify the change in value of a function with respect to changes in the independent variable. In hyperspectral imagery, spectral derivatives are estimated by obtaining the slope information from the reflectance curve over the available wavelengths in the spectrum. The process of estimating a derivative is called differentiation, and the order of the derivative is the number of times the function is being differentiated.

Hyperspectral data collected in real time uncontrolled conditions is bound to be contaminated with different types of noise. Apart from general random noise, thermal noise, shot noise, atmospheric effects causing path irradiance, viewing angles and illumination affects introduce additive noise into the data. Hyperspectral sensors when employed for longer periods of time in hot weather conditions tend to introduce some noise that typically spectrally coherent. Derivatives are very sensitive to these disturbances. Additive noise in the reflectance signatures usually gets severely magnified when derivatives are calculated on such data. Therefore it becomes necessary to pass the reflectance signatures through an appropriate filtering process that removes such disturbances. In this study, mean and median filtering algorithms were used to filter out the unnecessary noise present in the hyperspectral signatures.

### 3.2.1 Mean filtering

Mean filtering is a simple process of smoothing out unexpected reflectance variations between consecutive spectral bands. Replacing each reflectance value in a band with the mean (average) reflectance value in its neighboring bands forms the basis of a mean filter. This effectively removes any values which are unrepresentative of their neighborhood. Filter order is the number of surrounding neighbors that are considered when computing the average. Usually odd numbers starting from three are employed as filter orders. Increase in the order of the filter increases the smoothing effect on the signature, but can also blur out certain sharp features.

### 3.2.2 Median filtering

Median filtering is similar to mean filtering except that instead of replacing the value with the average of the surrounding reflectance values, it replaces it with the median of those values. Median filtering is considered to be more effective than mean filtering in terms of preserving the useful details (shape) present in the signatures.

### 3.2.3 Estimating spectral derivatives

In practice, computation of derivatives depends on the order of the derivative and the sampling order of the spectral measurements. The first derivative is calculated by using the formula below

$$\frac{\partial R}{\partial \lambda_i} = \frac{R(\lambda_j) - R(\lambda_i)}{\nabla \lambda}, \tag{3.1}$$

where $\nabla\lambda = \lambda_j - \lambda_i$ is the separation between the adjacent bands at $\lambda_j$ and $\lambda_i$, with $\lambda_j > \lambda_i$, $R(\lambda_j)$ is reflectance value at $\lambda_j$ and $\frac{\partial R}{\partial \lambda_i}$ is the first derivative at wavelength $\lambda_i$.

Similarly second and third derivatives are defined as

$$\frac{\partial^2 R}{\partial \lambda_j^2} = \frac{R(\lambda_k) - 2R(\lambda_j) + R(\lambda_i)}{\nabla \lambda^2}, \tag{3.2}$$

where $\frac{\partial^2 R}{\partial \lambda_j^2}$ is second derivative at wavelength $\lambda_j$, $\nabla\lambda = \lambda_j - \lambda_i = \lambda_k - \lambda_j$ and $\lambda_i < \lambda_j < \lambda_k$., and

$$\frac{\partial^3 R}{\partial \lambda_k^3} = \frac{R(\lambda_l) - 3R(\lambda_k) + 3R(\lambda_j) - R(\lambda_i)}{\nabla \lambda^3}, \tag{3.3}$$

where $\frac{\partial^3 R}{\partial \lambda_k^3}$ is the third derivative at wavelength $\lambda_k$, $\nabla\lambda = \lambda_j - \lambda_i = \lambda_k - \lambda_j = \lambda_l - \lambda_k$ and $\lambda_i < \lambda_j < \lambda_k < \lambda_l$.

This could be generalized to an $n^{th}$ order derivative as

$$\frac{\partial^n R}{\partial \lambda_i^n} = \frac{\sum_{j=i}^{i+n} (-1)^{i+n-j} * C(n,(j-i+n)) * R(\lambda_j)}{(\nabla\lambda)^n} \quad . \tag{3.4}$$

Here $C(n, j)$ is the combinatorial function of n and j. For every n, n+1 wavelengths are considered where $\lambda_i, \lambda_{i+1}, \lambda_{i+2}$ and so on are separated by the desired sampling order. Here $\nabla\lambda$ is the difference between any two adjacent wavelengths considered, with uniform sampling across the spectrum.

Sampling order is defined as the difference between two adjacent spectral bands, where reflectance value is available and also plays an important role while calculating the derivatives. Thus, three user defined parameters, filter order (of the smoothing filter), derivative order (differentiation) and sampling order (the separations between the wavelengths), are to be carefully selected before including derivatives into a feature space for classification.

Figure 3.1 shows the plot of experimental hyperspectral signatures of two aquatic plant species named American Lotus (a native species) and Water Hyacinth (an invasive species) collected using an ASD hyperspectral sensor. Details of the sensor used for collecting this data are provided in chapter 4. The mean signatures of the two classes are plotted in reflectance and derivative domain against the wavelengths.

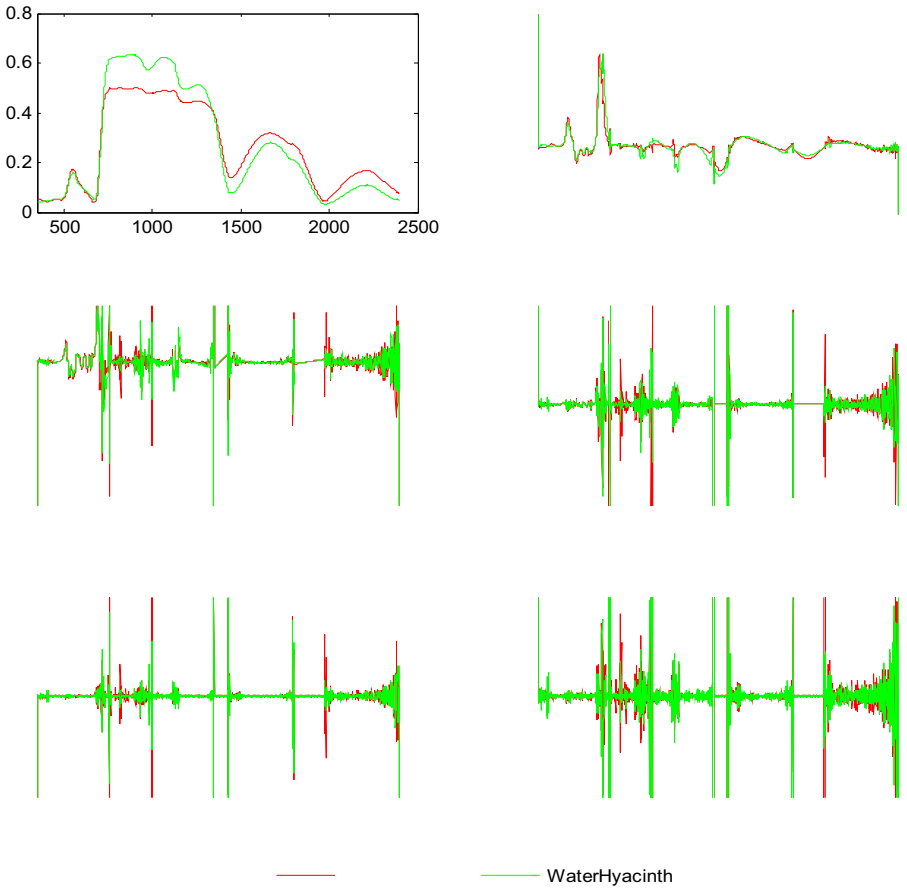


Figure 3.1 Plot of mean of the original signatures in reflectance domain and the derivative signatures of first five orders for a experimental hyperspectral two class data.

Figure 3.2 shows the plot of Bhattacharyya distance (A metric that quantifies class separation for Gaussian distributions – C.f. page 23, eq. 3.13) versus wavelength for the first order derivative features plotted against the Bhattacharyya distance calculated using the reflectance features for comparison. The original hyperspectral signatures in reflectance domain are from the previously mentioned American Lotus-Water Hyacinth data collected using an ASD sensor. Bhattacharyya distance is a metric for measuring class separation. For a two class problem (which is the case here), distance between the

classes is calculated for every feature, while for a multi-class problem (as will be reported in the next chapter) the minimum of the pair-wise distance measurements of all the classes is considered in finding the class separating distance.

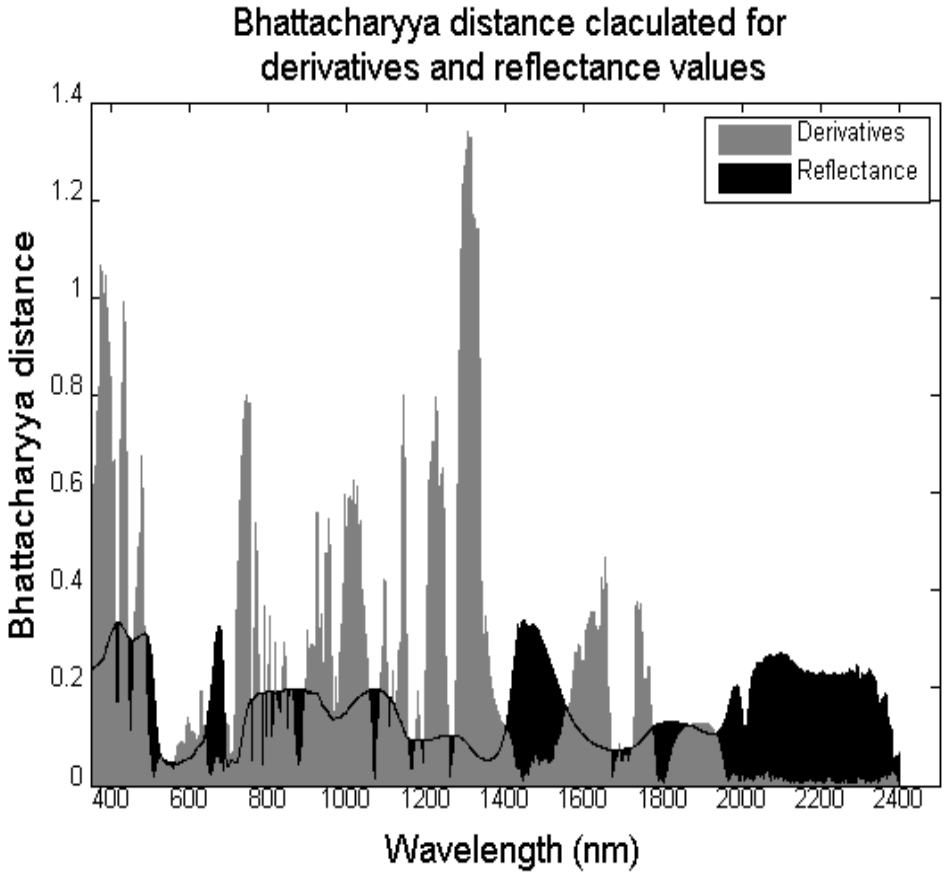


Figure 3.2 Plot of Bhattacharyya distance vs. wavelength for derivative features plotted against that of reflectance values.

The increased metric value with derivatives, when compared to that with raw reflectance values suggests that the inclusion of spectral derivatives into the feature space will result in improved classification performance. With an appropriate selection of these features with high Bhattacharyya distances into the feature set, we expect the classification accuracies to be higher than those achieved using the reflectance values

only. Further, at many wavelengths when the distance is small in the reflectance domain, the corresponding distance is high in the spectral domain, and vice-versa. This implies employing both reflectance and derivative features simultaneously is going to be beneficial.

### 3.3 Fisher's linear discriminant analysis (FLDA)

Discriminant analysis involves finding directions that are effective for discrimination. Linear discriminant analysis (LDA) constructs discriminant functions that are linear in the input variables, resulting in linear decision boundaries. LDA seeks to project a d-dimensional feature space onto a c-1 dimensional hyper plane in a direction or orientation that best separates the projected samples. Fisher's LDA method also finds the linear combination of inputs or, in some sense, a direction that best separates the classes under consideration. FLDA finds an optimal linear direction by maximizing the between class separability while simultaneously minimizing the within class variability.

The criterion maximizes the Rayleigh quotient, given by

$$J_F = \frac{\left|\vec{\omega}^T S_b \vec{\omega}\right|}{\left|\vec{\omega}^T S_w \vec{\omega}\right|}, \tag{3.5}$$

where ω is the optimum direction, $S_b$ is the between class covariance matrix and $S_w$ is the within class covariance matrix. This problem is solved by using the Eigen-analysis technique. For optimal projections that best separate the classes, it is required for the within class matrix to be symmetric and full ranked which requires ample amount of training data from which these projections are learned.

The within class covariance matrix is given by

$$S_\omega = \sum_{i=1}^{c} \sum_{i=1}^{n_k} (\overrightarrow{x_i} - \overrightarrow{\mu_k})(\overrightarrow{x_i} - \overrightarrow{\mu_k})^T , \quad (3.6)$$

where $\overrightarrow{\mu_k}$ is the within class mean for class k given by

$$\overrightarrow{\mu_k} = \frac{1}{n_k} \sum_{i=1}^{n_k} \overrightarrow{x_i} , \quad (3.7)$$

and $\overrightarrow{m}$ is the total mean given by

$$\overrightarrow{m} = \frac{1}{n} \sum_{i=1}^{n} \overrightarrow{x_i} , \quad (3.8)$$

where n is the total number of samples in all the classes, $n_k$ is the number of samples in class k and *c* is the number of classes.

The between class covariance matrix is given by

$$S_b = \sum_{i=1}^{c} n_k (\overrightarrow{\mu_k} - \overrightarrow{m})(\overrightarrow{\mu_k} - \overrightarrow{m})^T. \quad (3.9)$$

For a 'c' class distribution data, FLDA produces a transformation into a hyper space of dimension at most equal to c-1. The vector w that maximizes the equation (3.5) must satisfy the condition

$$S_b \overrightarrow{\omega} = \lambda S_\omega \overrightarrow{\omega}. \quad (3.10)$$

This takes the form of a generalized Eigen value problem given by

$$S_\omega^{-1} S_b \overrightarrow{\omega} = \lambda \overrightarrow{\omega}. \quad (3.11)$$

After ω is found, the optimal feature projection is calculated by taking a product of complete features with the optimal projection.

$$f_{opt} = \vec{x} * \overrightarrow{\omega}, \quad (3.12)$$

where $f_{opt}$ is the optimal reduced-dimensional feature space that has good class separation.

Figure 3.3 illustrates the operation of FLDA. In the figure an example two-dimensional, two class data is considered to explain the FLDA operation with data distribution along the two dimensions is shown on the left side of the figure. The main aim of FLDA is to find the direction of a projecting line (or a hyper-plane for a multi-class problem) that best separates both classes. Two directions X and Y, for the projecting line are marked on the figure. FLDA finds the direction Y (least overlap) with the help of scatter matrices as the optimal direction discarding the direction X (high overlap). Right side of the figure shows the distribution functions of both the classes when projected onto the line along the Y direction.

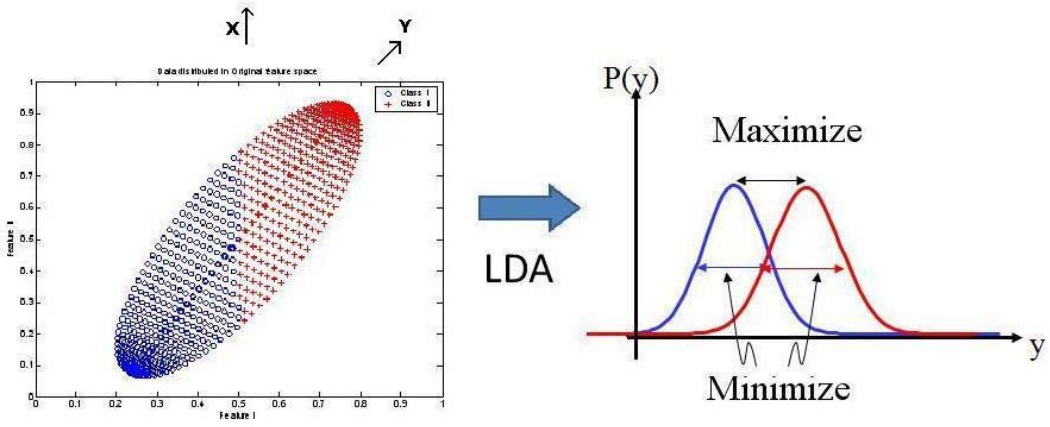


Figure 3.3 Fisher's LDA for a two class problem.

### 3.4 Stepwise-linear discriminant analysis (SLDA)

Stepwise LDA is a type of compromise for FLDA where, instead of finding an optimal projection from the original high dimensional feature space, a sub-optimal projection is found by working on a relatively smaller dimensional subset of the feature

space. This kind of feature optimization best suits situations where the feature space has high dimensions and the available training data sample size is low. Use of SLDA for such data not only decreases the number of computations but is also effective for classification tasks.

Performance of SLDA depends on the choice of the discriminating metric used for identifying an appropriate subset of the original high dimensional feature space. There are many metrics that could be considered for SLDA, including area under the receiver operating characteristics (ROC), Bhattacharyya distance (BD), and Jeffries Matusita (JM) distance. In this study, BD has been chosen as the metric.

### 3.4.1 Bhattacharyya distance

Bhattacharyya distance (BD) in general statistics is used to measure the similarities between two discrete probability distribution functions (PDF). In image processing, BD is used to assess the class separation capability in a feature space. For a two-class problem, BD is estimated as follows

$$\mathrm{BD} = \frac{1}{8}(\vec{\mu}_2 - \vec{\mu}_1)^{\mathrm{T}} * \left[\frac{\sum_1+\sum_2}{2}\right]^{-1} * (\vec{\mu}_2 - \vec{\mu}_1) * \ln\left(\frac{\frac{|\sum_1|+|\sum_2|}{2}}{|\sum_1|+|\sum_2|}\right), \qquad (3.13)$$

where $\vec{\mu}_1$ and $\vec{\mu}_2$ are the means and $\sum_1$ and $\sum_2$ are the covariance matrices for class 1 and class 2. One drawback of BD is that it assumes the PDF is Gaussian while calculating the distance.

### 3.4.2 Functioning of SLDA

For an N dimensional feature space, SLDA finds the BDs separating the classes along each of the N individual dimensions to yield a 1 X N BD vector. For a two class problem, Bhattacharyya distance between the two classes is taken and stored, while in the case of a multi-class problem, the minimum of all possible pair-wise class separations is used. The BDs are sorted in descending order. The rest of the functionality can be divided into two parts, namely, forward selection and backward rejection. Figure 3.4 shows a flowchart of the SLDA operations.

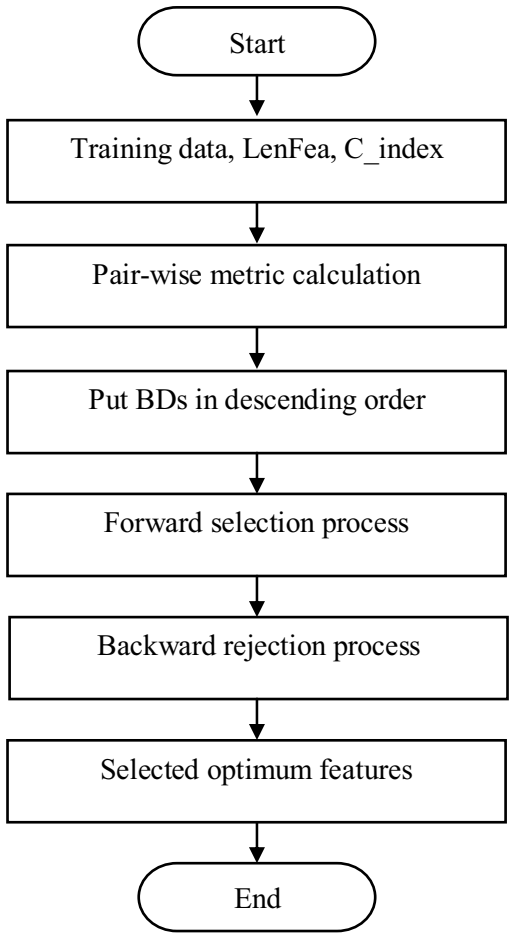


Figure 3.4 Flowchart of the Stepwise-linear discriminant analysis operations.

#### 3.4.2.1 Forward selection

Figure 3.5 shows the flowchart of the forward selection process. In forward selection, for a 'c' class problem, the first c features with highest BDs are selected and are then sent through an FLDA for reducing them to c-1 feature space. Then, BD is calculated for this reduced dimensional feature space and is saved as BD1. Then the feature with the next best BD is included into the original feature space, which goes through FLDA to get a new reduced dimensional feature space. BD calculated on the new space is now stored as BD2. A comparison operation performed between BD1 and BD2 decides the inclusion (BD1 < BD2) or rejection (BD1 > BD2) of the last added feature into the feature stream. This process goes on until either the entire feature space gets exhausted or a certain "maximum allowable" dimensionality of feature space (based on the user's choice of the variable LenFea) is reached.

#### 3.4.2.2 Backward rejection

Figure 3.6 shows the flowchart of the backward rejection process. In the backward rejection process, the final BD with the selected feature space from the forward selection process is calculated and stored as BD1. Then, of the N features selected, the first feature is removed from the feature space and is then sent through FLDA to get a reduced dimensional feature space, on which a new BD is calculated and saved as BD2. Rejection (BD1 < BD2) or retention (BD1 > BD2) of the removed feature from the feature space is performed. In this way, a final subset of the original feature space that best separates the classes under consideration is selected. Since this algorithm finally employs FLDA on a reduced subset of features (based on the outcome of the forward

selection, backward rejection), the resulting FLDA formulation is expected to be well-conditioned even when a relatively smaller amount of training data is available.

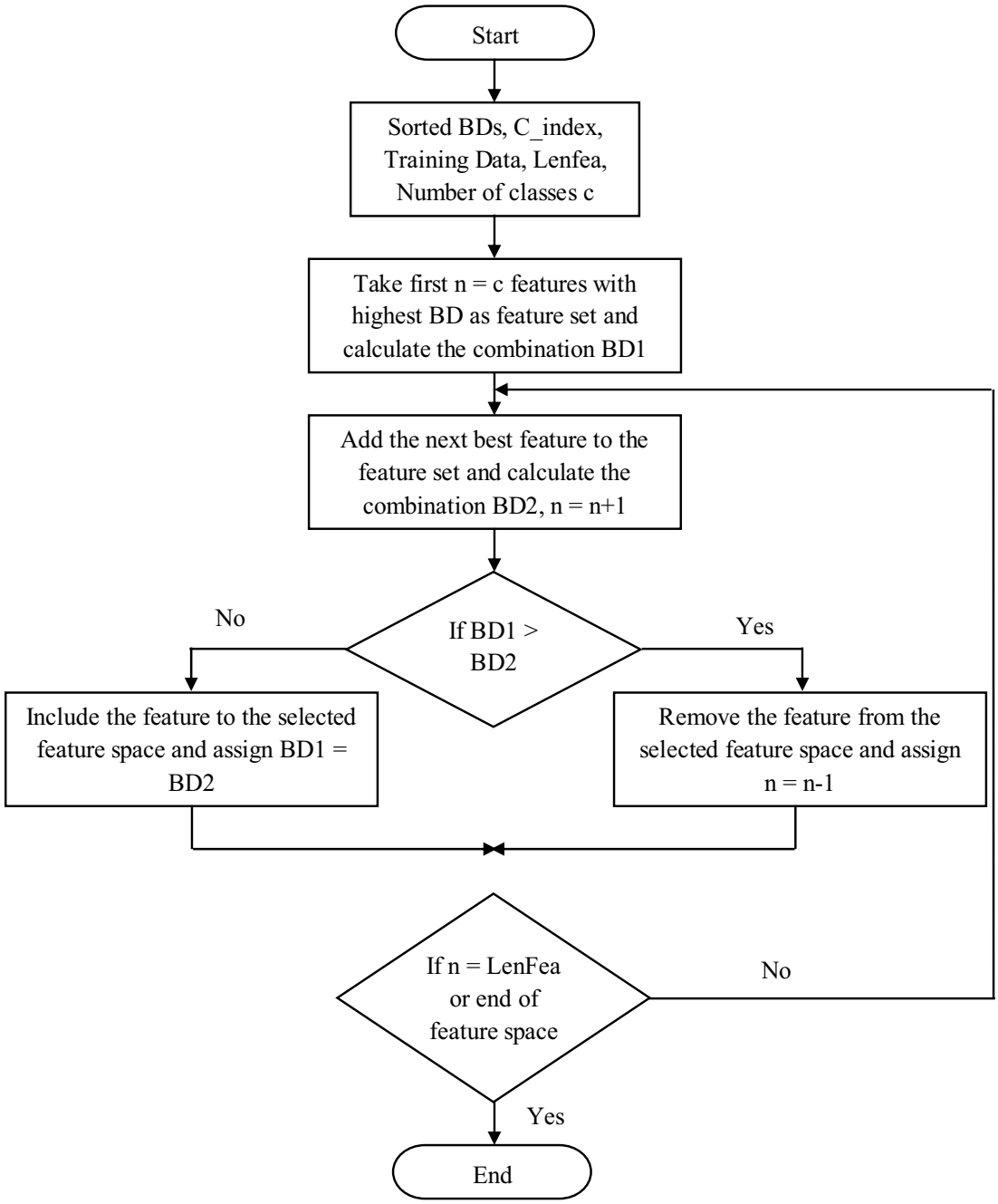


Figure 3.5 Flowchart of the forward selection process of the SLDA operations.

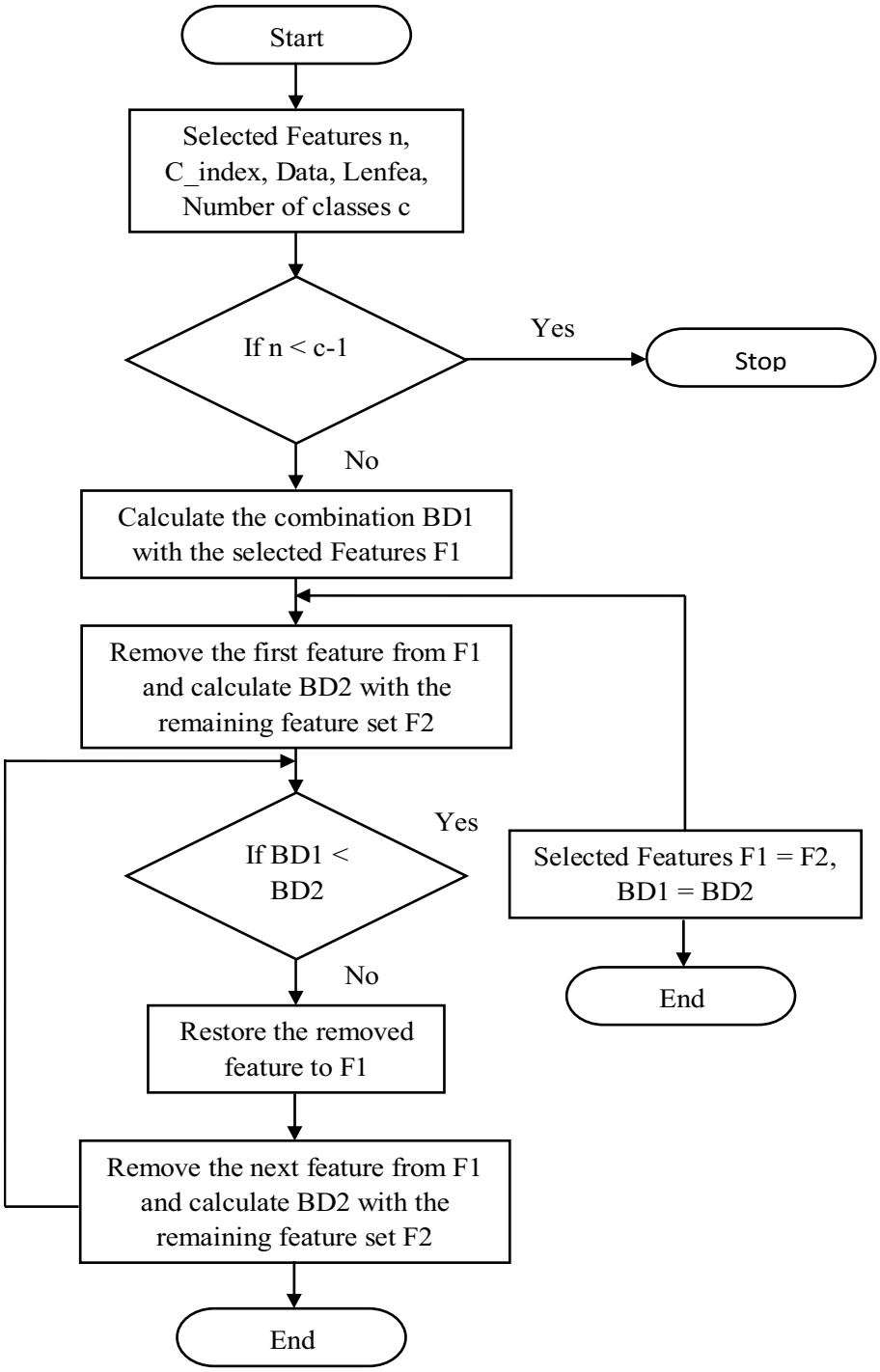


Figure 3.6 Flowchart of the backward rejection process of SLDA operations.

### 3.5 Maximum likelihood classifier

Maximum Likelihood classifier (ML) is defined as a decision rule that estimates the probabilities of the unlabeled data conditioned on all available classes and assigns a label to this data that results in the highest probability. Given its simplicity and efficacy, the maximum likelihood classifier is commonly used for image analysis.

The maximum likelihood decision rule is given by

$$\mathrm{P}(\mathrm{c}/\vec{x}) = \frac{1}{2} * \left(\vec{x} - \vec{\mu}_c\right) * \sum_c^{-1} * \left(\vec{x} - \vec{\mu}_c\right)^T + \frac{1}{2} * ln(|\sum_c|) - ln(P_c)\ , \qquad (3.14)$$

where the class label is decided by finding the highest posterior probability. In equation 3.14, $\vec{\mu}_c$ is the mean vector for the c'th class, $\sum_c$ is the covariance matrix for the c'th class, and Pc is the a-priori probability for the c'th class. The ML classifier assumes that the probability distribution function of each class is a Gaussian distribution with mean $\vec{\mu}_c$ and covariance $\sum_c$, given by $\mathrm{P}\left(\frac{\vec{\mathrm{x}}}{\mathrm{C}}\right) \sim \mathrm{N}(\vec{\mu_c}, \Sigma_c)$. The advantages of maximum likelihood estimation include simplicity in representation and good convergence for ample amounts of training data. A key drawback is that it assumes data to be normally distributed, which is not always the case.

In situations where the amount of training data is much less than the dimensionality of the data, the covariance matrices estimated by the classifier are usually ill-conditioned (not being full ranked). This affects the matrix inverse calculations, thereby making classification unreliable under such situations.

### 3.6 Multi-classifier decision fusion

Figure 3.7 shows the flowchart of the MCDF operations. Multi-classifier decision fusion (MCDF) involves partitioning a high dimensional feature space into groups of

smaller dimensions, assigning a dedicated classifier per group, and fusing results from these classifiers into a single class label per data sample. Functioning of a MCDF framework can be divided into three parts – subspace identification, multiple classification and decision fusion.

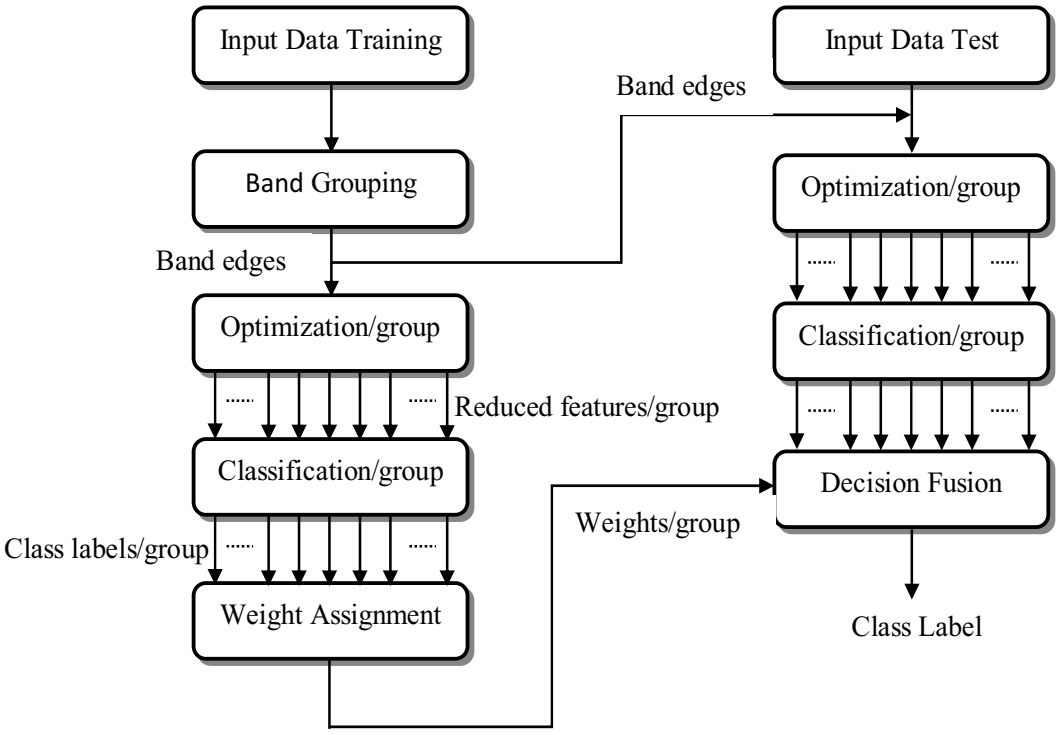


Figure 3.7 Flowchart of a Multi classifier decision fusion framework with adaptive weight assignment.Subspace identification.

Subspace identification, also called band grouping, involves grouping of the hyperspectral feature space into contiguous subgroups so that each subspace/subgroup possesses good class separation and correlation between the subspaces is minimum. Using training data, the band grouping algorithm first keeps on adding individual spectral bands (features) into a subgroup until a metric governing the class separation does not

change with further addition of features. Along with maintaining good class separation, the partitioning should ensure minimum intergroup correlation. Good class separation within the subspace ensures accurate classification in further stages, reducing the local classification errors. Minimum correlation between the subspaces further ensures robust decision fusion, avoiding propagation of correlated errors.

Three parameters governing the subspace partitioning are employed in this study, and they are threshold, size bound, and minimum size. Threshold is used to monitor the sensitivity of partitioning to changes in the separation metric. It acts as a stopping criterion for the growing group size. For example for threshold 't', band grouping of the current group stops when

$$\frac{\text{metric}_n - \text{metric}_{n-1}}{\text{metric}_{n-1}} < t \,, \tag{3.15}$$

and at this point, the next grouping starts. In this study we maintained the threshold 't' as zero.

Size bound lays an upper bound on the number of bands grouped into one subspace. This ensures that the size of each group is not so large that it breaks the feature optimization and classification steps that follow band grouping. Minimum size keeps a check on the number of band groups formed.

### 3.6.1 Metric

In this study the metric used for the band grouping process is BDCorr. BDCorr is the product of BD and the Correlation Coefficient. BD is given by equation 3.13, and is defined in the previous section. Correlation coefficient is a measure of the second-order

statistical relationship between two random variables. Correlation coefficient for two variables x and y is given by

$$R([x\ \ y]) = R(i,j) = \frac{\mathrm{C(i,j)}}{\sqrt{\mathrm{C(i,i)} * \mathrm{C(j,j)}}}, \tag{3.16}$$

where C = cov(x) is the covariance matrix.

For a data set with N variables, a correlation coefficient matrix of dimensions 2 x 2 is calculated for all $N^2$ paired variables, of which the minimum value is taken and multiplied with the BD to obtain BDCorr.

### 3.6.2 Multi-classifiers

With partitioned subgroups available, an individual classifier system is run on each subgroup for obtaining local per group decisions, which are then sent to a decision fusion system for obtaining a global class label per data sample. The multi-classifier system is essentially a bank of classifiers operating locally on the partitioned subspaces. For this study, an LDA based feature reduction, followed by a maximum likelihood classifier is used to make these local decisions.

Recall that feature reduction using LDA tends to be suboptimal in the case of available training data being less than the dimensionality of the feature space. That is not the case here, as the data is being partitioned into subgroups with bounds imposed on the minimum and maximum sizes of each subgroup. LDA offers the best optimization at the local subspace level because the scatter matrices are likely to be well conditioned (assuming uni-modal class conditional distributions).

The maximum likelihood classifier, which estimates the label, is also expected to work properly at the subspace level. The covariance matrices are likely to be full ranked,

ensuring reliable estimation, even with little amounts of available training data. Membership function and details of ML classifier are explained previously in section 3.5.

### 3.6.3 Decision fusion

Two kinds of decision fusion mechanisms are employed in this study - majority voting and linear opinion pool.

#### 3.6.3.1 Majority Voting (MV)

Majority voting is a kind of hard decision fusion mechanism where the final classification label is assigned based on a vote over individual class labels coming out from the "local" classifiers. A simple MV decision fusion is given by:

$$\omega = \text{argmax}\, \mathrm{N}(\mathrm{i}), \quad where \quad \mathrm{N}(\mathrm{i}) = \sum_{\mathrm{j}=1}^{\mathrm{N}} \mathrm{I}(\omega_{\mathrm{j}} - \mathrm{i}) \,. \tag{3.17}$$

One advantage of MV over soft decision fusion techniques is that it is not sensitive to the inaccuracies in the estimates of posterior probabilities. The above equation assumes equal weights to all the individual subspace classifiers. An adaptive weight majority voting ensures greater priority for strong classifiers, and is given by:

$$\omega = \text{argmax}\, \mathrm{N}(\mathrm{i}), \quad where \quad \mathrm{N}(\mathrm{i}) = \sum_{\mathrm{j}=1}^{\mathrm{N}} \alpha_{\mathrm{j}} * \mathrm{I}(\omega_{\mathrm{j}} - \mathrm{i}) \,. \tag{3.18}$$

The weight assignment ($\{\alpha\}$) is based on the class separation metric, indicating the strength of that subspace/classifier (Bhattacharyya distance in this work) that each subspace provides.

#### 3.6.3.2 Linear opinion pool:

Linear opinion pool is a soft decision fusion algorithm where the global class membership function is generated using the individual posterior probabilities coming out

from each of the ML classifiers. The global class membership function is a weighted average of the 'local' class membership functions given by:

$$c(\omega_i/x) = \sum_{j=1}^{n} \alpha_j \; p_j(\omega_i/x)\,, \quad \omega_i = \text{argmax}\, c(\omega_i/x)\;. \tag{3.19}$$

Uniform classifier weights assigned to each classifier make the above equation an ordinary LOP, while non-uniform weight assignment, based on the class separation metric (Bhattacharyya distance), make it a weighted LOP.

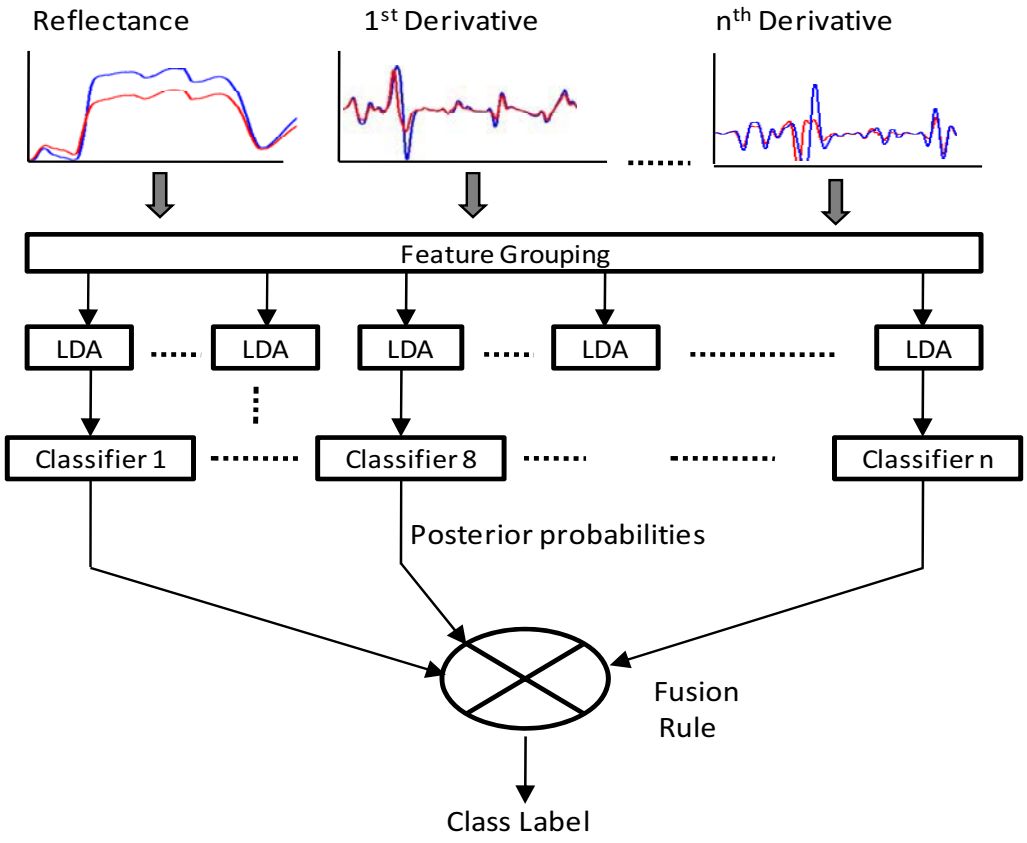


Figure 3.8 Block diagram of the MCDF classifier system used in this study.

Figure 3.8 shows the block diagram of the MCDF system used in this study. Spectral derivatives are concatenated with the reflectance values as shown in the figure. The concatenated data is then partitioned into contiguous subspaces so that each subspace possesses good class separation and the correlation between subspaces is minimized. This is achieved using the feature-grouping method described above. After the feature grouping process, LDA is employed for feature optimization within each group. A Maximum likelihood (ML) classifier per subspace/group is then used to classify the data. Decision fusion techniques, LOP and MV, are used to fuse the different class labels coming from multiple classifiers.

# CHAPTER 4

# EXPERIMENTAL SETUP AND RESULTS

In this chapter, the benefits of including spectral derivatives for classification of hyperspectral data are tested and quantified for three different hyperspectral datasets. First a set of tuning experiments are carried out for obtaining "optimal" system parameters such as derivative order, sampling order and filter order for calculating the appropriate derivative features. A "combined" analysis is then carried out, where reflectance features are combined with derivative features and fed to traditional single classifier systems, and to the proposed MCDF system for classification. Finally, sensitivity of the proposed and traditional classification approaches to the amount of training data employed for classification is presented.

## 4.1 Experimental hyperspectral data

Three different datasets are used as case studies in this thesis. Each sensor is on-board a different platform, representing data collection and acquisition in different scenarios.

### 4.1.1 Dataset 1 – handheld hyperspectral data

The first dataset consists of hyperspectral signatures collected from a corn crop treated with six different levels of herbicide concentrations, along with a part of it being left untreated. This represents a 7-class dataset representing a corn crop under varying

severity of chemical stress (ranging from no stress to severe stress). The data is collected using an Analytical Spectral Device (ASD) Fieldspec Pro handheld spectroradiometer [28], with specifications shown in Table 4.1.

An average of ten samples collected every second is recorded and stored as one sample. The sensor (aboard a tractor) is held 4 feet above the vegetation canopy using a $25^0$ instantaneous field-of-view while collecting the data. This data is collected from Brooksville, Mississippi in good weather conditions (during clear sky in summer) on $2^{nd}$ of June 2008 and is used to test the multi-class classification accuracy of the system.

The six different herbicide concentrations sprayed on the field are 1, 0.5, 0.25, 0.125, 0.0625 and 0.03125 times the standard concentration, measured in fluid ounces per area (fl. oz./a.). The data represents seven classes with one class per treatment and an additional class for no treatment. The classes are labeled as control (untreated), rate1x, rate05x, rate025x, rate0125x, rate00625x and rate003125x.

Table 4.1 Specifications of the handheld ASD sensor.

| Sensor Type | Handheld ASD Sensor |
|---|---|
| Spectral Range | 350 nm to 2500 nm. |
| Spectral Resolution | 3 nm at 700nm and 10nm at 1400/2100 nm |
| Sampling interval | 1 nm |
| Scanning time | 100 milliseconds |
| Detectors | 0ne silicon and two InGaAs photodiodes |

With high levels of noise creeping into the signatures, especially at higher wavelengths due to extended usage time of the sensor in higher temperatures, the signatures are used after truncating them to 1800 nm with the noise from the water band absorption at 1350 to 1430 nm removed using water band interpolation.

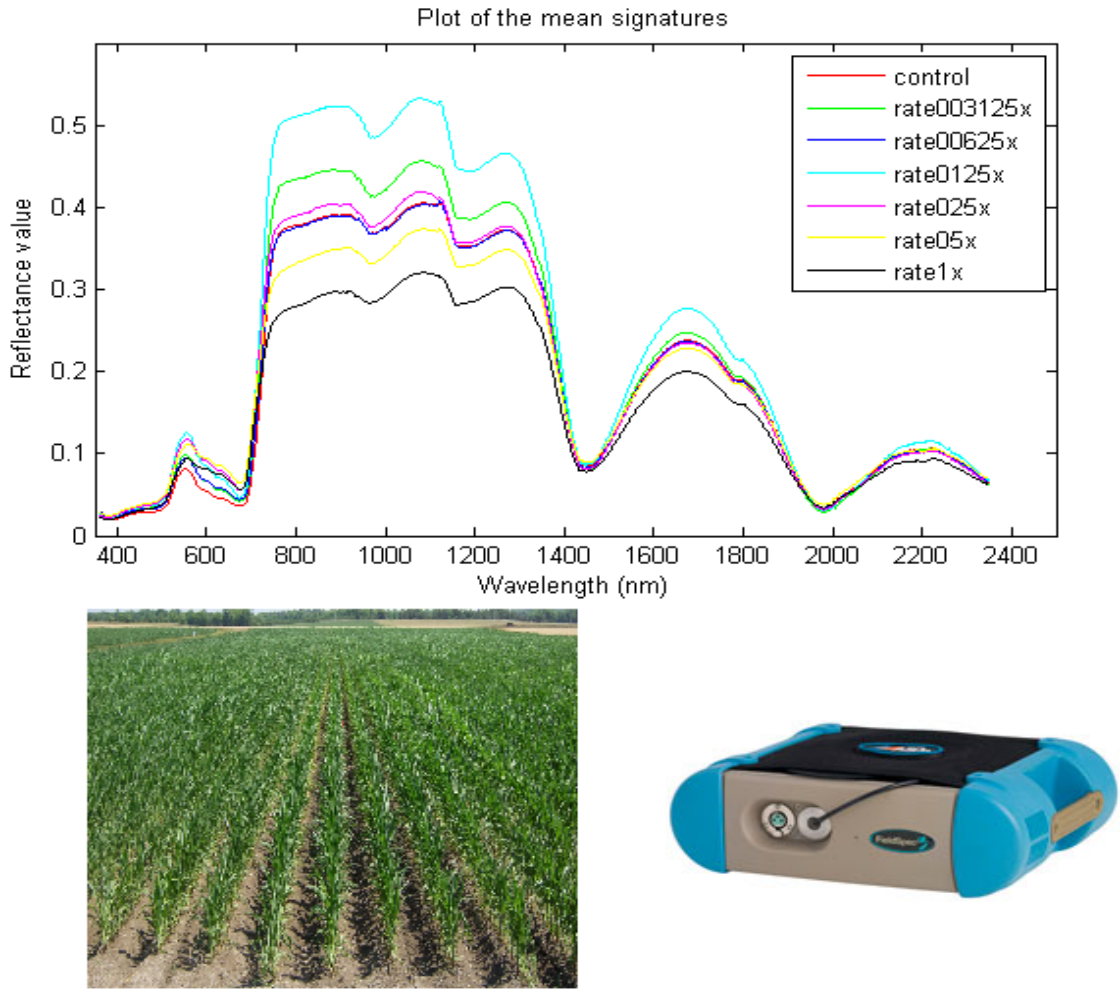


Figure 4.1 Top: Plot of mean signatures for the seven class ASD data; Bottom-Left: Photograph of the corn field in Mississippi State University experimental fields at Brooksville, MS. Bottom-Right: photograph of the ASD Feildspec Pro sensor used.

The dataset consists of approximately 180 (per class) finely resolved reflectance signatures that can be used for classification of the seven closely related classes. A two-fold cross-validation technique (Jackknifing) is employed, where the data is equally divided into two groups called training data and test data for experimental analysis on this dataset. The setup represents a rapid crop stress detection and classification task. Figure 4.1 shows a plot of mean signatures associated with the seven classes.

### 4.1.2 Dataset 2 – spaceborne hyperspectral data

The second dataset is from a spaceborne hyperspectral image acquired using the HYPERION sensor. The data is collected over some areas in Colorado, where the affect of invasive species on the native vegetation is predicted to be very high. HYPERION is a spaceborne hyperspectral sensor aboard NASA Earth-Orbiter I satellite [29]. Sensor specifications are shown in Table 4.2.

Table 4.2 Specifications of the spaceborne Hyperion sensor.

| Sensor type | Space borne Hyperion |
|---|---|
| Spectral range | 400 nm to 2500 nm. |
| Number of bands | 200 contiguous bands |
| Spectral resolution | 10 nm |
| Swath | 7.5 km |
| Spatial resolution | 30 m |

All the reflectance signatures in this dataset are grouped into two classes; 1) Tamarisk (*Tamarix ramosissima*) and 2) Non-Tamarisk (a collection of native vegetation

signatures in the vicinity, such as those of cottonwood and willow). Tamarisk, also known as salt cedar, is considered as an invasive species, which suppresses the growth of native vegetation by aggressively consuming the available water supply.

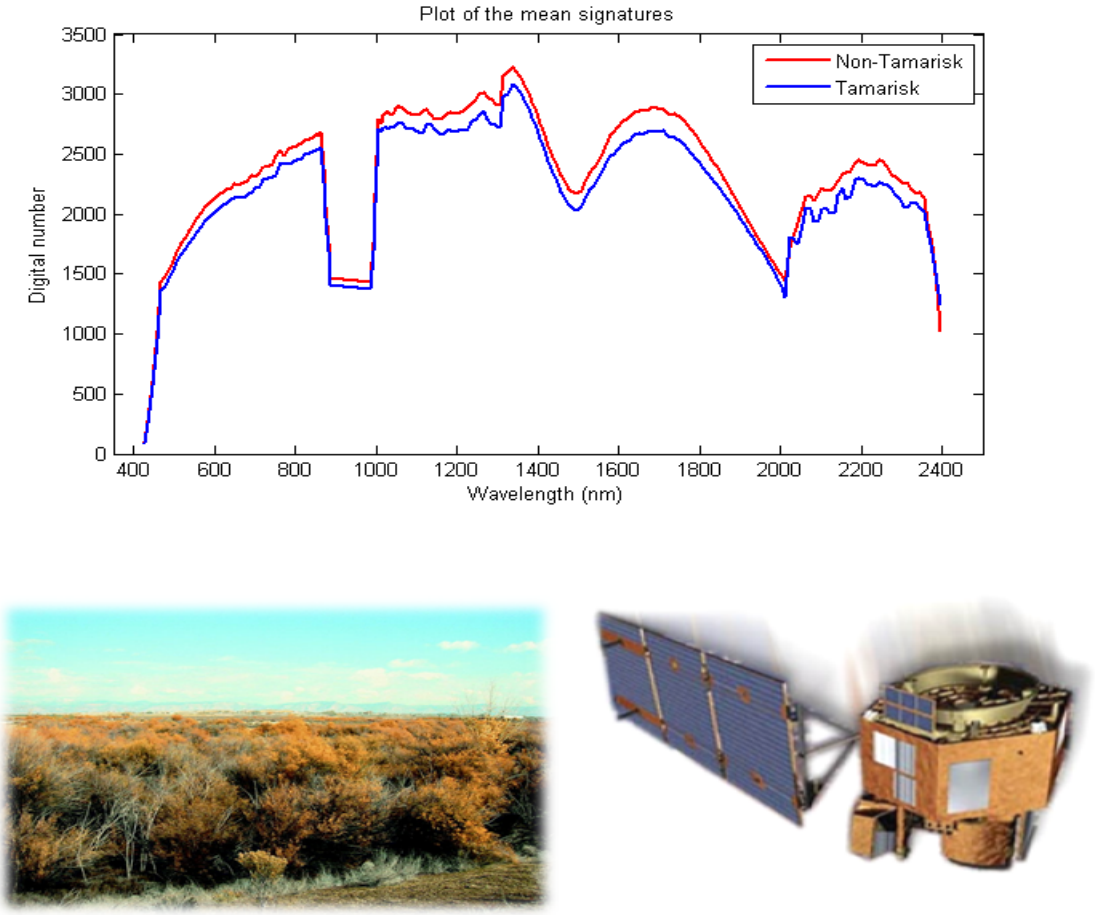


Figure 4.2 Top: Plot of mean signatures for the two class Hyperion data; Bottom-Left: Photograph of the tamarisk stand in Colorado State; Bottom-Right: Hyperion sensor aboard the EO-1 mission.

The data consists of 115 signatures of Tamarisk and 65 signatures of Non-Tamarisk. Because of limited amounts of ground truth available, an n-fold cross-validation technique, also called Leave-one-out, is used for carrying out the experimental

analysis. With the aim of recognizing Tamarisk from other species, this classification task forms a good example of a typical hyperspectral target recognition system. Figure 4.2 shows a plot of mean signatures of the two classes present in the data.

### 4.1.3 Dataset 3 – airborne hyperspectral data

The third dataset is an airborne hyperspectral image data acquired using a Pro-SpecTIR-VNIR sensor. Data is collected on 6$^{th}$ June 2008, over the same corn field where dataset 1 (using the ASD sensor) is collected. Pro-SpecTIR-VNIR is a hyperspectral sensor with specifications shown in Table 4.3 [30].

Table 4.3 Specifications of the airborne SpecTIR sensor.

| | |
|---|---|
| Sensor type | Air borne Pro-SpecTIR-VNIR |
| Spectral range | 400 nm to 994 nm. |
| Number of bands | 128 contiguous bands |
| Spectral resolution | 2.3 - 20 nm |
| Sampling intervals | 4.6 nm |
| Spatial resolution | 1 m |

The ground spatial distance (equivalent to spatial resolution) was 1m.With the help of the global positioning system (GPS) coordinates recorded while collecting the ground truth data with the ASD handheld sensor (Dataset 1), corresponding signatures from the airborne imagery are separated and grouped into “ground truth data” for the SpecTIR imagery. This collection consists of approximately 400 signatures for each class which are used as training data for classifying the entire imagery. The ground truth data

from the SpecTIR imagery is also used to perform the experimental analysis for this study. A two-fold cross-validation technique, (Jackknifing), is used for dividing the available ground-truth data into training and test data. Figure 4.3 shows a plot of mean signatures of the seven classes present in the data.

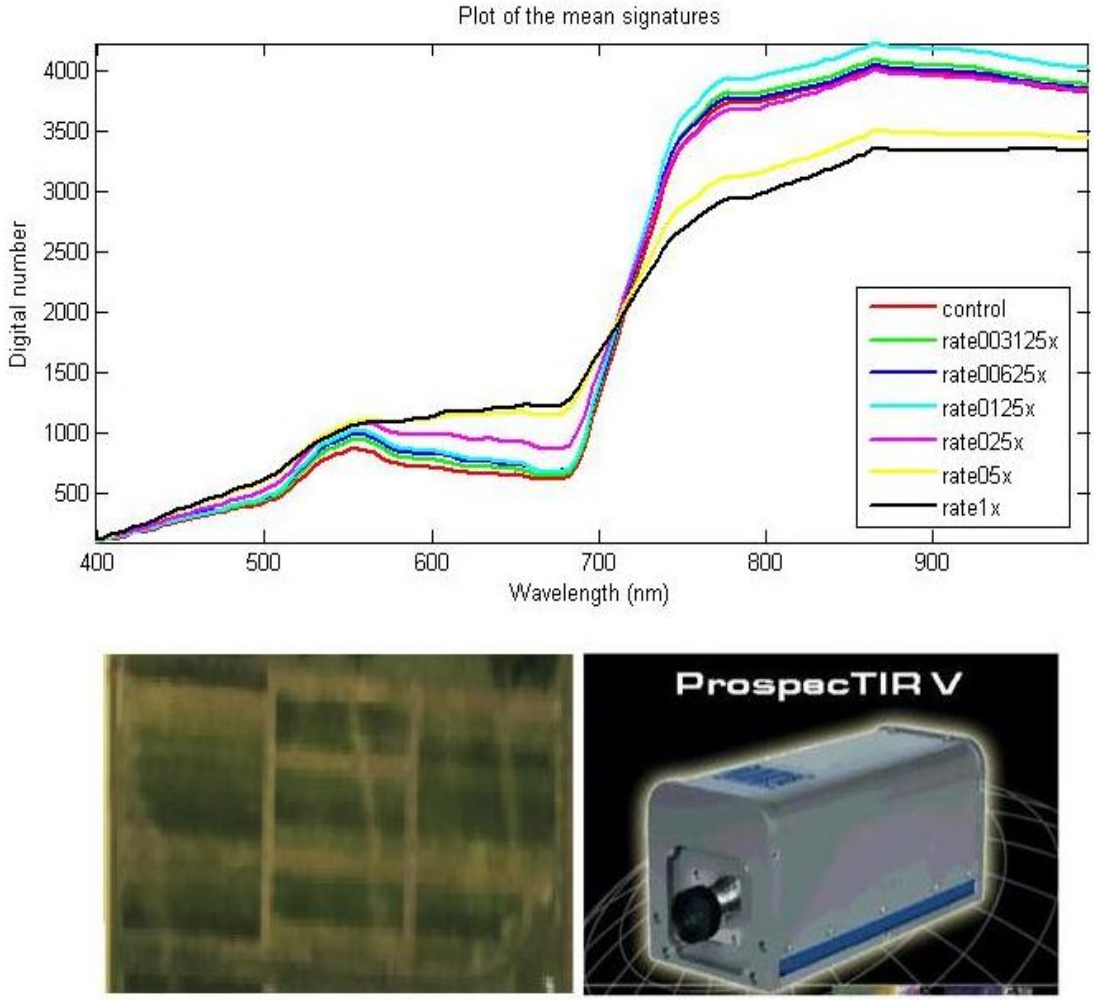


Figure 4.3: Top: Plot of mean signatures for the seven class SpecTIR data; Bottom-Left: Hyperspectral image of the corn field in Mississippi State Universitys' experimental fields at Brooksville, MS; Bottom-Right SpecTIR-VNIR sensor.

## 4.2 Preliminary experiments for parameter tuning

In this section, some tuning experiments are conducted to identify an appropriate set of system parameters that will be used for all experiments. Parameters such as derivative order 'd', filter order 'f' and the sampling order 's' (used in calculation of derivatives), that yield the best classification accuracies with the available development data are found and are used on the 'training' and 'testing' data. Development data is derived from available "training data" by further partitioning the training data into training and test data, for tuning the system parameters.

### 4.2.1 Dataset 1

ASD data having a total of approximately 1,200 signatures with 2,151 dimensions is divided into two groups called 'training' and 'testing' with 600 signatures each. The training group, with a total of 600 training samples from all the seven classes, is further divided into two groups called 'training_trn' and 'training_tst' each with 300 samples. This is the development data we use to "tune" the system. Experiments for finding the appropriate parameters (particular to this data) yielding good classification accuracies are performed by using these 300 'training_trn' samples for training the classifier and then testing it on the 'training_tst' samples. Per class mean signatures calculated for all the seven classes using the reflectance features and first five derivative features are shown in the Figure 4.4.

A set of classification accuracies (final) are determined using the SLDA-ML classifier system by varying the derivative order from 1 to 6, sampling order from 1 to 30 and filter order from 3 to 10. In this way the SLDA-ML classifier is run on 1,440 different kinds of derivative data. The classification accuracies are stored in a 6 x 30 x 8

ordered three-dimensional matrix. Now the parameter combination yielding the highest accuracy for all the six derivative orders is found by looking for the local maxima in the 3D matrix. The parameter set and the best accuracies for dataset1 are shown in Table 4.4.

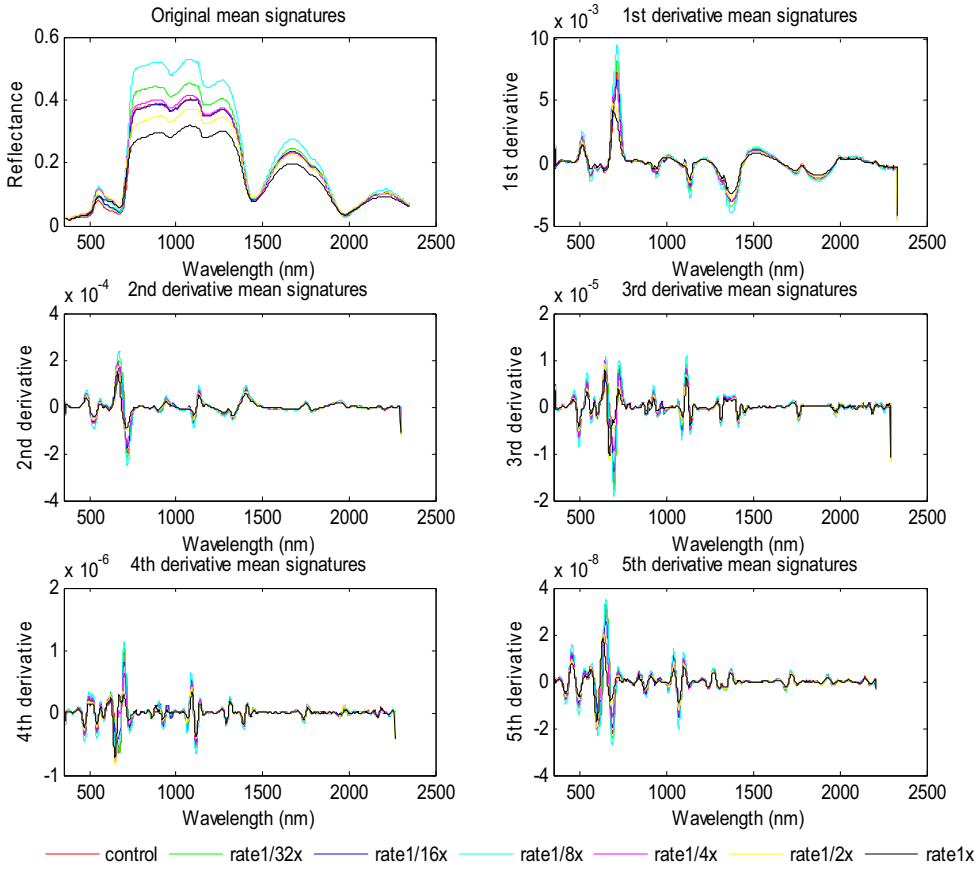


Figure 4.4: Plot of mean signatures for the handheld ASD data. Mean of the seven classes taken in reflectance domain and the derivative domain for the first 5 derivative orders are plotted in subplots.

It can be seen from the table that different derivatives yield different classification accuracies depending on the class separation provided by the individual derivative features. Figure 4.5 shows plots of class separation metric (Bhattacharyya distance),

calculated using different derivative features (different derivative orders), versus the wavelength, compared against the metric calculated using reflectance features. The plot shows a clear difference in the distances provided by the derivative features of different orders, which is reflected in the final accuracy variation.

Table 4.4 Accuracies for ASD data with different derivative orders.

| **Derivative order** | **Sampling order** | **Filter order** | **Accuracy % ± CI** |
|---|---|---|---|
| 1 | 16 | 4 | 72.0 ± 1.1 |
| 2 | 24 | 9 | 79.5 ± 1.1 |
| 3 | 18 | 7 | 81.8 ± 1.0 |
| 4 | 20 | 9 | 82.8 ± 0.9 |
| 5 | 28 | 8 | 81.1 ± 1.0 |
| 6 | 14 | 3 | 79.5 ± 1.0 |

Higher values of sampling order are expected to be better with ASD data (sampled at 1nm wavelengths), since calculating derivative at lower sampling rates means finding the slope information between bands very close on the wavelength spectrum. With increasing sampling order and filter order, the variation in the accuracies with higher order derivative features is minimal, which can be seen in the mesh plots shown in Figure 4.6.

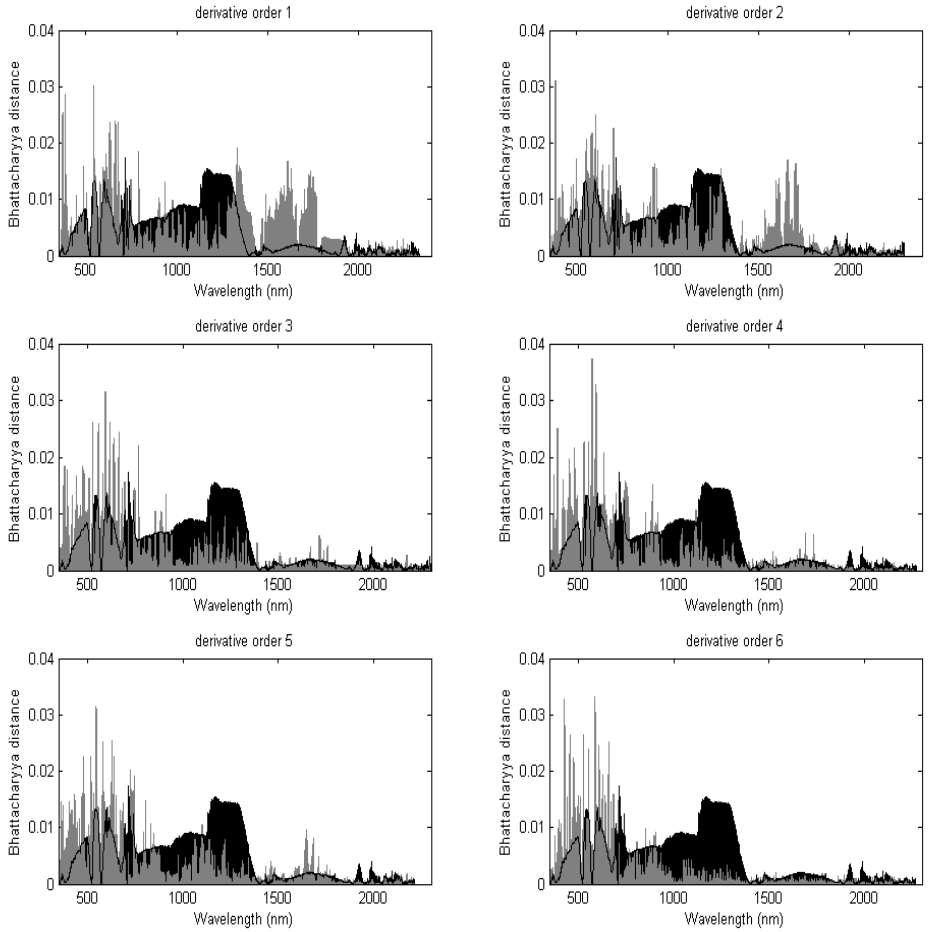


Figure 4.5: Plot of Bhattacharyya distance vs. wavelength for the first 6 derivative orders plotted against the Bhattacharyya distance with the reflectance values for ASD data.

Table 4.4, though, shows variation in the sampling orders with increasing derivative order, and this is because a local maxima was picked from the closely varying accuracies. It is also evident from the Figure 4.5 that derivative features seem to be performing better than the reflectance values in providing good class separation. The fact that SLDA selects the features based on a class separation metric, correlates with the increase in the

classification accuracies with derivative features over reflectance values, as shown in the Table 4.5. Note that for this dataset, we found MV based decision fusion to outperform LOP, and have hence used MV in all experiments reported with this dataset.

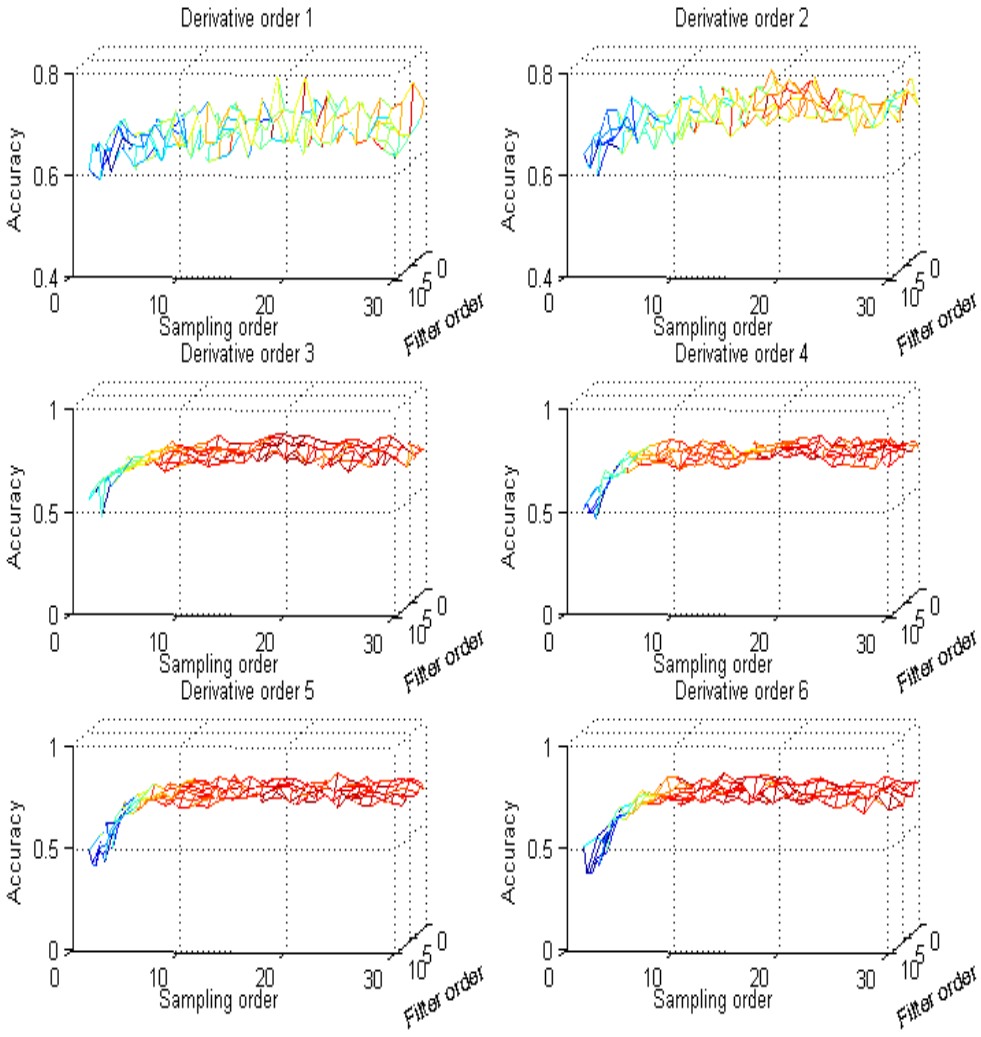


Figure 4.6: Mesh plot of accuracies with varying sampling rates and filter orders for first 6 derivatives orders for ASD data using an SLDA-ML classifier.

Table 4.5 Accuracies with full training and testing data for both the classifiers using the reflectance values and derivative features calculated using the combination found previously for each derivative order for ASD data.

| Derivative order | SLDA- ML<br>Accuracy % ± CI | MCDF<br>Accuracy % ± CI |
|---|---|---|
| Reflectance values | 66.2 ± 1.1 | 68.2 ± 1.5 |
| 1 | 66.5 ± 1.1 | 68.5 ± 1.5 |
| 2 | 71.8 ± 1.1 | 71.3 ± 1.4 |
| 3 | 74.1 ± 1.0 | 72.8 ± 1.4 |
| 4 | 77.8 ± 0.9 | 71.5 ± 1.5 |
| 5 | 72.3 ± 1.0 | 67.8 ± 1.4 |
| 6 | 72.5 ± 1.0 | 70.5 ± 1.4 |

### 4.2.2 Dataset 2

The Hyperion dataset has a total of 180 signatures of which 115 are Tamarisk and 65 are Non-Tamarisk. The available data being insufficient, relative to the dimensionality of the dataset, forces the use of an N-fold cross-validation technique (Leave-one-out) for performing experiments on this dataset. Therefore, parameter tuning for this dataset is made by considering all the available data – that is, we do not partition training data into further training and test data, and instead tune the system with the entire dataset. The classification accuracies with this dataset are hence expected to be slightly biased, but the results will nevertheless provide valuable insight into the efficacy of the proposed

approach with such data. Per class means calculated for both the classes using the reflectance features and first five derivative order features are shown in the Figure 4.7.

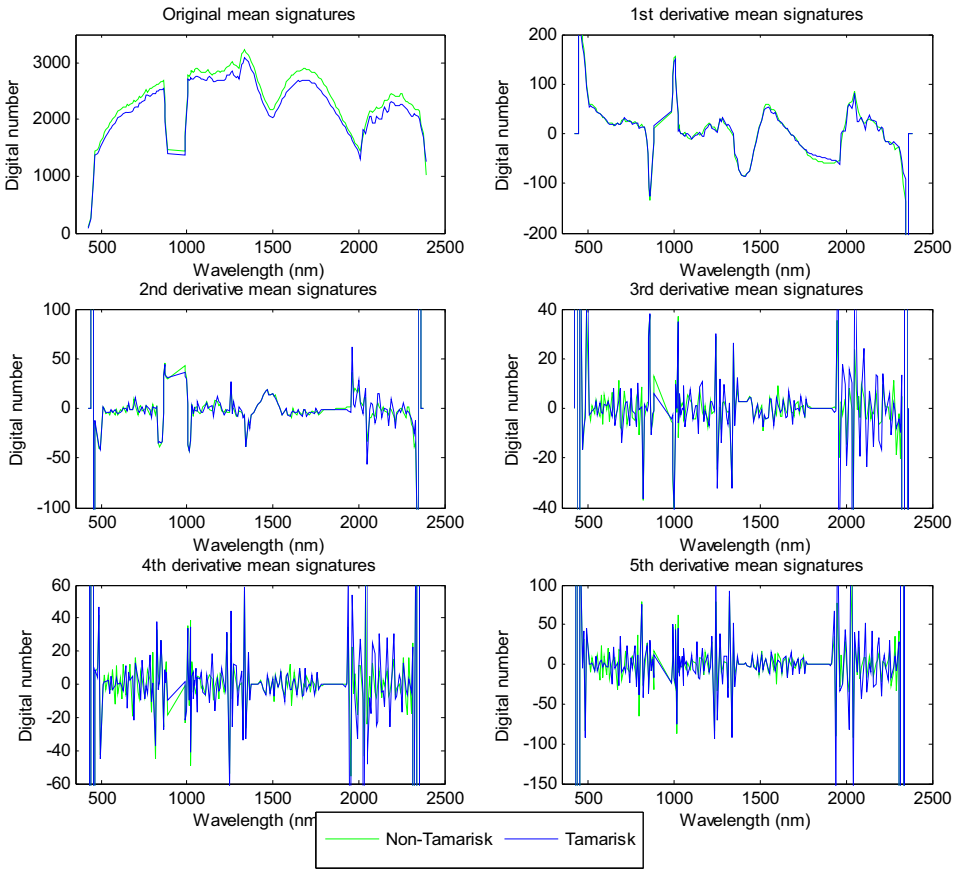


Figure 4.7: Plot of mean signatures for the spaceborne Hyperion data. Mean of the two classes taken in reflectance domain and the derivative domain for the first 5 derivative orders are plotted in subplots.

Classification accuracies (final) are determined using the SLDA-ML classifier system by varying the derivative order from 1 to 6, sampling order from 1 to 20 and filter order from 3 to 10. In this way, the SLDA-ML classifier is run on 960 different combinations of derivative data. Classification accuracies are stored in a 6 x 20 x 8

ordered three-dimensional matrix. Now the parameter combination yielding the highest accuracy for all the six derivative orders is found using the 3-D matrix. The parameter set and the best accuracies for the dataset 2 are shown in Table 4.6.

Table 4.6 Accuracies for Hyperion data with different derivative orders.

| **Derivative order** | **Sampling order** | **Filter order** | **Accuracy % ± CI** |
|---|---|---|---|
| 1 | 16 | 4 | 72.0 ± 1.1 |
| 2 | 24 | 9 | 79.5 ± 1.1 |
| 3 | 18 | 7 | 81.8 ± 1.0 |
| 4 | 20 | 9 | 82.8 ± 0.9 |
| 5 | 28 | 8 | 81.1 ± 1.0 |
| 6 | 14 | 3 | 79.5 ± 1.0 |

The fact that Hyperion data is sampled at 5nm spectral resolution, higher than that of ASD data explains the lower of sampling order for this dataset. Also, filter orders being constant across the varying derivative order indicates the invariance of derivative data to filter order. Selection of the parameters for each derivative is done based on the accuracy each combination yields, and the combinations shown in the table are the ones which gave the highest accuracies for the corresponding derivative order. The table shows an increase in the accuracies for derivative features for this data. A pattern similar to that with ASD data (dataset 1) is observed with the mesh plots drawn using the 3-D matrix for the Hyperion data. Figure 4.8 shows the bar graph of Bhattacharyya distance (a

measure of class separation) provided by each derivative feature plotted against that provided by the individual reflectance values over the available spectral range.

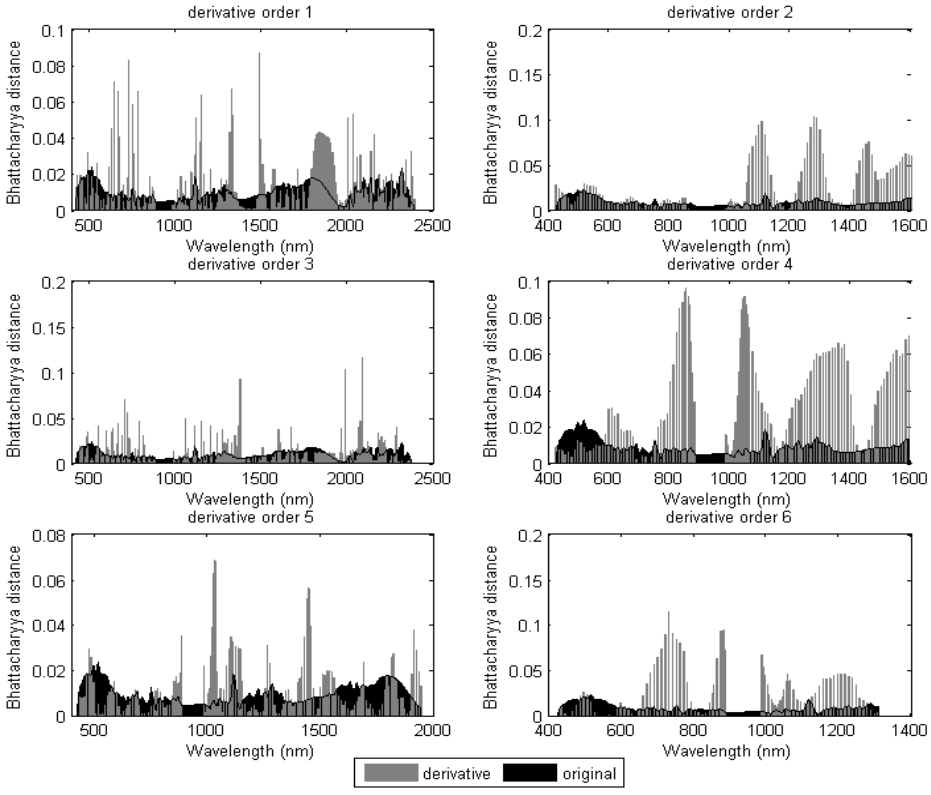


Figure 4.8: Plot of Bhattacharyya distance vs. wavelength for the first 6 derivative orders plotted against the Bhattacharyya distance with the reflectance values for Hyperion data.

The plot shows a clear domination of derivative features over the reflectance values in providing a greater class separation, and this is one reason for derivatives providing higher classification accuracies than the reflectance values. Table 4.7 shows the final accuracies with and without derivatives included into the feature space. Note

that for this dataset, we found LOP based decision fusion to outperform MV, and have hence used LOP in all experiments reported with this dataset.

Table 4.7 Accuracies with full training and testing data for both the classifiers using the reflectance values and derivative features calculated using the combination found previously for each derivative order for ASD data

| **Derivative order** | **SLDA- ML**<br><br>**Accuracy % ± CI** | **MCDF**<br><br>**Accuracy % ± CI** |
|---|---|---|
| Reflectance values | 67.8 ± 2.1 | 78.7 ± 2.3 |
| 1 | 72.8 ± 2.0 | 78.8 ± 2.0 |
| 2 | 75.0 ± 2.3 | 80.5 ± 2.4 |
| 3 | 75.0 ± 2.5 | 82.1 ± 2.3 |
| 4 | 75.6 ± 2.3 | 77.1 ± 2.5 |
| 5 | 72.8 ± 2.1 | 77.6 ± 2.5 |
| 6 | 73.9 ± 2.0 | 77.8 ± 2.6 |

### 4.2.3 Dataset 3

The SpecTIR dataset consists of 2,590 signatures of ground truth data points with 128 dimensions and is divided into 'training' and 'testing' datasets, each with 1,245 samples. As was done with the ASD dataset, the 'training' dataset is further divided into 'training_trn' and 'training_tst' datasets (our "development" data) with 622 and 623 samples in the respective groups. Tuning experiments are performed on these 'training_trn' and 'training_tst' datasets. Per class mean signatures calculated for seven classes using the reflectance features and first 5 derivative order features are shown in Figure 4.9.

Table 4.8 Accuracies for SpecTIR data with different derivative orders.

| Derivative order | Sampling order | Filter order | Accuracy % ± CI |
|---|---|---|---|
| 1 | 13 | 8 | 56.7 ± 1.1 |
| 2 | 12 | 5 | 60.3 ± 1.1 |
| 3 | 13 | 10 | 59.8 ± 1.1 |
| 4 | 12 | 10 | 55.0 ± 1.1 |
| 5 | 12 | 10 | 54.2 ± 1.1 |
| 6 | 10 | 6 | 52.8 ± 1.1 |

SLDA-ML classifier is run on different derivative datasets formed by varying the derivative order from 1 to 6, sampling order from 1 to 15 and filter order from 3 to 10. In this way the SLDA-ML classifier run on 720 different kinds of derivative data yielded a classification accuracy matrix of the order 6 x 15 x 8. Now the parameter combination

yielding the highest accuracy for all the six derivative orders is found using this 3-D matrix. The parameter set and the best accuracies for the dataset 3 are shown in Table 4.8.

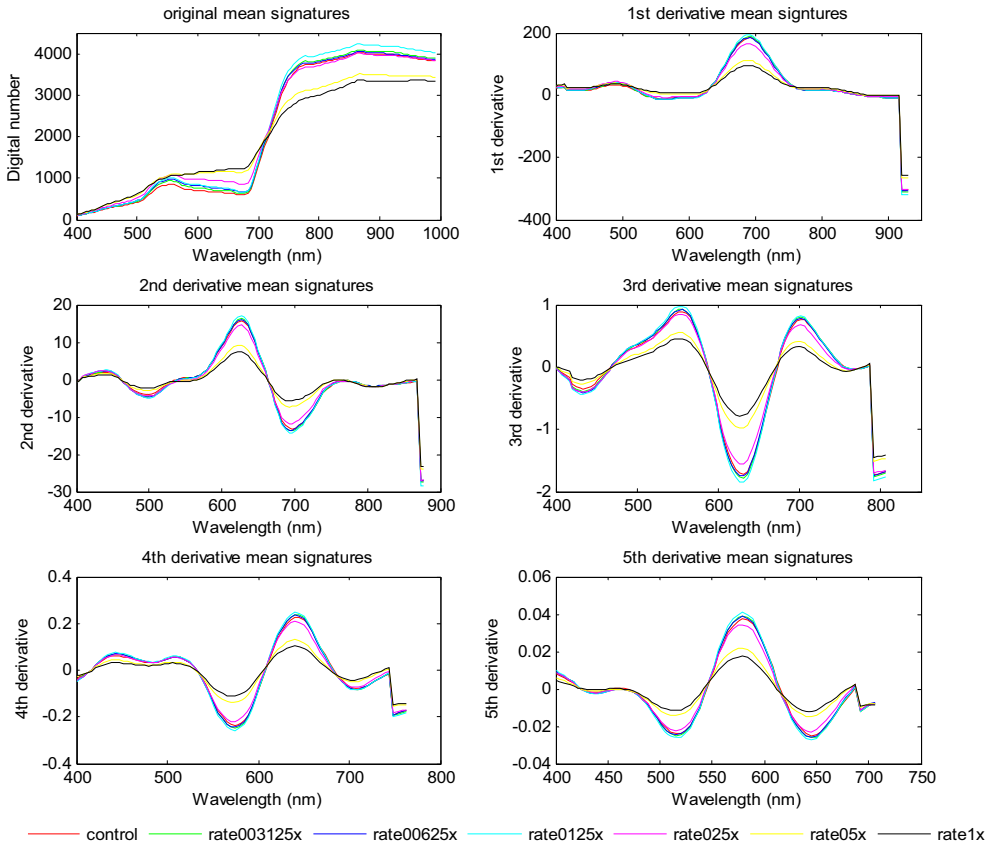


Figure 4.9: Plot of mean signatures for the handheld SpecTIR data. Mean of the seven classes taken in reflectance domain and the derivative domain for the first 5 derivative orders are plotted in subplots.

Higher band sampling for the SpecTIR data (4.6 nm) accounts for the lower values of "optimal" sampling orders and filter orders. Combinations of parameters shown in Table 4.8 for different derivative orders are the ones that yielded highest accuracy

when compared to all possible combinations of these parameters. Mesh plots plotted using all the parameter combinations, showed a pattern similar to that of the previous two datasets (ASD and Hyperion). Bhattacharyya distance graphs for different derivative orders plotted against the distances calculated using the reflectance values at the same wavelengths are shown in Figure 4.10.

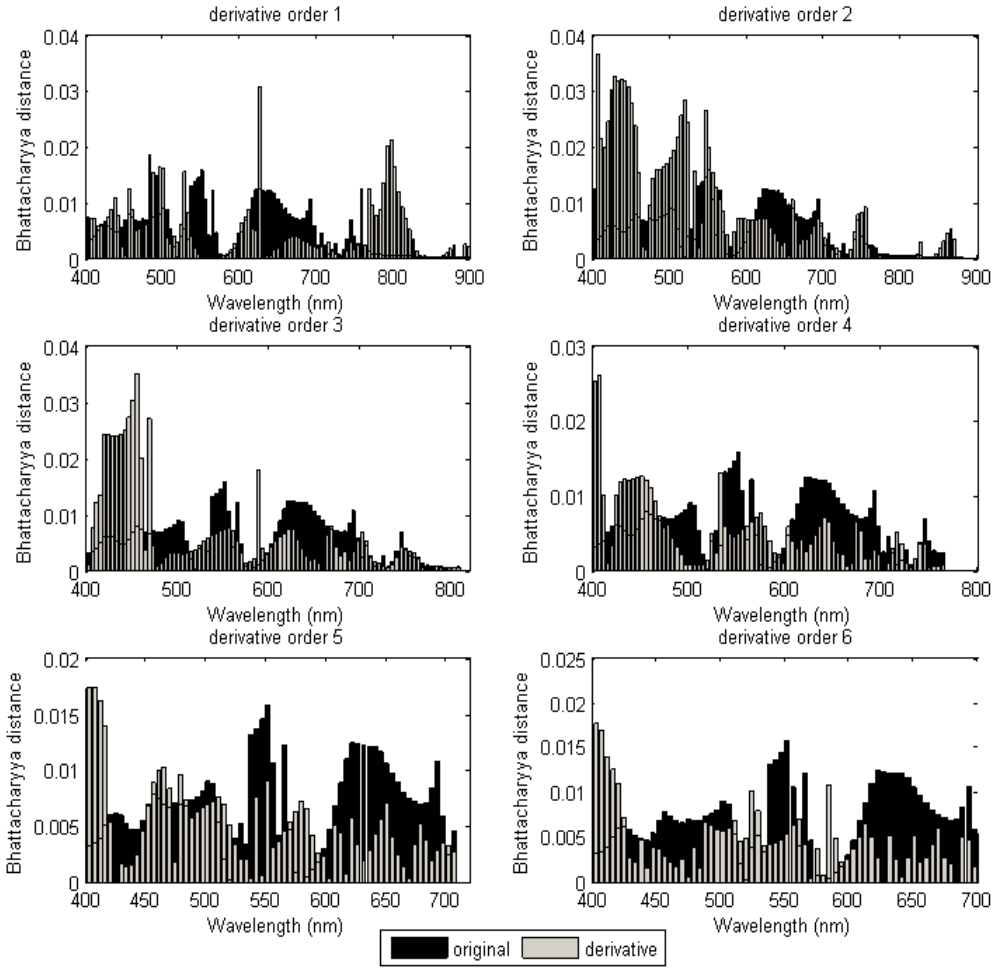


Figure 4.10: Plot of Bhattacharyya distance vs. wavelength for the first 6 derivative orders plotted against the Bhattacharyya distance with the reflectance values for SpecTIR data.

It can be seen from the plot that there is a clear increase in the distance values with derivative features when compared to the ones provided by the reflectance values, which explains the increase in the final classification accuracies with derivative features over reflectance features as shown in Table 4.9. Note that for this dataset, we found LOP based decision fusion to outperform MV, and have hence used LOP in all experiments reported with this dataset.

Table 4.9 Accuracies with full training and testing data for both the classifiers using the reflectance values and derivative features calculated using the combination found previously for each derivative order for SpecTIR data.

| **Derivative order** | **SLDA- ML**<br>**Accuracy % ± CI** | **MCDF**<br>**Accuracy % ± CI** |
|---|---|---|
| Reflectance values | 58.9 ± 0.7 | 63.1 ± 0.7 |
| 1 | 62.5 ± 0.7 | 64.2 ± 0.7 |
| 2 | 61.3 ± 0.7 | 62.8 ± 0.7 |
| 3 | 62.8 ± 0.7 | 62.6 ± 0.7 |
| 4 | 60.5 ± 0.7 | 63.3 ± 0.7 |
| 5 | 58.9 ± 0.7 | 63.5 ± 0.7 |
| 6 | 61.9 ± 0.7 | 63.1 ± 0.7 |

### 4.3 Classification experiments after parameter tuning:

Derivative features calculated using the parameter combinations identified from the tuning experiments are combined with the existing feature space (reflectance values) simultaneously (one after the other) to analyze the benefits of combining derivatives with the reflectance values using the available classifier systems.

#### 4.3.1 Dataset 1

Figure 4.11 shows the accuracies obtained using the SLDA-ML classifier and the MCDF classifier.

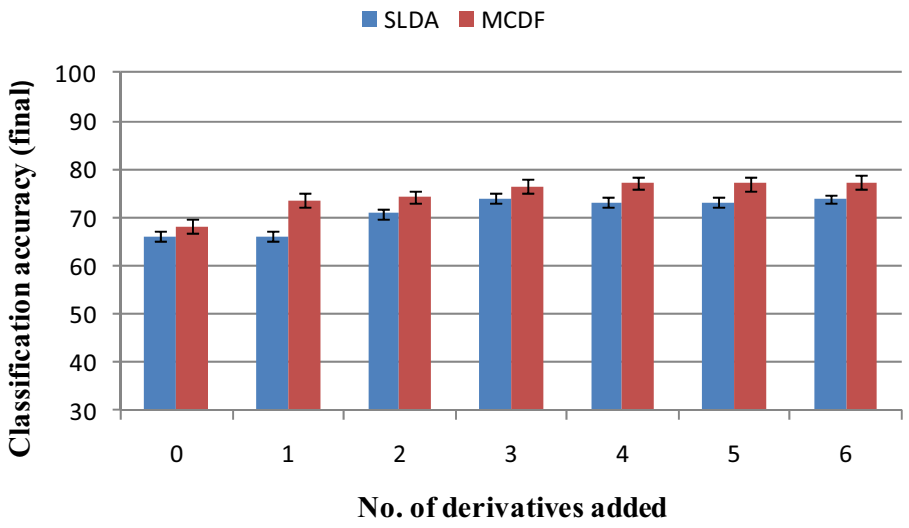


Figure 4.11: Bar graph comparing the accuracies, in percentage, for SLDA-ML and MCDF classifier systems, with the addition of each successive higher order derivative into the feature space for ASD data.

It can be seen from the graph that inclusion of additional derivatives features into the feature space has a positive effect on the overall classification accuracy with both the

classifier systems. One point that can be noted from the graph is the invariance of SLDA-ML accuracies to further addition of a derivative features after a certain number being added which actually highlights the drawback of SLDA – its inability to utilize all the features available (restricted to select a small subset of all available features only). On the other hand, the MCDF system shows a steady increase, hence handling the over-dimensionality problem from the additional derivatives more effectively, which is the key advantage of the system.

### 4.3.2 Dataset 2

Figure 4.12 shows the accuracies obtained using the SLDA-ML classifier and the MCDF classifier.

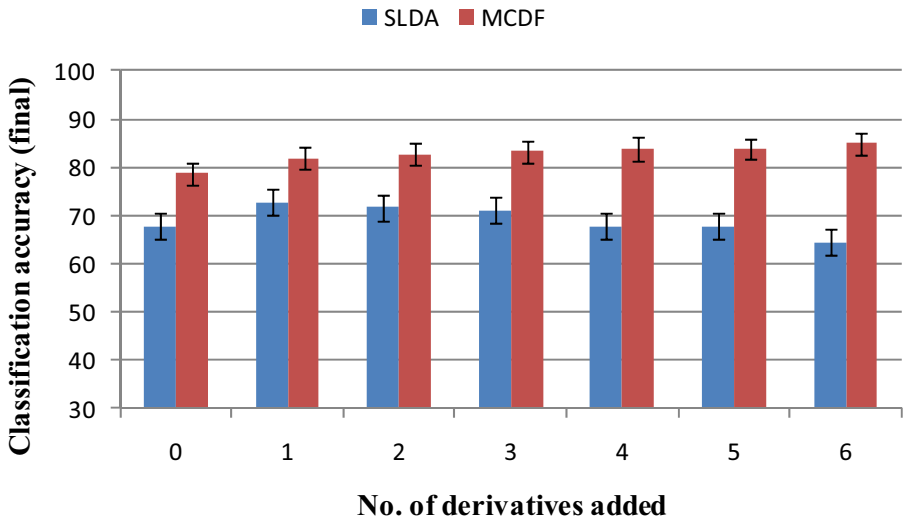


Figure 4.12: Bar graph comparing the accuracies, in percentage, for SLDA-ML and MCDF classifier systems, with the addition of each successive higher order derivative into the feature space for Hyperion data.

Both the systems show a similar pattern discussed in the previous section. The overall accuracy with the SLDA-ML classifier system is invariant to the additional derivatives sent into the system, instead there is a slight drop in the accuracy due to the over dimensionality of the input data. The drop is because of the insufficient training data available when compared to that available with the ASD data. The MCDF framework, which can work well in small-sample-size conditions, exhibits a steady increase in the accuracy.

### 4.3.3 Dataset 3

Figure 4.13 shows the shows the accuracies obtained using the SLDA-ML classifier and the MCDF classifier.

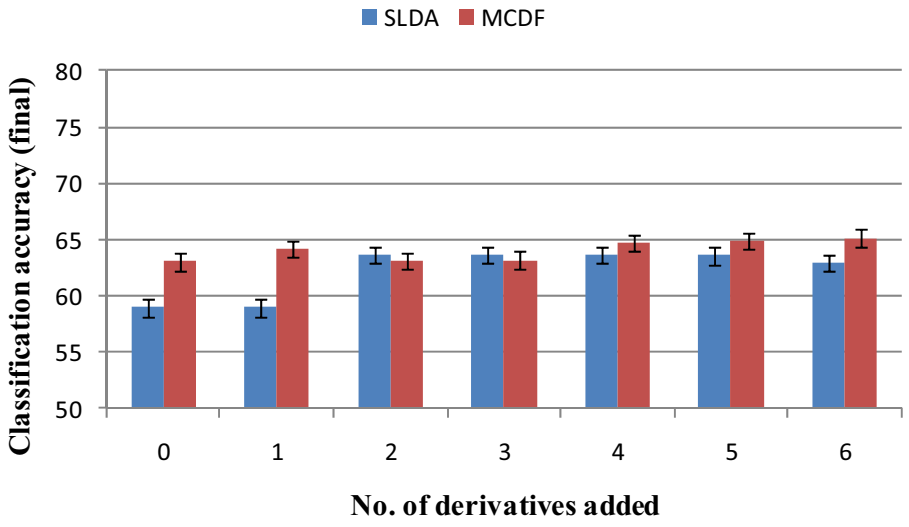


Figure 4.13: Bar graph comparing the accuracies, in percentage, for SLDA-ML and MCDF classifier systems, with the addition of each successive higher order derivative into the feature space for SpecTIR data.

As before, the MCDF system outperforms the SLDA system, and the overall classification with derivative features included is higher than that obtained with just the reflectance features.

## 4.4 Stress classification maps

In this section, the hyperspectral imagery for the corn field acquired by SpecTIR™ for Mississippi State University is employed to come up with a ground-cover classification map, indicating the variation in chemical stress on the corn crop, by using the combination of reflectance values and derivative features that gave the best classification accuracy on the ground truth data (performed in the previous section). Here, the available ground truth data is used to train the classifier and is then used to label every pixel present in the imagery. Results accomplished using the SLDA-ML classifier and MCDF classifier systems by considering previously mentioned feature space is compared with that achieved using just reflectance values as feature space. Table 4.10 shows the original spray rate distribution map of the corn field.

Table 4.10 Original spray rate distribution map of the corn field.

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

Figure 4.14 shows the classification maps obtained using the SLDA-ML classifier system. Original optical image [top] is also shown along with the maps achieved with [bottom] and without [middle] using the derivative features. Similarly the classification map obtained using the MCDF classifier system is shown in Figure 4.15.

It can be seen from figure 4.14 that including derivative features into the feature stream resulted in superior performance compared to reflectance features, resulting in a sharper separation between different spray rates (especially between the rows sprayed with relatively closer herbicide concentrations). Regions of the image labeled 1and 2 on both the maps (middle and bottom) highlight this:

Region 1) Rows of corn sprayed with rates ½ and ¼ are segregated well in the bottom image.

Region 2) The salt and pepper noise (in the row of corn sprayed with rate ½ at the edge of the field) due to misclassification is reduced in the bottom image.

The same pattern is evident when an MCDF classifier system is employed, as shown in Figure 4.15. Regions of the image labeled 1, 2 and 3 on both the maps (middle and bottom) highlight this:

1) Misclassification noise (spray rate ½ being mislabeled as spray rate 1) is significantly reduced with derivatives features.

2) Rows of corn sprayed with rates 1 and ½ are well defined (demarked) in the image obtained using the derivatives features.

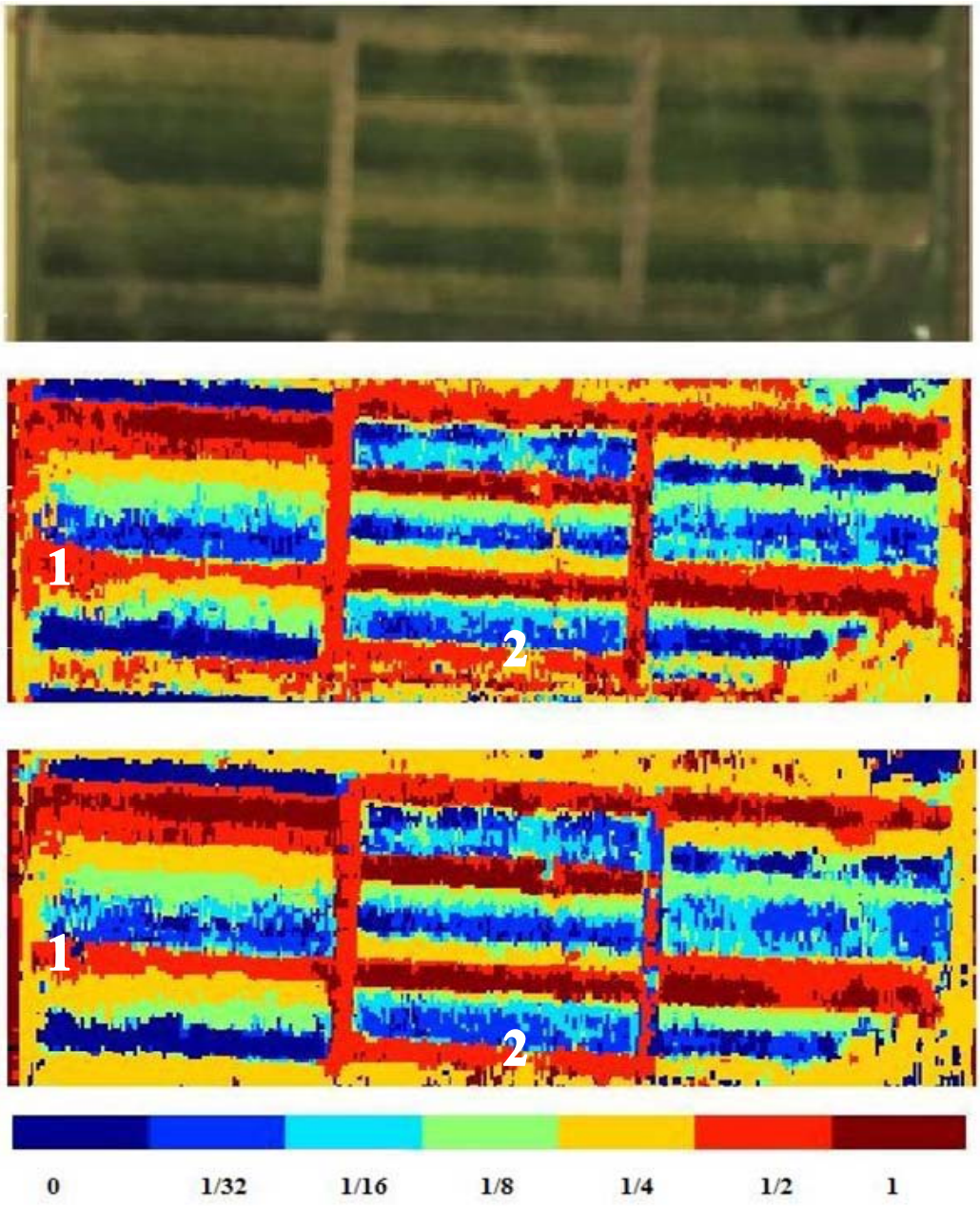


Figure 4.14: [Top]: Original RGB colored map of the corn field taken using the SpecTIR sensor. [Middle]: Classification map using the original reflectance features with SLDA-ML classifier. [Bottom]: Classification map using the derivative features with SLDA-ML classifier.

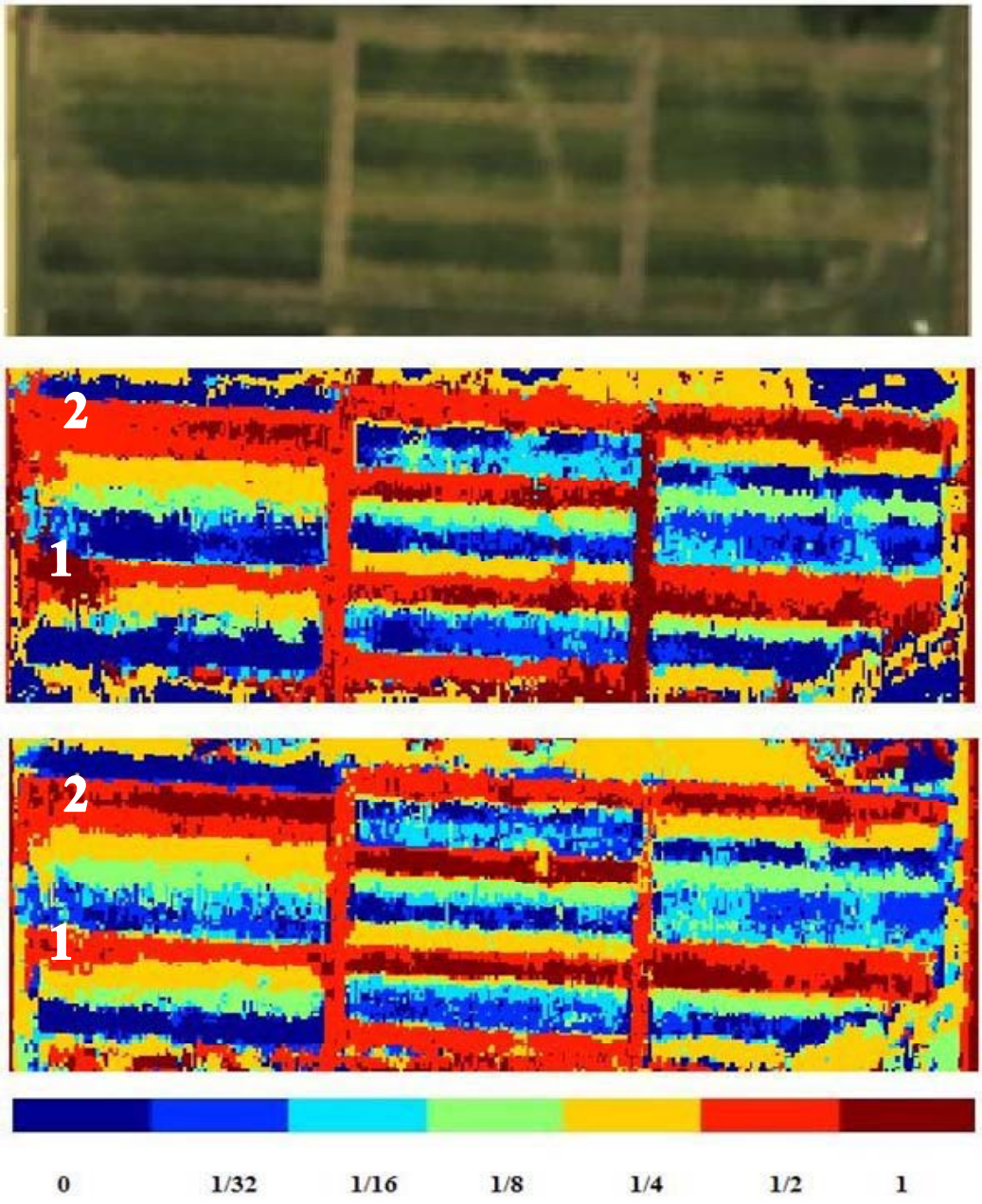


Figure 4.15: [Top]: Original RGB colored map of the corn field taken using the SpecTIR sensor. [Middle]: Classification map using the original reflectance features with MCDF classifier. [Bottom]: Classification map using the derivative features with MCDF classifier.

## 4.5 Sensitivity analysis

Through all of these experiments, the inclusion of derivatives into the feature space is proven to be very effective in improving the classification accuracies of the hyperspectral data. In this section, sensitivity of these classification systems to variations in the amount of data employed for training the system is studied with derivatives as the feature space and is compared with when reflectance values are considered as the feature space. Dataset 2 is not considered for this experiment as the data available is insufficient for performing this study.

Figure 4.16 and Figure 4.17 show a line plot of the accuracies obtained with and without derivatives included into the feature space for dataset 1 and dataset 3 respectively. The plot on the top shows the accuracies obtained using the SLDA-ML system and the bottom plot shows accuracies with the MCDF system. Both the classifier systems perform better with derivatives included. The SLDA-ML classifier with derivatives at lower percentages of training data is no better than the one without derivatives. MCDF on the other hand performs well at all times being somewhat independent of the training data availability. These observations can be made with both datasets.

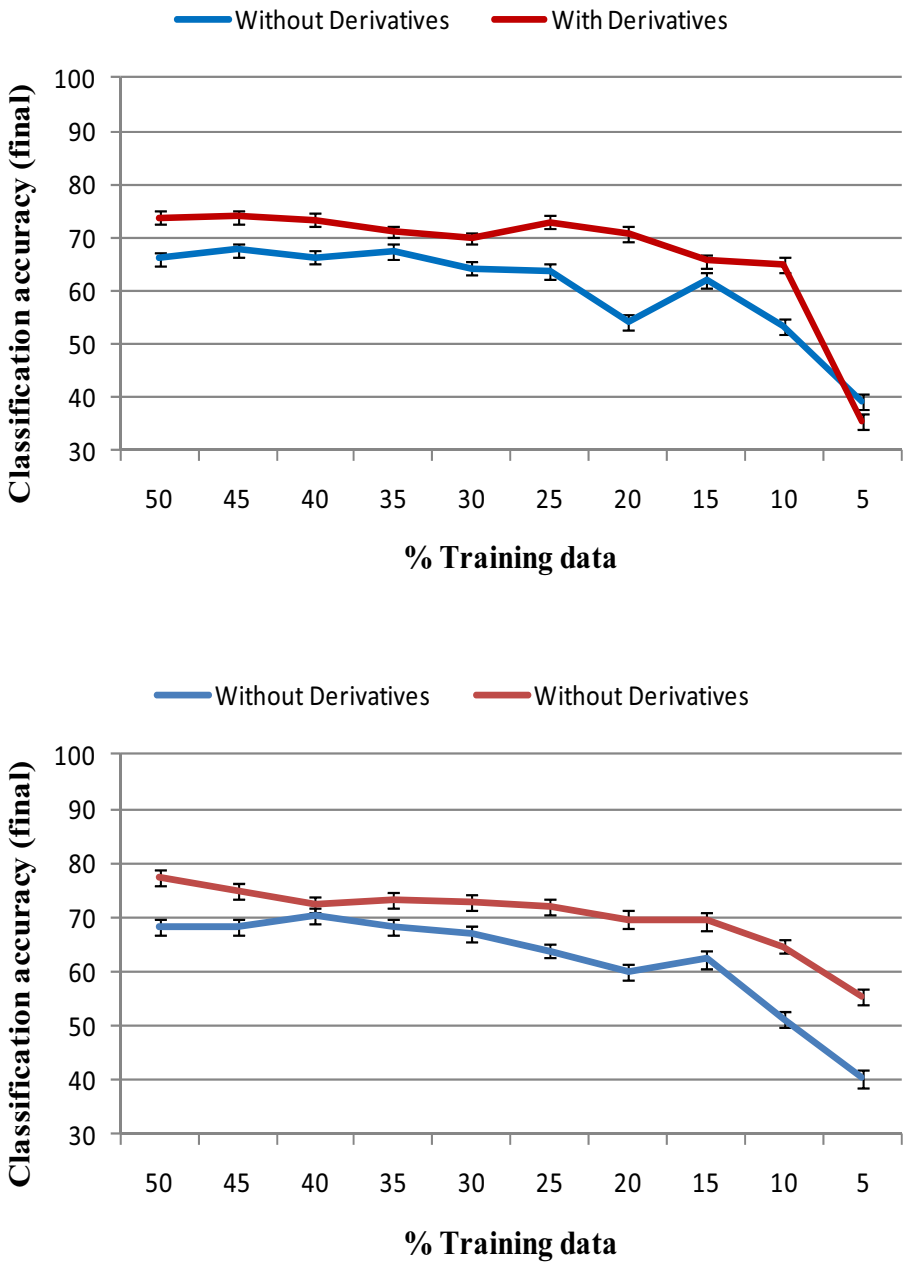


Figure 4.16: Final accuracy, in percentage, for dataset 1 (ASD data) using SLDA [Top] and MCDF [Bottom] for varying training sizes.

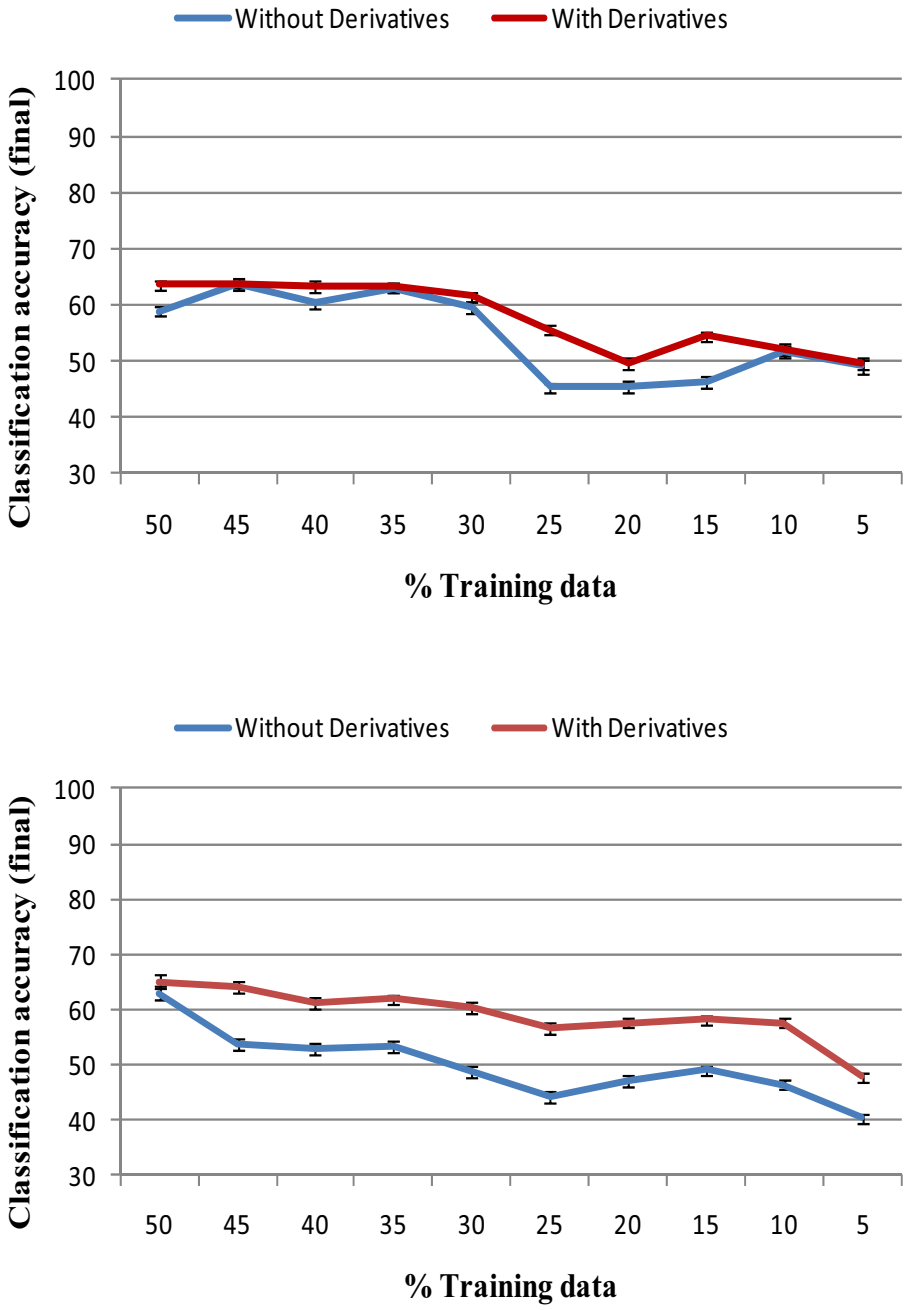


Figure 4.17: Final accuracy, in percentage, for dataset 3 (SpecTIR data) using SLDA [Top] and MCDF [Bottom] for varying training sizes.

## 4.6 Adaptive classifier

In this section, a new technique is proposed for further improving the overall classification accuracy of a multi-class classifier system. In this extension, individual class producer accuracies in the confusion matrix are used for adaptively including features that best separate the classes that are most confused. The efficacy of this adaptive classifier on the overall classification accuracy is studied over the normal classification system.

### 4.6.1 Functioning of the Adaptive classifier:

To start with, the adaptive multi classifier system takes in the training and test datasets, and the training data is further divided into two sets called 'training_trn' and 'training_tst'. Now the classifier is trained on the 'training_trn' data, which is then used to label the 'training_tst' dataset. A confusion matrix is computed based on the labeling given to the 'training_tst' data by the classifier. Table 4.10 shows a typical confusion matrix for a five class problem.

Table 4.11 Typical confusion matrix for a five class problem.

| Class Name | Class1 | Class2 | Class3 | Class4 | Class5 | Producer accuracy |
|---|---|---|---|---|---|---|
| **Class1** | 127 | 42 | 2 | 0 | 19 | 66.84 |
| **Class2** | 116 | 32 | 9 | 3 | 5 | 19.39 |
| **Class3** | 76 | 38 | 19 | 27 | 0 | 11.88 |
| **Class4** | 43 | 3 | 14 | 117 | 16 | 60.62 |
| **Class5** | 0 | 1 | 6 | 0 | 177 | 96.20 |
| **User Accuracy** | 35.08 | 27.59 | 38.0 | 79.59 | 81.57 | **52.91** |

It can be seen from the confusion matrix that even though the overall accuracy is 53 percent this accuracy is largely contributed by the classification accuracies from Class1, Class4 and Class5. The misclassification rates in Class2 and Class3 are high, and their classification accuracies are not in harmony with the overall accuracy. In such cases, the adaptive algorithm searches for such sources of 'confusion' in the confusion matrix by comparing the overall accuracy with the individual producer accuracies. In this case, the algorithm finds two instances of severe confusion, which are Class2 and Class3.

In the next step, the algorithm searches for the class names that confuse the previously found confused classes (with low producer accuracies). This is done by examining the total number of pixels from the confused classes, which are classified or labeled into other classes. In this case, the algorithm finds Class1 as the confusing class for Class2 and for Class3. After finding the confused and confusing class name pairs, the algorithm now finds the features that best separate these class pairs (excluding the ones selected previously by the optimizer) and includes them into the optimal feature space selected by the feature optimizer present in the classifier system. At this point the algorithm tests the classifier trained on the new optimal feature space with the available 'training_tst' data and also the 'testing' data and generates new confusion matrices. All the above described processes continue until all the producer accuracies get into harmony with the overall accuracy or until the optimal feature space becomes too large for the classifier to produce a reliable classification.

A flowchart explaining the procedure is shown in Figure 4.18 for the two classifier systems employed in this thesis.

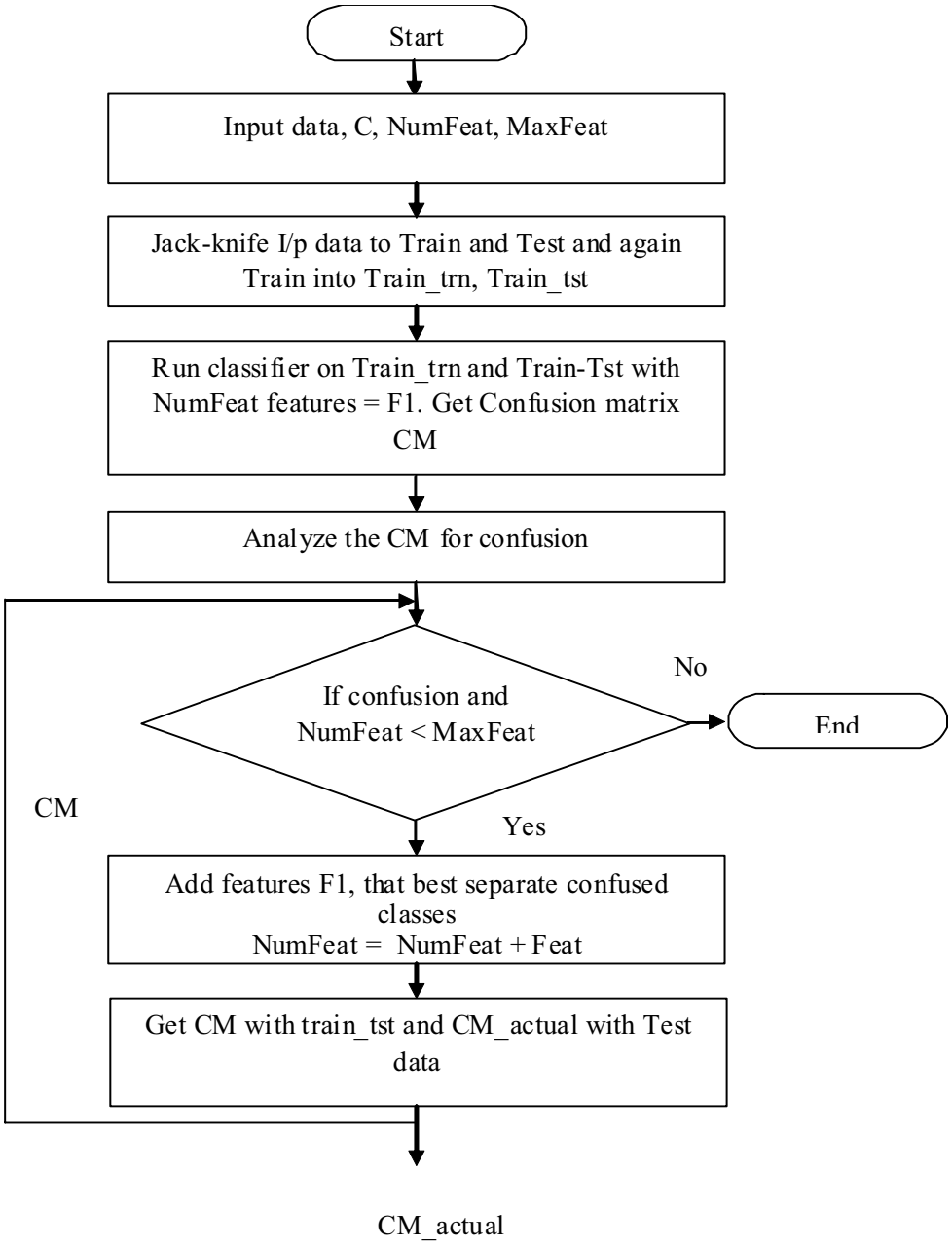


Figure 4.18: Flowchart of the operations of an adaptive classifier system used in this study.

The flowchart can be used with the SLDA-ML classifier and the MCDF classifier, with a small change in the step where the additional features picked up by this algorithm,

denoted by the feature vector F1, instead of being added to the previous optimal feature vector as was done in SLDA, are added to every subgroup that comes out of the subspace identification (band-grouping) process of the MCDF framework.

### 4.6.2 Results and discussion:

The proposed technique is implemented with the SLDA-ML and the MCDF classifiers. The efficacy of the adaptive system is then tested on two datasets, 1) ASD corn data (Dataset 1) and 2) SpecTIR corn data (Dataset 3). Figure 4.19 shows the bar plot of classification accuracies achieved using the adaptive classifier compared with that obtained using the normal classification process for both the datasets and both the classifier systems with and without using the derivative features. In all cases, the adaptive classification process outperforms the normal classification approach.

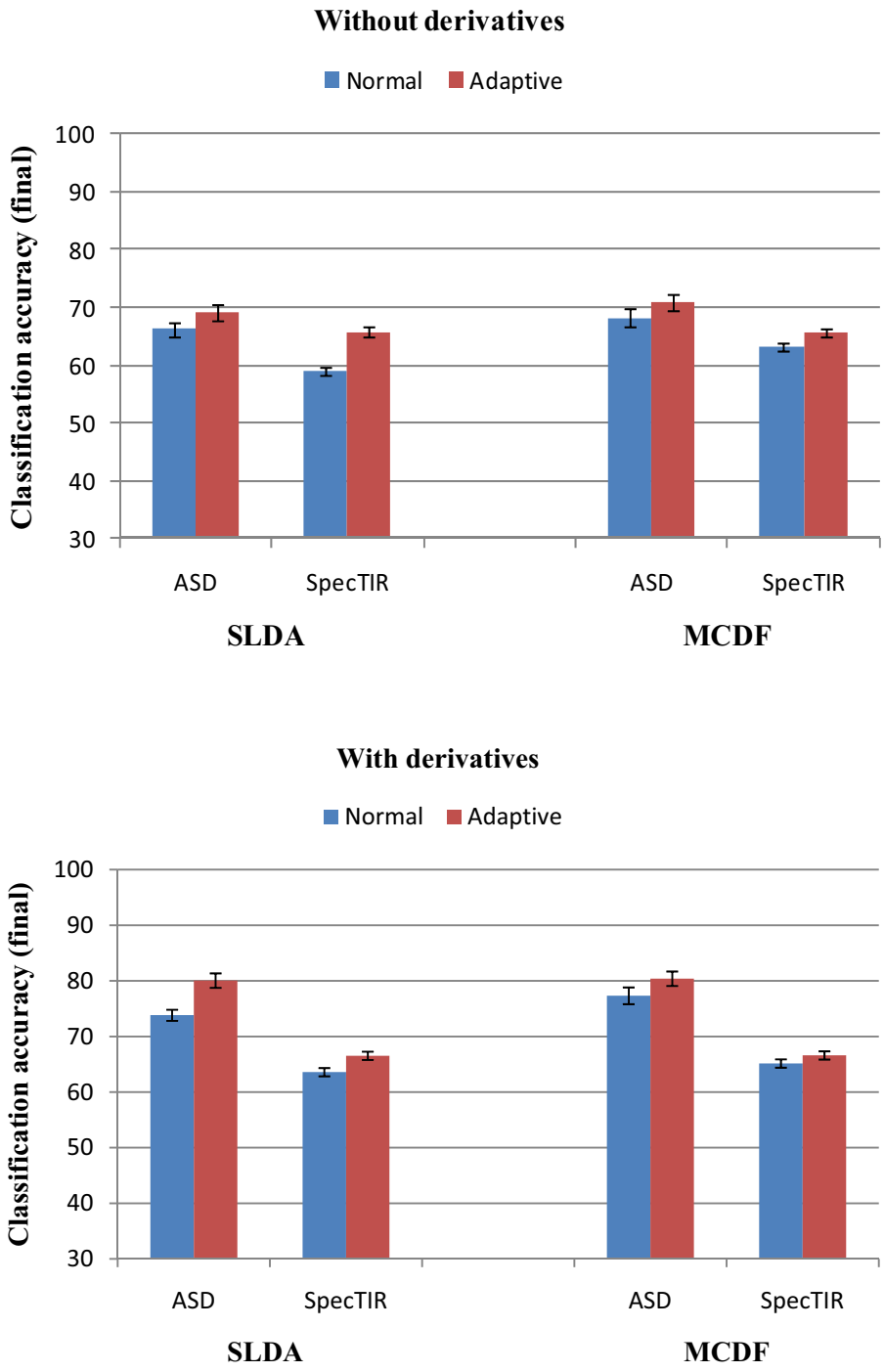


Figure 4.19: Final accuracy charts for dataset 1 and 3 using SLDA and MCDF using the normal classification process and adaptive process with and without using derivatives.

# CHAPTER 5

# CONCLUSION AND FUTURE WORK

## 5.1 Conclusions

In this thesis, the effects of including the spectral derivative information into feature space of hyperspectral data were investigated. The study was conducted on three different kinds of experimental hyperspectral data sets collected using different sensor systems. The derivative information is extracted from the reflectance information measured in different spectral bands that are specific to each sensor. For each data, derivative orders up to six are used. Class separation efficacy using the derivative features was found to be better than that achieved using the reflectance features. The performance of the spectral derivatives in providing higher class separation was found to be dependent on the derivative order chosen which in turn depended on the sampling order and the filter order (used to filter the hyperspectral data). The parameter combination defined by derivative order, sampling order and filter order are first obtained for all the three data sets by performing some preliminary experiments (parameter tuning) using the development data and are then used for the further study. An SLDA based feature selection/reduction algorithm followed by a ML classifier is used to perform the preliminary experiments.

After finding the parameter combinations for all the three data sets, the performance of spectral derivatives is then tested using two classifier systems, they are: 1) SLDA-ML classifier and 2) MCDF classifier system. The testing is carried out in three phases. 1) The classifiers trained using the development data are tested on testing data for every derivative order separately. 2) The features with increasing order of derivative combined with the reflectance features are then used for training and testing the data. 3) The sensitivity of the classifier systems to the amounts of training data is then tested using the feature space yielding the highest final accuracy (from the previous experiments) and is compared to that obtained by considering only reflectance values as features. An adaptive classifier system for improved classification accuracies is introduced and its benefits are studied.

From the experimental results, the following conclusions are made. In the first set of experiments, where the efficacy of each derivative order is tested using both single and multiple classifier systems, the performance of both the classifiers is better when using derivative features instead of reflectance features. This is the case with every derivative order considered in this study and for all the three data sets. In the second set of experiments, where the feature set formed by adding derivative features with increasing derivative order to the reflectance features, the SLDA-ML classifier showed an increase in the classification accuracy for the first few additions of higher order derivative information, which then either remained constant or started to decline with the further additions. This can be attributed to the fact that SLDA provides a sub-optimal dimensionality reduction to alleviate the high dimensional problem, and actually discards a majority of potentially useful features. The MCDF classifier on the other hand exhibits

a steady and consistent increase in the classification accuracy with the addition of derivative features into the feature space. This is the case with all the three data sets. In the third set of experiments, sensitivity of both the classifier systems to the variations in the amount of data available for training the classifier is studied. The SLDA-ML classifier system with spectral derivatives as features is seen to be performing better compared to when reflectance values are used. For the extreme case where the training data availability is 5%, SLDA-ML with derivatives is no better than without derivatives. MCDF classifier with spectral derivatives outperforms the one without derivatives in all the cases, which proves the ability of MCDF classifier to be somewhat invariant to the availability of training data.

Finally, the adaptive classification algorithm, which seeks to find additional features adaptively based on the confusion matrices calculated from the development data performed better than the previous classifier methods. The adaptive versions of the SLDA-ML and MCDF classifiers outperformed the standard implementations. This was observed with and without derivatives features included in the feature set.

## 5.2 Future work

In this work, Gaussian class distributions were assumed for representing and classifying features. It is expected that nonlinear analysis methods, such as support vector machines (SVMs) [31] will further improve classification and target recognition performance. Such classification paradigms can model more complex decision boundaries, and are hence expected to provide further robustness under severe operating conditions. Finally, it would be interesting to study the efficacy of spectral derivatives and their ratios to alleviate problems arising due to illumination variations and

atmospheric distortion. Previous pilot studies indicate an improvement in classification performance when using spectral derivative ratio features under such conditions, and it is likely that the proposed classification methods will outperform conventional methods under such scenarios.

REFERENCES


[1]. John R. Jensen, “Remote Sensing of the Environment an Earth Resource Perspective,” 2nd edition, A Prentice Hall publication, 2007.

[2]. S. Prasad and L. M. Bruce, "Overcoming the Small Sample Size Problem in Hyperspectral Classification and Detection Tasks," in *IEEE International Geosci. and Remote Sens. Symposium*, pp. V - 381-V - 384, 2008.

[3]. G.F. Hughes, “On the Mean Accuracy of Statistical Pattern Recognizers,” *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 55-63, Jan. 1968.

[4]. K. Fukunaga, “Introduction to Statistical Pattern Recognition”. NewYork: Academic, 1990.

[5]. M. D. Farrell and R. M. Mersereau, “On the Impact of PCA Dimension Reduction for Hyperspectral Detection of Difficult Targets,” *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 192-195, Apr. 2005.

[6]. S. Prasad and L. Mann Bruce, “Limitations of Principal Components Analysis for Hyperspectral Target Recognition,” in *proc. IEEE Int. Geosci. Remote Sens. Lett.*, pp. 625-629, Oct. 2008.

[7]. R. O. Duda, P. E. Hart and D. G. Stork, “Pattern classification,” 2nd edition, A Wiley-Interscience publication, 2001.

[8]. S. Prasad and L. Mann Bruce, “Limitations of Subspace LDA in Hyperspectral Target Recognition Applications,” in *proc. IEEE Int. Geosci. Remote Sens. Symp.*, pp. 4049-4052, Jul. 2007.

[9]. S. Prasad and L. Mann Bruce, “Multiple Kernel Discriminant Analysis and Decision Fusion for Robust Sub-Pixel Hyperspectral Target Recognition,” in *proc. IEEE Int. Geosci. Remote Sens. Symp.*, pp. II-45-II-48, Jul. 2008.

[10]. S. Prasad and L. M. Bruce, "Information Fusion in Kernel-Induced Spaces for Robust Subpixel Hyperspectral ATR," *Geosci. and Remote Sens. Lett., IEEE,* vol. 6, pp. 572-576, 2009.

[11]. Chulhee Lee and David A. Landgrebe, "Feature Extraction Based on Decision Boundaries," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15*,* pp. 388-400, April 1993.

[12]. Stefan A. Robila and Lukasz Maciack, "New Approaches for Feature Extraction in Hyperspectral Imagery," in *proc IEEE Trans. Systems, Applications and Tech. Conference,* pp. 1-7, May 2006.

[13]. Groves, P. and Bajcsy, P., "Methodology for Hyperspectral Band and Classification Model Selection," in *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp. 120-128, October 2003.

[14]. M. P. Derde and D. L. Massart, "Evaluation of the Required Sample Size in Some Supervised Pattern Recognition Techniques," *Analytica Chimica Acta*, vol. 223, no. 1, pp. 19–44, 1989.

[15]. Anuradha Agatheeswaran, "Analysis of the Effects of JPEG2000 Compression on Texture Features Extracted from Digital Mammograms," MS Thesis, Mississippi State University, 2004.

[16]. John E. Ball, "Three Stage Level Set Segmentation of Mass Core Periphery, and Spiculations for Automated Image Analysis of digital Mammograms," PhD Dissertation, Mississippi State University, 2007.

[17]. N. Memarsadeghi, J. Le Moigne, D. M. Mount, and J. Morisette, "A New Approach to Image Fusion Based on Cokriging," in *Proc. 8th Int. Conf. Inf.Fusion*, pp. 25–28, Jul. 2005.

[18]. M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision Fusion for the Classification of Urban Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2828–2838, Oct. 2006.

[19]. J. Chanussot, G. Mauris, and P. Lambert, "Fuzzy Fusion Techniques for Linear Features Detection in Multitemporal SAR Images," *IEEE Trans.Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1292–1305, May 1999.

[20]. S. Prasad and L.Mann Bruce, "Information Theoretic Partitioning and Confidence Based Weight Assignment for Multiclassifier Decision Level Fusion in Hyperspectral Target Recognition Applications," in *Proc. SPIE—Defense Security Symp.*, Orlando, FL, Apr. 2007.

[21]. S. Prasad and L. Mann Bruce, "Hyperspectral Feature Space Partitioning via Mutual Information for Data Fusion," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Barcelona, Spain, Jul. 23–27, 2007, pp. 4846–4849.

[22]. S. Prasad and L. Mann Bruce, "Decision Fusion with Confidence-Based Weight Assignment for Hyperspectral Target Recognition," *IEEE Trans. Geosci. Remote Sens.,* vol. 46*,* pp. 1448-1456, May. 2008.

[23]. T. H. Demetriades-Shah, M. D. Steven and J. A. Clark, "High-Resolution Derivatives Spectra in Remote Sensing," *Remote Sens. Environ.*, vol. 33, pp. 55-64, 1990.

[24]. D. G. Goodin, Luoheng Han, Fraser R. N. and Rundquist D., "Analysis of Suspended Solids in Water Using Remotely Sensed High-Resolution Derivative Spectra," *Photogramm. Eng. Remote Sens.*, vol. 59, no. 4, pp. 505–510, 1993.

[25]. William D. Philpot, "The Derivative Ratio Algorithm: Avoiding Atmospheric Effects in Remote Sensing," *IEEE Trans. Geosci. Remote Sensing*, vol. 29, no. 3, pp. 350-357, May 1991.

[26]. F. Tsai and W. D. Philpot, "Derivative Analysis of Hyperspectral data," *Remote Sens. Environ.*, vol. 66, no. 1, pp. 41-51, 1993.

[27]. Begum Demir and Sarp Erturk, "Spectral Magnitude and Spectral Derivative Feature Fusion for Improved Classification of Hyperspectral Images," in *proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, pp. 1020-1024, Jul. 2008.

[28]. *Analytical Spectral Devices FieldspecPro FR specifications* [Online]. http://asdi.com/productspecifications-FSP.asp.

[29]. HYPERION instrument specifications, available: http://eo1.gsfc.nasa.gov/Technology/Hyperion.html

[30]. *SpecTIR – End to End Hyperspectral Solutions* [Online]. http://www.spectir.com/assets/Images/Capabilities/ProspecTIR%20specs.pdf.

[31]. M. Chi and L. Bruzzone, "Semisupervised Classification of Hyperspectral Images by SVMs Optimized in the Primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pp. 1870-1880, Jun. 2007.