CLASSIFICATION OF REMOTELY SENSED DATA BY USING 2D LOCAL DISCRIMINANT BASES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÇAĞRI TEKİNAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

JULY 2009

Approval of the Graduate School of Informatics

_____

Prof. Dr. Nazife BAYKAL

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. Yasemin YARDIMCI

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Prof. Dr. Yasemin YARDIMCI

Supervisor

Examining Committee Members

Asst. Prof. Altan KOÇYİĞİT (METU, II) _____

Prof. Dr. Yasemin YARDIMCI (METU, II) _____

Dr. Ahmet ÇİZMELİ (METU, GGIT) _____

Dr. Habil KALKAN (KKK) _____

Asst. Prof. Alptekin TEMİZEL (METU, II) _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Çağrı TEKİNAY

Signature : _____

ABSTRACT

CLASSIFICATION OF REMOTELY SENSED DATA
BY USING 2D LOCAL DISCRIMINANT BASES

Tekinay, Çağrı
M.Sc., Department of Information Systems
Supervisor: Prof. Dr. Yasemin Yardımcı

July 2009, 104 pages

In this thesis, 2D Local Discriminant Bases (LDB) algorithm is used to 2D search structure to classify remotely sensed data. 2D Linear Discriminant Analysis (LDA) method is converted into an M-ary classifier by combining majority voting principle and linear distance parameters. The feature extraction algorithm extracts the relevant features by removing the irrelevant ones and/or combining the ones which do not represent supplemental information on their own. The algorithm is implemented on a remotely sensed airborne data set from Tippecanoe County, Indiana to evaluate its performance. The spectral and spatial-frequency features are extracted from the multispectral data and used for classifying vegetative species like corn, soybeans, red clover, wheat and oat in the data set.

Keywords: Remote Sensing, Local Discriminant Bases, Linear Discriminant Analysis, Hyperspectral Imaging, M-ary Classification.

ÖZ

UZAKTAN ALGILANAN VERİLERİN 2 BOYUTLU YEREL AYIRTAÇ TABANLARI İLE AYRILMASI

Tekinay, Çağrı
Yüksek Lisans, Bilişim Sistemleri Bölümü
Tez Danışmanı: Prof. Dr. Yasemin Yardımcı

Temmuz 2009, 104 sayfa

Bu tezde, 2B Yerel Ayırtaç Tabanları algoritması kullanılarak uzaktan algılanan verinin sınıflandırılması sağlanmıştır. İkili sınıflandırma formundaki Doğrusal Ayırtaç Analizi yöntemi, çoğunluk analizi ve doğrusal uzaklık bileşenleri birlikte kullanılarak, ikiden fazla sınıfın sınıflandırılması işlemine uygun hale getirilmiştir. Kullanılan öznitelik çıkarımı algoritması, hiperspektral görüntüleri oluşturan çok sayıda spektral bandın içerisinden, ayırt edici bantların seçilmesi, gereksiz olanların elenmesi veya tek başlarına ayrımsallığı olmayanların birleştirilerek daha yüksek ayrımsallığa sahip bantların oluşturması işlemlerini gerçekleştirmektedir. Algoritmanın performansı, Tippecanoe County, Indiana'ya ait uzaktan algılanmış bir görüntü üzerinde test edilmiştir. Multispektral veriden spektral ve uzamsal-

frekans öznitelikler çıkarılarak görüntü içerisindeki mısır, soya fasülyesi, buğday, kızılyonca ve yulaf ayrıştırılmıştır.

Anahtar Kelimeler: Uzaktan Algılama, Yerel Ayırtaç Tabanları, Doğrusal Ayırtaç Analizi, Hiperspektral Görüntüleme.

ACKNOWLEDGEMENTS

I thank my supervisor Prof. Dr. Yasemin Yardımcı for her patience, encouragement, valuable guidance and support in supervising my thesis. Under her guidance, I have chance to develop myself both academically and personally.

I would also like to thank Dr. Habil Kalkan for his suggestions, valuable comments and feedbacks throughout the thesis. His helps and guidance was very valuable for me.

I would like to thank Asst. Prof. Dr. Alptekin Temizel, Asst. Prof. Dr. Altan Koçyiğit, and Dr. Ahmet Çizmeli for their valuable suggestions. I would also like to thank them for reviewing my work.

I would like to thank to my friends for their encouragement, and suggestions and I also thank to my girlfriend Arzu Burçak Sönmez for her patience, suggestion and understanding during this thesis.

I want to dedicate this thesis to

My Family

for their endless support …

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BTES : Binary Tree Edge Sensing Demosaicking

BTBI : Binary Tree Bilinear Interpolation

CFS : Correlation-based Feature Selection

DAFE : Discriminant Analysis Feature Extraction

DBFE : Decision Boundary Feature Extraction

DMCF : Discrete Measurement Criteria Function

ECHO : Extraction and Classification of Homogenous Objects

EDM : Euclid Distance Measurement

FDB : Fisher Distance Based

FLC1 : Flight Line C1

FLL : Fisher Linear Likelihood

FT : Fourier Transform

HDT : Hybrid Decision Tree

HMM : Hidden Markov Model

KNN : K-Nearest Neighbor

LDA : Linear Discriminant Analysis

LDB : Local Discriminant Bases

MD : Minimum Distance to Mean

MDE : Minimum Differentiated Entropy

MDT : Multivariate Decision Tree

ML : Maximum Likelihood

PCA : Principle Component Analysis

PDC : Probability Distance Criteration

TM : Thematic Mapper

QDT : Quadrature Discriminate Analysis

SVM : Support Vector Machine

UDT : Univariate Decision Tree

WT : Wavelet Transform

CHAPTER 1

INTRODUCTION

Hyperspectral and multispectral images, which are obtained from hyperspectral sensors located in planes, aircrafts and satellites, have a variety of usage like food safety, both military and civilian purposes. Nowadays, the development in the sensor technology leads to the problem that the amount of the provided information and the complexity of the data are rapidly increasing. However, the extensive availability and accessibility of such data encourage the researchers to understand the necessity of machine learning algorithms. In order to analyze the data, automated feature extraction algorithms are used to obtain the necessary content from the hyperspectral images.

The hyperspectral imaging devices are called spectroradiometers. Spectroradiometers not only collect the visible sunlight energy reflected from the surfaces, but also detect and gather a substantial amount of information revealed from the reflected energy in the infrared spectrum that is not visible to the human eye. This widely covered spectrum of

reflected energy which forms the hyperspectral images consists of hundreds of adjacent bands [1].

Hyperspectral images are represented as high dimensional data which can be used to obtain both spectral and spatial information of a scene. Furthermore, hyperspectral imagers reveal considerably more detailed spectral information about the scene of interest than the traditional multispectral sensors [2].

In land use/cover classification problems in remote sensing, the high dimensional data is expected to increase the effectiveness of the classification. However, when the training data is insufficient, increasing the number of spectral bands causes a decrease in the classification accuracy and an increase in the computational complexity and processing time. This is termed by Bellman [3] as “curse of dimensionality” which leads to the “Hughes phenomenon” [4] when the design of classifiers is matter. Additionally, the great amount of information obtained from hyperspectral images causes redundancies due to the repetition of spectral data [2, 5, 6].

In order to overcome these drawbacks, dimension of the available data set should be optimized by not only eliminating the irrelevant bands but also by combining the ones which have no additional information on their own.

1.1 Thesis Statement

In this thesis, Local Discriminant Bases based feature extraction method is applied to classify a land cover data in order to locate the most discriminative contents out of a 12-band multispectral remote sensing image. The dimensionality reduction process is applied to the data set by pruning the spectral and spatial-frequency features along both spectral and spatial axes. The extracted most discriminative features are used for the classification of five different vegetation species inside the data set: corn, soybeans, wheat, oat and red clover.

1.2 Thesis Overview

Chapter 2 introduces the common literature about feature extraction methods. It also presents the Best Bases algorithm and the Local Discriminant Bases algorithm for feature extraction. Some feature selection algorithms and classification methods are also reviewed

Chapter 3 provides a brief overview on the feature extraction and classification techniques used in remote sensing.

Chapter 4 presents the methodology behind the feature extraction, selection and classification steps used in this study. This section describes the specific steps and explanations about the proposed algorithm.

Chapter 5 provides detailed information about the multispectral data set used in the experiment. This section describes the specific properties of the data set and gives information about the pre processing phase.

Chapter 6 presents the outputs of feature extraction step. The classification results are given for the LDB approach for two different test and training sets.

Chapter 7 presents conclusions and recommendations for future works.

CHAPTER 2

Literature Review and Background

2.1 Feature Extraction and Feature Selection

The process of dimensionality reduction in the literature can be grouped as feature extraction and feature selection. The feature extraction can be defined as the transformation of an N dimensional feature to an M dimensional feature vector [7] where N is greater or equal than M.

In one perspective, feature selection can be thought as a subset of feature extraction [8] through a selection procedure realized by assigning ones and zeros to the feature coefficients. Although it minimizes/optimizes the feature set, this condition may not always lead to desired classification accuracy [9, 10]. The accuracy and the development of the feature extraction algorithms mostly depends on the structure of the problem and the type of the data [11].

There is an extensive literature on feature extraction and feature selection algorithms. Not only both supervised and unsupervised feature extraction algorithms exist [12] and used for dimensionality reduction, but also this separation can be extended by grouping the supervised and unsupervised methods into linear and non-linear feature extraction methods [13].

The commonly used linear methods for extracting features from an original data set are the common-mean feature extraction [14], the decision boundary feature extraction (DBFE) [6] and, probably the most widely preferred linear method, Linear Discriminant Analysis Feature Extraction (LDA or DAFE) [15,16,17]. Based on the Fisher distance, the LDA is a fast and simple method. These methods are also known as statistic-based feature extraction methods because the extraction of features relies on the statistical theory. However, some of the well known drawbacks of LDA and training set dependencies [18, 19, 20, 21] reduces its usability.

Another popular feature extraction algorithm in remote sensing applications is the Principle Component Analysis (PCA) [22]. The PCA algorithm converts the correlated variables into a small size of uncorrelated variables by using Karhunen-Loeve transformation. The basic PCA has the drawback of using all of the data set for converting process. Instead the segmented PCA method is proposed by Jia and Richards [23]. Also a Kernel PCA was introduced for remote sensing purposes by Fauvel, Channussot and Benediktsson [12]. PCA is also

known as a statistic-based feature extraction method. Statistic-based feature extraction methods are evaluated by the results of classification process [24].

Additionally, a valuable method for feature extraction is the Fourier Transfrom (FT). Fourier Transform is a frequency based feature extraction method. Basically, frequency based methods are used to represent the hyperspectral signal, which is infact a hyperspectral curve of each pixel, from spectrum based to frequency based [25]. Similarly, another frequency based method for feature extraction is the wavelet transform (WT) [26]. One of the most important reason why the wavelet features can be suitable for image classification is its multiresolution approach [27]. When extracting the wavelet features, mostly orthogonal and representatively unique wavelet basis functions are used and low and high pass filters are applied along the specified rows and columns of the image for decomposition. Each combination of bidirectional filtering gives the information about the image. For example applying high-pass filters in both directions (HH) gives details along the diagonal direction. Also HL, LH filtering gives horizontal, vertical details of the image. LL filtering enhances low frequencies.

Feature selection algorithms can also be grouped as filter and wrapper methods. The wrapper methods use a predefined induction algorithm and cross-validation to select the final set of features out of the whole set. Filter methods, on the other hand, use general properties of the features and they do not use any learning algorithm [28]. Sanmay

proposed a new boosting-based hybrid feature selection method in 2001 which combines the filter and wrapper methods [29]. Another way of applying feature selection is searching for maximum difference between the classes. Method based on Fisher distance can be used for such discrimination and can be used as a basis for feature selection [30].

Commonly used feature selection methods based on filter approach are discretization based feature selection [31], correlation based feature selection (CFS) [31]. Some feature selection methods based on information theory are also used [32].

2.2 Best Basis Algorithm

A best orthogonal basis algorithm was introduced by Coifman and Wickerhauser to reach the goal of efficient signal compression [33]. As a first step, the original signal is represented by an orthonormal wavelet-packets or trigonometric basis in a binary tree. At the second step, the algorithm computes the value of each node of the binary tree to reach minimum entropy transform. Then, using a leaf to root approach, the tree is pruned by using the entropies of nodes. Three main parts of the algorithm can be summarized as [34, 35, 36]:

1. Expansion of the original signal in a binary tree after selecting a decomposition method (orthonormal wavelet-packets or localized trigonometric functions) to obtain a coefficient vector.

2. Computation of the entropy of each node using the coefficients determined in the previous step. The entropy H of a sequence {p} with $\sum p_i = 1$ is :

$$\boldsymbol{H}(\boldsymbol{p}) = \sum_{\boldsymbol{i}} \mathrm{p}i \; \boldsymbol{log}(\mathrm{p}i) \qquad (\textbf{Equation 2.1})$$

3. Comparison of frequency subbands with respect to their entropy and pruning of the binary tree in order to achieve the minimum cost.

The parent and the children nodes change adaptively during the pruning process. At the beginning, the children nodes are the bottom level leaves of the binary tree and the parent nodes are the ones located at the one upper level of these children nodes. Then, the decision is made according to the cumulative value of the children nodes and compared with the entropy of the parent node. If the parent has greater entropy than the cumulative value of its children, then the children nodes are omitted and the parent node is assigned as a new child for upper levels of the binary tree. At the end of the pruning

process, the remaining nodes are the best bases for signal representation and compression.

2.3 Local Discriminant Basis Algorithm

Different from the best-basis algorithm [33] which uses the entropy values of the binary tree components as an information measurement at the pruning phase to obtain the local information out of the signals representation, the LDB algorithm maximizes the distinguishing ability of classes in time-frequency energy distributions by using a distance function [37].

At the beginning of the LDB algorithm, the given training signal is decomposed into a dictionary of orthonormal bases by using either localized wavelet packets or local trigonometric functions in binary tree form. After the decomposition of the signal, the discrimination power of each binary tree component is calculated by using a distance function. Most commonly used distance (dissimilarity) measures are J-divergence, Fisher Distance, Euclidean Distance and Hellinger Distance. The LDB algorithm locates the best bases by pruning the binary tree.

Eliminating a parent node or eliminating the children from the binary tree depends on their discriminative powers. Child nodes are eliminated only if their cumulative discrimination power is smaller than their parent node. In that case, the parent node remains as a child node of an upper level node. Otherwise, the child nodes remain in the binary tree whenever their cumulative discrimination power is greater than their parent. In this situation, the discrimination power will be carried by the parent node.

2.4 Classification

Classification process in image analysis can be defined as assigning samples to groups based on their features and parameters [38]. Classification methods can be categorized as supervised and unsupervised methods. The aim of supervised classification is to match the samples to the previously selected classes. On the other hand, unsupervised classification algorithms promote the use of clustering based approaches because they try to define the class boundaries by grouping the samples with similar characteristics together [39]. If a classifier takes advantage of predefined labeled data, it is considered as a supervised classifier. For example, a simple LDA classifier can be considered as a supervised classification algorithm because it uses the labeled class samples as a training data. On the other hand, if a classifier has a tendency to ignore the labels when classifying, it can be

defined as an unsupervised classifier. A commonly used PCA classifier can be an example of an unsupervised classification algorithm.

In the literature, there is a variety of classification algorithms like k-Nearest Neighbor (kNN), Hidden Markov Model (HMM), Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), etc... These classifiers require working with relevant and orthogonal feature sets. In this thesis, a simple LDA classifier is used for the analysis of the multispectral data because the main focus of this study is the extraction of most discriminative features for classification purposes.

As a simple and one of the most common statistics based classifiers [40], LDA computes the best hyperplane to separate the N dimensional data which is assumed to be linearly separable from each other. An example decision boundary line for a linearly separable sample class 1 and class 2 is given in Figure 2.1.

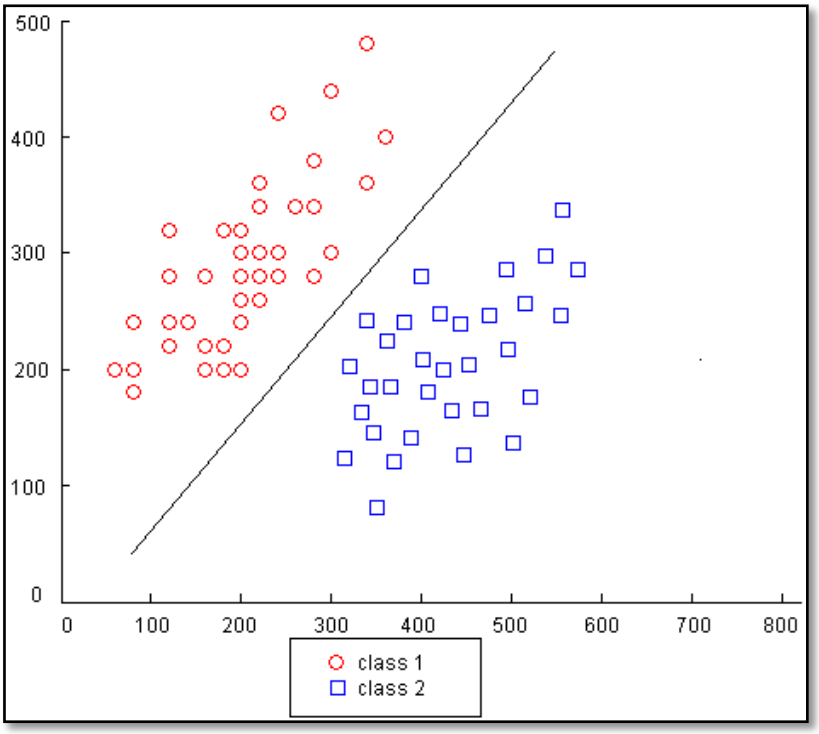

Figure 2.1 – The decision boundary for the class 1 and class 2 obtained by LDA.

Although the LDA algorithm has low computational complexity and it is easy to implement, it has some major drawbacks which can directly affect the classification accuracy. One of the most important disadvantages of applying LDA is the small sample size problem. In order to make the LDA applicable and prevent the between-class scatter matrix from becoming singular, the number of the inputs should be smaller than the number of samples [21].

Furthermore, the LDA algorithm can only separate the classes with hyper-planes only if they are linearly separable. A more sophisticated Quadrature Discriminate Analysis (QDA) is proven to be more effective [20] because of its ability to separate classes with parabolas and hyperbolas. Also Support Vector Machine (SVM) classifier statistically outperformed the LDA classifier in [20] experimentally on a number of data sets. In order to overcome the requirements of training samples and complete class knowledge, the original algorithm was modified by Qian Du [19].

CHAPTER 3

Hyperspectral and Multispectral Imaging in Remote Sensing Literature

3.1 Multispectral and Hyperspectral Imaging for Land Cover/Use Studies

Classification and mapping studies from remotely sensed hyperspectral and multispectral satellite and airborne imageries have long been used by the remote sensing community for land use/cover analysis, instead of the conventional classification methods requiring land surveys [41, 42, 43, 44]. Classification of the land cover areas is one of the most important study areas in remote sensing [45, 46]. Some of the image classification studies use the spectral discrimination of classes because most of the land cover types have their own characteristic intensity values [46, 47]. Although the intensity values of the land cover types is

an important parameter, spatial composition can also be used for a deeper analysis and better accuracies [48].

Multispectral and Hyperspectral sensors can scan the surface of the earth from satellites and aircrafts. One of the main differences between these two platforms is the spatial resolutions that they provide. Different from the spectral resolution which gives the number and positions of the spectral bands located in the electromagnetic spectrum, the spatial resolution represents the level of spatial detail presented in an image. Spatial resolution not only has a relation with the image pixel size but also dependent of the viewing angle and altitude of the sensor systems [49, 50].

Sensor systems receive the reflectance energy from small areas on the earth surface known as patches. Size of an individual patch gives the spatial resolution. That means smaller patches provide higher spatial resolution [50]. Nowadays, the commercial satellite systems may have spatial resolutions in the order of 50 cm and better.

Multispectral sensor systems located inside different platforms like aircrafts and satellites have been used for agricultural and other classification studies over a half century [51].

On the other hand, the traditional multispectral systems have a limited number of spectral bands which can only be increased up to six spectral bands covering visible and near-infrared part of the spectrum.

However, there are some available multispectral systems adding thermal infrared bands to their spectrum coverage [49]. Because of those limitations of multispectral systems, hyperspectral technologies have become the area of interest for almost over two decades.

The results from different studies which compare the accuracy of multispectral and hyperspectral sensors for agricultural and other parametric classifications is given at [52]. Offering the ability to collect data from more than 200 spectrally adjacent bands, which are also ordered by their wavelengths, hyperspectral data cover the visible, near-infrared, mid-infrared, and short-wave infrared domains [53]. On the other hand, the difficulty of processing the huge amount of data, limited training samples comparing to the high dimension and high data correlation between the adjacent spectral bands are the concerns of hyperspectral classification [4, 54].

3.2 Feature Extraction and Feature Selection from Remotely Sensed Hyperspectral/Multispectral Data

Feature extraction from remotely sensed data is important since it extracts the key attributes for an area of interest before the classification step and it helps to discriminate the land cover areas in the scene. In land use/cover classification problems, availability of high

dimensional data is expected to increase the effectiveness of the classification. On the other hand, when the training data is insufficient, increasing number of spectral bands causes a decrease in the classification accuracy and an increase in the computational complexity [5, 6].

Hyperspectral sensors gather images from the earth surface by picking great number of adjacent spectral bands continually. Although this high dimensional data seems very beneficial for remote sensing, the information in close spectral bands is highly correlated and the images often tend to be highly redundant. Also this high dimensional data consumes great amount of computational time and performance [8, 55].

To solve this limitation, feature vector size is decreased by removing the irrelevant features obtained from the raw data and/or combining the ones together which do not represent supplemental information on their own [56, 23, 36]. There are a great number of studies in remote sensing literature using different kinds of feature extraction and feature selection problems. The performances of feature extraction methods are usually compared by their classification accuracies [24].

Hui Hsu and Hsing Tseng [25] compare the accuracy of frequency based feature extraction algorithms, best basis algorithm and two versions of LDB algorithm (cross entropy based and L^2 norm) to classify the four vegetation classes in the Aviris Indiana Land Cover data set. The results show that non-linear methods give better results than linear methods. Also LDB algorithm classifies the classes with 93%

accuracy by using 12 extracted features and also reduces the data dimension dramatically.

Pal [57] proposed a margin-based feature selection method and compared it with the SVM based algorithm in order to classify two different hyperspectral data. DAIS data consists of 65 spectral bands containing 8 vegetation classes and Aviris data consists of 185 bands containing 9 vegetation classes. The results show that two margin-based feature selection methods reduced the data dimension from 65 to 24 and 185 to 65 by reaching 92.6% and 82.4% classification accuracy, respectively.

Bruce, Koger and Li [26] proposed a dyadic discrete wavelet transform to classify the vegetative classes from hyperspectral data taken with a portable hand-held spectroradiometer. Although the reflectance signal between 350 nm and 2500 nm contains 2151 contiguous spectral bands, only the first 1000 bands were used for the experiment. The algorithm reaches 95% and 80% classification accuracy for end member and mixed-signature applications by using the extracted features from the proposed algorithm.

Fauvel, Channussot and Benediktsson [12] developed an unsupervised Kernel Principle Component Method to analyze the hyperspectral remote sensing land cover data. The proposed method increases the classification accuracy from 79% to 96% for one data set. Borra [58] used the Robust Principle Component Analysis method to a 60 bands

hyperspectral land cover data. The presented results show that the 60 bands were reduced to 17 by using a hierarchical dimensionality reduction method. After applying PCA to the reduced data set, the number of bands dropped to 5.

Tian, Guo and Lyu [59] evaluated five feature extraction methods in the order of the classification results. The evaluated feature extraction methods are Euclid Distance Measurement (EDM), Discrete Measurement Criteria Function (DMCF), Probability Distance Criterion (PDC), Minimum Differentiated Entropy (MDE) and Principle Component Analysis. For the experiment, 5 different Landsat-5 thematic mapper (TM) images of Beijing, China with 7 spectral bands were used. After the experiments, the data dimension was reduced to 3 bands and all the classification accuracies were over 90% for all the images. The results show that the EDM and PDC algorithms were better when extracting the key features while MDE reached an average performance and DMCF and PCA methods were the worst among all.

3.3 Classification Problem in Hyperspectral and Multispectral Remotely Sensed Data

Classification of remotely sensed data is a very important procedure as it provides results relevant for many environmental and economical applications. Although the number and the variety of researches on

classification of remotely sensed data using statistical, structural methods is rapidly increasing [60, 6], the recent study of Wilkinson [44] stated that the classification rates of the presented methodologies in the last 15 years showed no significant improvements.

A widely encountered problem which was studied by Hughes [4] for remotely sensed data classification is the inconsistency in the classification accuracy when the number of available training samples is limited and the dimension of the data is increasing. The situation where an increase which is followed by a decrease in the classification performance caused by the small ratio of number of training samples to the number of spectral features is known as the Hughes Phenomenon.

Different from the Hughes Phenomenon, selection of training samples that are poorly representative of the features of interest is another common problem in remote sensing studies. This kind of drawback occurs when the training samples of a class are gathered from a limited region, although the class samples are distributed all over the region of interest. In order to overcome the limited and unrepresentative training sample problem, Jackson and Landgrebe developed and proposed an adaptive classifier and improved the classification accuracy up to 93.4% for the Flight Line C1 data set after 5 iterations [61].

In pixel-based approaches, each pixel is evaluated individually based on its spectral reflectance value in order to classify the whole remotely sensed data [62]. The comparison is made between the reflectance

value of a single pixel and the spectral representation of a predefined land cover area.

Since the spectral information gathered from a single pixel does not have the ability to represent the remotely sensed land covers by itself and some complex land cover areas can only be identified by the combination of many pixels, the spatial information is a very important component for remotely sensed image classification. Image segmentation techniques are the approaches for the usage of spatial information to classify the land cover types [60, 63]. Some of the most common image segmentation approaches are region growing, Markov random field model and fuzzy rule-based classification [64, 65, 66].

There is an extensive literature on land cover classification in remote sensing. Some of the studies for land cover classification using hyperspectral and multispectral data are given as an example. Jackson and Landgrebe [61] presented a self-learning and self-improving adaptive classifier for 12 vegetative classes from the 12-band Flightline C1 dataset and reached 93.4% classification accuracy with only 5 iterations.

Brodley, Friedl and Strahler [67] compared the classification accuracies of three different decision tree classifiers using three different hyperspectral land cover data sets. The Univariate Decision Tree (UDT), Multivariate Decision Tree (MDT) and Hybrid Decision Tree (HDT). A linear classifier was used as a reference and the results show

that three decision tree algorithms outperformed the linear classifier. The classification accuracies for UDT, MDT, HBT and linear classifier were 85.61%, 86.39%, 87.70% and 78.66% respectively for the first data set. The other results for the other two data sets were also similar.

Dundar and Landgrebe [68] developed a supervised classification algorithm based on Bayes rule with kernel to reach high classification accuracy by using 3 different hyperspectral data sets. The data sets are the 12-band Flightline C1 vegetation data set, the 220 contiguous banded Washington DC mall HYDICE data set and the 126-band Purdue University West Lafayette Campus data set. The proposed algorithm produced a considerably low mean error after being applied to all data sets.

Karakahya, Yazgan and Ersoy [62] proposed a new spectral-spatial classifier for remote sensing data and compared the results over traditional per-pixel classifiers and Extraction and Classification of Homogenous Objects (ECHO) algorithm. The accuracies were tested using 12-band Flightline C1 dataset and the results show that the algorithm outperformed the Minimum Distance to Mean (MD), Maximum Likelihood (ML) and Fisher Linear Likelihood methods and was slightly better than the ECHO algorithm by reaching 92.1% classification accuracy using only the first 3 features.

CHAPTER 4

METHODOLOGY

4.1 Feature Extraction from Hyperspectral Data by 2D Structured LDB Algorithm

Most discriminative features in hyperspectral data are extracted by the LDB algorithm which is used as a representation of 2D signals. The previous version of the LDB algorithm finds the discriminative features by decomposing the time axis into local cosine packets or frequency axis into wavelet packets. However, in the hyperspectral imaging perspective, both the local cosine packets and wavelet packets are valuable and instead of the time axis the spectral axis is used [60, 70, 71].

Without knowing any prior spectral and spatial-frequency location information, the modified 2D structured LDB algorithm gets the exact location of these discriminative features. After pruning both axes, the algorithm combines the irrelevant features which do not contain useful

information on their own, to get the most discriminative features. After the pruning step, the feature selection algorithm selects the most discriminative features for classification.

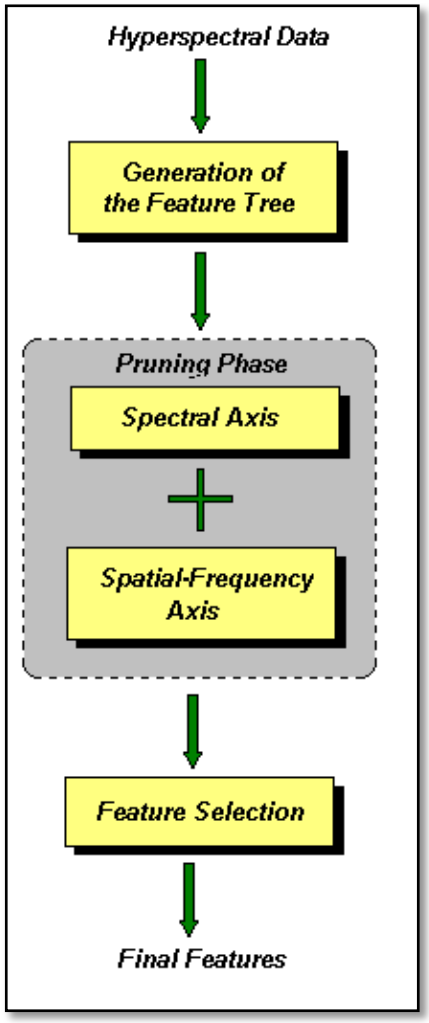

Figure 4.1: The block diagram of the 2D LDB based feature extraction for hyperspectral data [11].

4.1.1 Generation of the Feature Tree

Although the features are important for the classification of hyperspectral images, repetitive features from the adjacent spectral bands should be eliminated in order to increase the classification accuracy of the algorithm. Furthermore, reduction in the dimensionality also plays an important role. During the feature tree generation process, two feature trees on both axes are generated in order to get the candidate feature set.

4.1.2 Generation of the Spectral Feature Tree

In the generation of the spectral feature tree phase of the algorithm, the reflectance energies of images are placed at the bottom (n^{th}) level of the binary tree from left to right. If the number of spectral bands in the hyperspectral image are smaller than the number of leaves at an n^{th} level binary tree, remaining leaves should be set to null. An n=4 levels spectral feature binary tree with 16 spectral bands is shown in Figure 4.2.

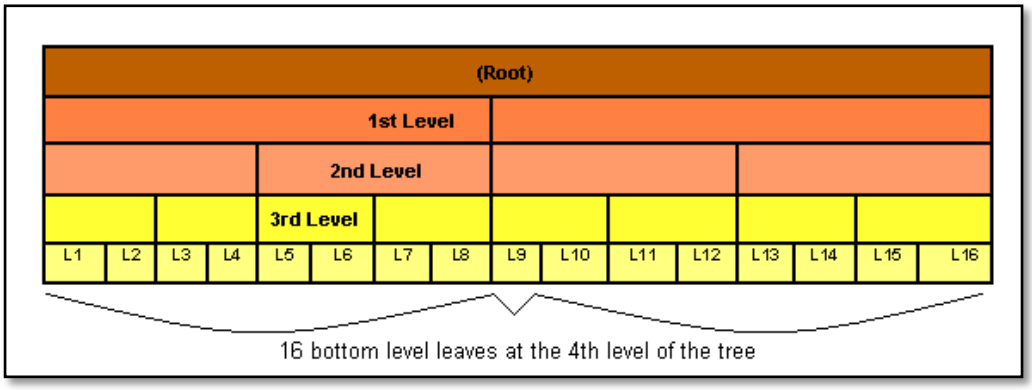

Figure 4.2: Spectral Band Binary Tree for 4 levels.

4.1.3 Generation of the Spatial-Frequency Feature Tree

The second sub step of the feature tree generation phase is decomposing the spectral segments into frequency subbands in order to get the frequency information of the signals. The reason behind that is to locate and process the local patterns in signals. Therefore, the spectral images separated into k level wavelet subbands of LL, LH, HL and HH in a quad tree formation to generate the spatial-frequency feature tree. L stands for “low” which represents row directional filtering and H stands for “high” which shows column directional filtering. For (k = 4) level full wavelet decomposition of a spectral image results in obtaining ($2^{k=4}$ = 341) features and a 3 level decomposition generates 85 features. A 4 level full wavelet decomposition quad tree illustration is given in Figure 4.3.

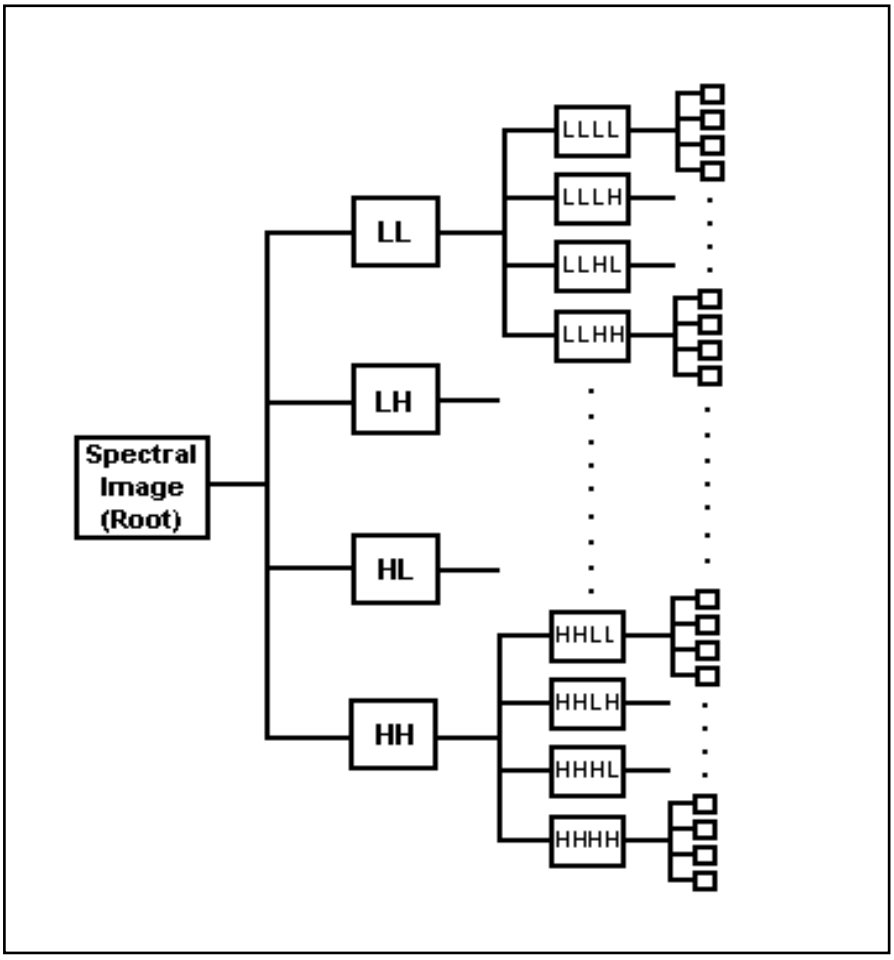

Figure 4.3: Three level full wavelet decomposition quad tree with 85 subbands.

4.1.4 Spectral Axis Pruning

A leaf to root approach is used to prune the binary spectral tree to obtain the spectral bands which have better discrimination power. In order to reach that goal, the algorithm combines some of the spectral bands which have lower classification accuracy on their own. The energy of the frequency components of the spectral bands which were combined in spectral pruning are averaged before the spatial-frequency quad tree pruning.

The binary spectral tree pruning algorithm consists of two main steps:

1. **If exists** $\max\{ d_{child\_node1} , d_{child\_node2} \}$ **where**

   $$\max\{ d_{child\_node1} , d_{child\_node2} \} > d_{mother\_node}$$

2. **Assign** $\max\{ d_{child\_node1} , d_{child\_node2} \}$ **as new** d_{mother_node}

   else delete children nodes

In the pruning algorithm, d_i stands for “the distance of i^{th} node feature between the classes”. The cumulative probability distribution of the nodes is calculated by using Euclidean Distance. The Euclidean Distance is calculated by using:

$$Euclidean\ Distance:\ D(x,y) = \sum_{n=1}^{m}(x_n - y_n)^2 \qquad \text{(Equation 4.1)}$$

where x_n and y_n are normalized energy distributions of features from class 1 and class 2.

4.1.5 Spatial - Frequency Axis Pruning

Because of the spatial-frequency feature tree's quadrature structure, the pruning algorithm for spectral tree given in Section 2.1.3 is modified to four children scales. The modified leaf to root approach can be formulated as:

1. **If exists** max{ d_{child_node1} , $d_{child\_node2,}$ d_{child_node3} , d_{child_node4}} **where**

   max{ d_{child_node1} , $d_{child\_node2,}$ d_{child_node3} , d_{child_node4}} > d_{mother_node}

2. **Assign** max{ d_{child_node1} , $d_{child\_node2,}$ d_{child_node3} , d_{child_node4}}

   as new d_{mother_node}

   else delete children nodes

In this modified pruning algorithm *child_node1*, child_*node2*, child_*node3* and child_*node4* represents the LL, LH, HL and HH wavelet subbands of the parent node. The algorithm compares the discrimination accuracy of the wavelet subbands to their mother node and decides to keep them if their cumulative potential is greater than that of the mother node. Otherwise, the child nodes will be deleted from the spatial-frequency feature quad tree.

4.1.6 Selection of Located Spectral & Spatial-Frequency Features

Although the feature extraction process prunes both the spectral and spatial-frequency axes and obtains the best combination of features located in both axes, the process does not remove the irrelevant features. For this reason, the feature extraction process is followed by a feature selection algorithm that selects relevant features and feeds them into the linear classifier one after another. This feature selection phase finds out the optimal feature subset.

The feature extraction phase should minimize the classification error by using this best feature subset. Approaching the ideal of reaching a higher classification accuracy with as fewer number of features as possible. In this thesis, the LDA statistical classifier is used to reach this high classification accuracy.

In the beginning of the 2D structured LDA method, two classes out of ***N*** number of classes and an observation data with ***j*** number of features like $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$, ... , $\mathbf{f}_j$ are known. The classes are called dependent variables and the features are called independent variables. In this phase, a solid line ***L*** which maximizes the separation between class projections, the decision boundary, is identified as a hyperplane. Using this hyperplane, the algorithm gives a positive or negative value to the observation pattern based on whether the given observation pattern is located on its positive or negative side, respectively.

Based on the result, the observation pattern is assigned to the first class or the second class. This can be called as the membership parameter. Furthermore, the method not only determines the membership of the pattern to one of these classes, but also reveals the distance of the observation pattern to the decision hyperplane. This distance parameter is used as a complementary parameter to the membership parameter to evaluate the strength of the assumption of this membership.

In this thesis, we deal with more than two classes; therefore, the 2D structured LDA algorithm is modified to a multi-class separation structure. The modified version of LDA basically has the same fundamental steps that 2D structured LDA has. The main difference is in the membership decision step. The differences are:

1. The original 2D structured LDA identifies the location and orientation of the hyperplane, the boundary line, by only using the class projections. After the hyperplane is identified, the observation pattern is assigned to a class based on its location with respect to the hyperplane. The result shows the membership of the pattern to a class.

   However, there are more than two different classes in the dataset. Therefore, a final decision can be made by applying majority voting rule to the final result set in order to obtain the membership information for a given observation pattern. The result set is in N-to-N matrix where N is the number of classes. This matrix is called the decision matrix, M_R. The decision matrix only contains $\binom{N}{2}$ number of relevant results because some elements of the matrix contain duplications and some are useless for the case. For example, both $M_D(j,m)$ and $M_D(m,j)$, identify the same selections where m and k represent different classes. Also the results located on the diagonal axis of the decision matrix are irrelevant for this particular application like $M_D(m,m)$ or $M_D(j,j)$.

   For the three-class LDA classifier problem, the decision matrix should be 3x3 matrix. For example, $M_D(1,3)$ represents the membership result, obtained after applying 2D structured LDA between the first and third classes. Also, $M_D(2,3)$ contains the membership result, after applying 2D structured LDA between the second and third classes for a given pattern. The elements of

the decision matrix are either +1 or -1 depends on the observation pattern's location with respect to the hyperplane.

At this stage, we made an assumption that the small numbered class should always be located at the negative side of the hyperplane for not causing any misconceptions. For instance, class1 is always located at the negative side when it is compared to class2 or class3, as well as class 2 is located at the negative side when it is compared to class 3 in a three class LDA integration. An example 3x3 decision matrix is given in below for explanatory purposes:

$$\mathbf{M_D} = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

The elements located at the diagonal axis of the decision matrix and located below the diagonal axis are assigned as 0 (zero). This means that these members are irrelevant for our application. The reason behind this assignment is that, for example, $M_D(2,3)$ and $M_D(3,2)$ represents the same pairwise selection by choosing both the second and the third class. Also the combinations located along the diagonal axis of the decision matrix; $M_D(1,1)$, $M_D(2,2)$ and $M_D(3,3)$ are irrelevant for this problem. After eliminating these matrix elements, the only remaining combinations are $M_D(1,2)$, $M_D(1,3)$ and $M_D(2,3)$ for this particular case.

The result “+1” at the $M_D(1,2)$ means that after applying the 2D structured LDA algorithm to the observation pattern for the first and second classes, the pattern is located at the positive side of the hyperplane where the second class is located. The result “+1” at the $M_D(1,3)$ means that, after applying the 2D structured LDA algorithm to the observation pattern for the first and third classes, the pattern is located at the positive side of the hyperplane where the third class is located. Finally, the result “-1” at the $M_D(2,3)$ means that, after applying the 2D structured LDA algorithm to the observation pattern for second and third classes, the pattern is located on the negative side of the hyperplane where the second class is located. The negative and positive side placement decisions are based on our previous assumption about assigning the lower numbered classes to the negative side of the hyperplane.

Application of the majority voting rule to the decision matrix is expected to give the sought solution. The modified 2D structured LDA algorithm decides that the pattern belongs to the second class for the example decision matrix M_D. This decision is made because the second class has 2 votes while the first class has 1 vote and the third class has no (zero) vote. The majority of the votes lead to this decision.

2. If there is more than one class that has the same number of votes, then the majority voting could create a deadlock. Another parameter is needed for verifying the membership decision. Therefore, not only the location of the observation pattern is used, but also the distance with respect to the hyperplane is taken into consideration when deciding the membership of a pattern. Also for a multi-class comparison, the distance parameter helps to increase the certainty of membership assignments.

   The distance result set is also an N-to-N matrix, D_D, which contains the axial distance of patterns with respect to the LDA hyperplane. Distance matrix has exactly the same properties that decision matrix, M_D, has. A different 3x3 decision matrix and its 3x3 distance matrix are given below for explanatory purposes:

$$\mathbf{M_D} = \begin{bmatrix} \mathbf{0} & \mathbf{-1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad \mathbf{D_D} = \begin{bmatrix} \mathbf{0} & \mathbf{1.7128} & \mathbf{0.7031} \\ \mathbf{0} & \mathbf{0} & \mathbf{2.0012} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

   The "-1" value at the $M_D(1,2)$ and $M_D(2,3)$ matrix elements means, applying the 2D structured LDA algorithm to the observation pattern for the first and second and second and third classes shows that the pattern could be a member of the first class or the third class. Furthermore, the "+1" value at the $M_D(1,3)$ matrix element means, applying the 2D structured LDA algorithm to the observation pattern for the first and third classes

shows that the pattern could be a member of the third class. It is observed that each class has the same number of votes which will lead the application to a decision deadlock. In order to avoid this deadlock, the distance parameter is used.

For the complementary distance parameter, it is assumed that maximizing the axial distance for the observation pattern with respect to the hyperplane maximizes the certainty of the membership. $D_D(2,3) = 2.0012$ has the maximum axial distance for the above example. For that reason, the class which the observation pattern belongs to is given by $M_D(2,3) = -1$ which corresponds to the 2^{nd} class.

CHAPTER 5

THE EXPERIMENTAL DATASET

5.1 The Dataset

Flightline C1 (FLC1), a historically significant 12-band multispectral data set, is used to evaluate the performance of the proposed algorithm [72]. This RGB color coded multispectral data covers the southern part of Tippecanoe County, Indiana. The flight line follows a county road from the Grandville Bridge over the Wabash River just south of South River Road (West Lafayette) to near State Highway 25. Although the data were obtained by using the M7 airborne scanner in June 1966, the data remains contemporary because of its key attributes.

The FLC1 data set not only contains a satisfactory number of different vegetation classes but also consists of more than a few spectral bands (12 bands) which make it a very valuable for illustrative purposes. The

12-band multispectral image which consists of 949 scan lines with 220 pixels per scan line was collected by scanning the terrain at an altitude of 2600 ft. More information about this data set can be gathered from [73]. Table 5.1 shows the spectral bands of the FLC1 data set:

Table 5.1 - Spectral Sensitivity of the multispectral Flightline C1 data set.

Band Number	Wavelength Interval
1	0.40 - 0.44 μm
2	0.44 - 0.46 μm
3	0.46 - 0.48 μm
4	0.48 - 0.50 μm
5	0.50 - 0.52 μm
6	0.52 - 0.55 μm
7	0.55 - 0.58 μm
8	0.58 - 0.62 μm
9	0.62 - 0.66 μm
10	0.66 - 0.72 μm
11	0.72 - 0.80 μm
12	0.80 - 1.00 μm

The images belonging to each band of the FLC1 data set are given in Figure 5.1, Figure 5.2 and Figure 5.3 respectively.

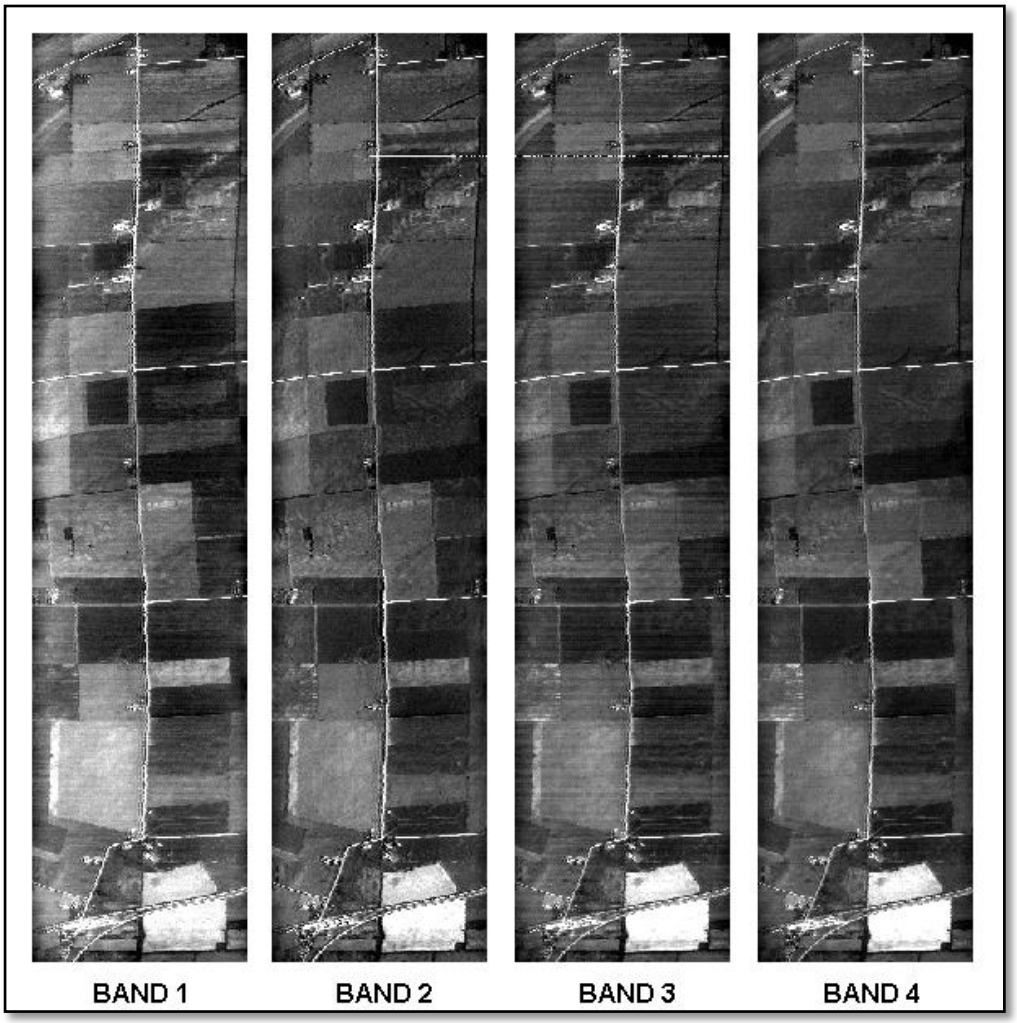

Figure 5.1 - Four spectral band images between 0.40 μm - 0.50 μm in grayscale format are given respectively.

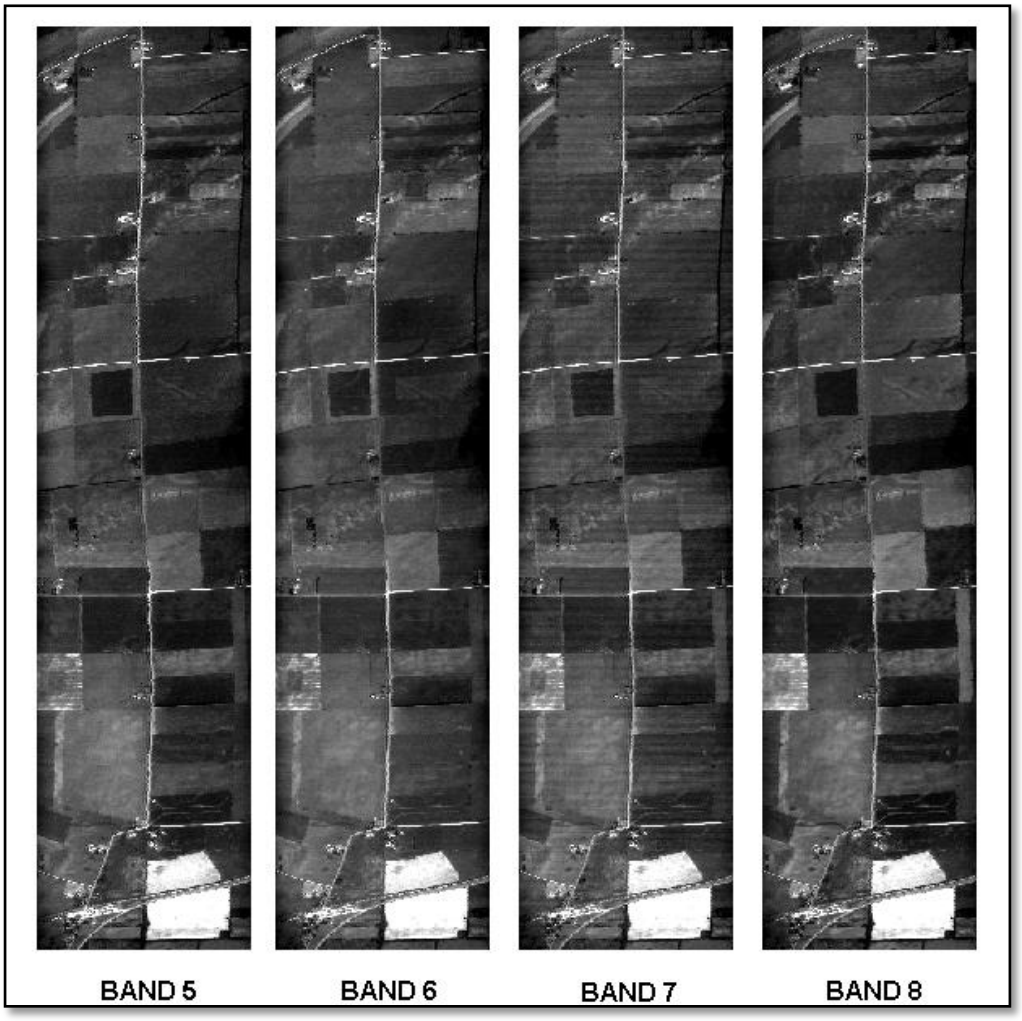

Figure 5.2 - Four spectral band images between 0.50 μm - 0.62 μm in grayscale format are given respectively.

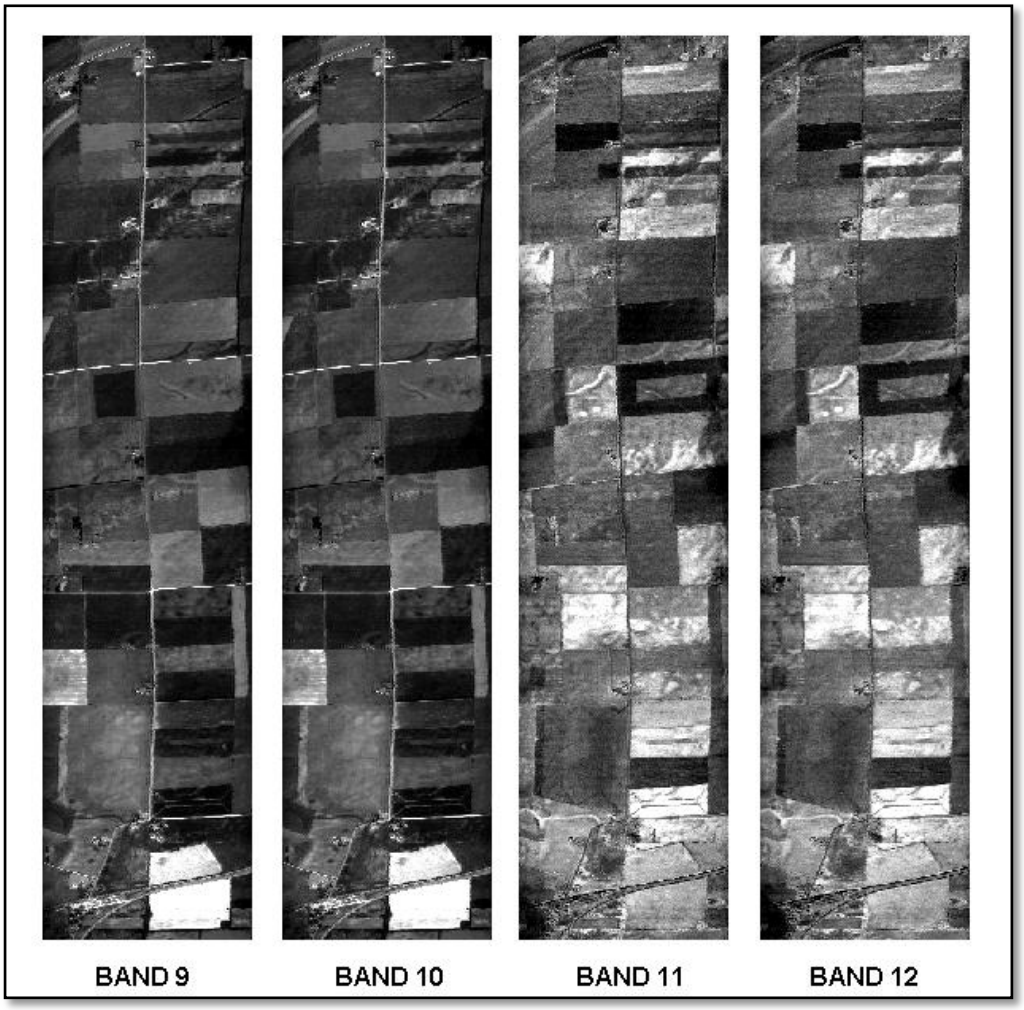

Figure 5.3 - Four spectral band images between 0.62 μm - 1.00 μm in grayscale format are given respectively.

The availability of ground truth information makes the FLC1 data set ideal for the application of feature extraction and classification algorithms. Based on the ground truth data, the data set was divided into five classes. Every class in the experimental data set represents different vegetative species. These five vegetative classes are corn, soybeans, wheat, oat and red clover.

5.2 Preprocessing the Data Set

In order to prepare the test and training images, binary masks for each class are generated. These masks are used to extract the class samples from the whole image and separate this class samples from the other class samples. This process is applied to every class and every spectral band. In fact, a binary mask of a specific class is a ground truth map of this class. Figure 5.4 and Figure 5.5 shows the masks for every class.

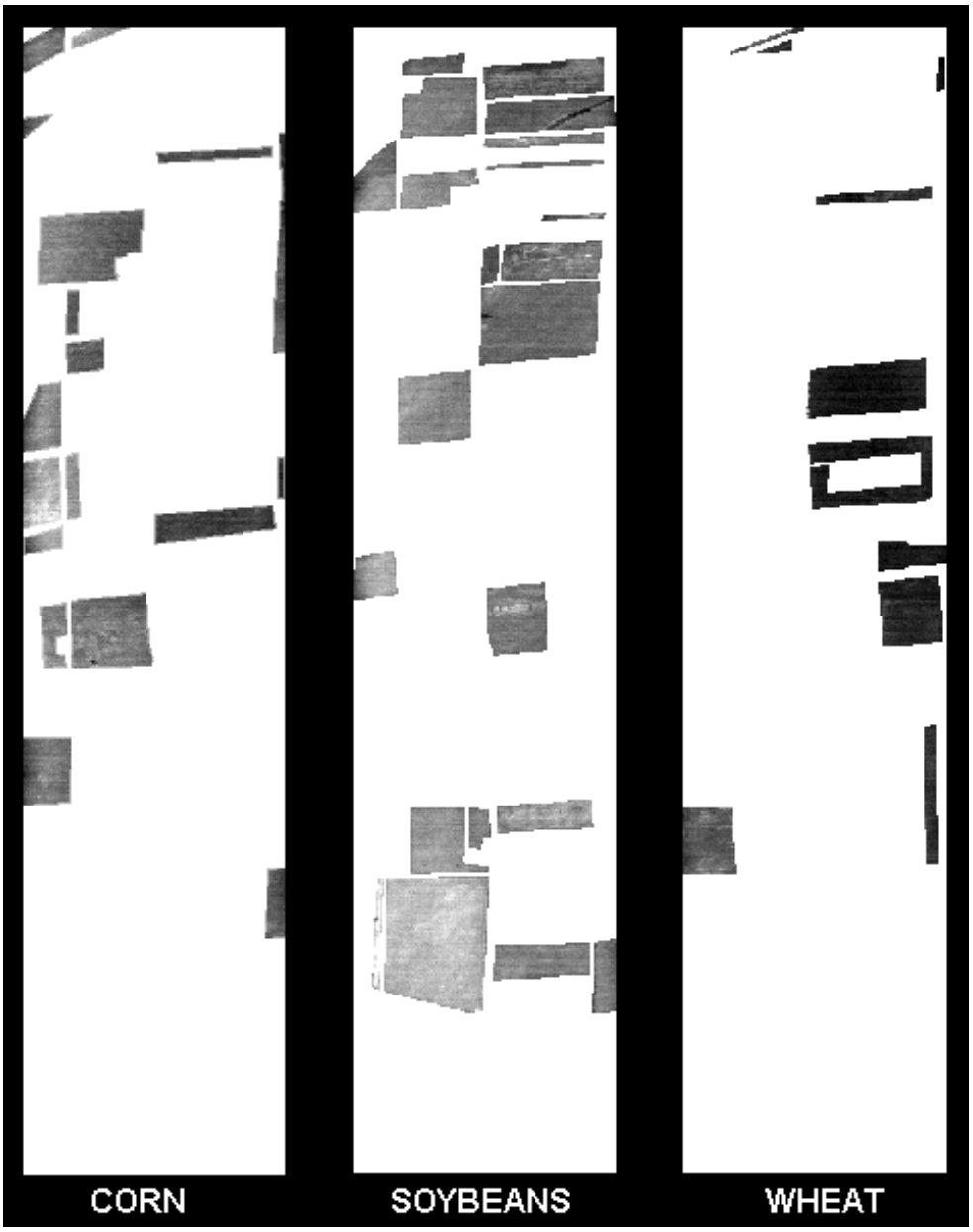

Figure 5.4 – Ground Truth Masks of corn, soybeans and wheat classes extracted from spectral band 1 image.

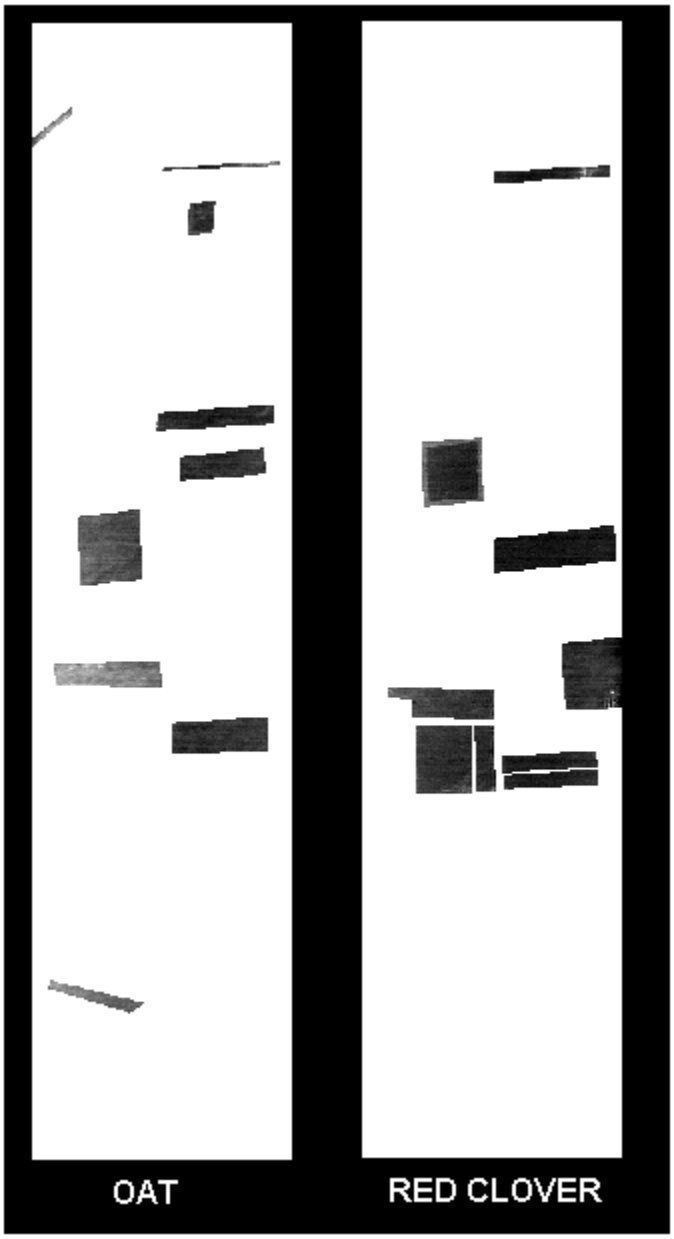

Figure 5.5 – Ground Truth Masks of oat and red clover classes extracted from spectral band 1 image.

CHAPTER 6

RESULTS AND DISCUSSIONS

6.1 Classification of the Data Set

In order to find the most discriminative features for the classification from the multispectral airborne image (FLC1), the developed algorithm based on the 2D Structured LDB approach is implemented. Flightline C1 (FLC1) data set is divided into five vegetative species: corn, soybeans, wheat, oat and red clover. In the first step, a binary spectral tree is generated and each spectral band of the image (12 bands between 0.40 μm – 1.0 μm) is placed at the bottom level (4^{th} level) of this spectral tree.

The images are placed based on their reflectance values from left to right. However, the bottom level of a binary tree with four depth consists of 16 nodes. Therefore, the remaining 4 nodes are assigned as null. Assigning null to these four remaining nodes sets their discriminative power as 0 (zero) and makes them ineffective in the pruning process. In

the spectral axis pruning process, the merged spectral features are averaged according to their energies before beginning the spatial-frequency axis pruning. Figure 6.1 shows the spectral pruning.

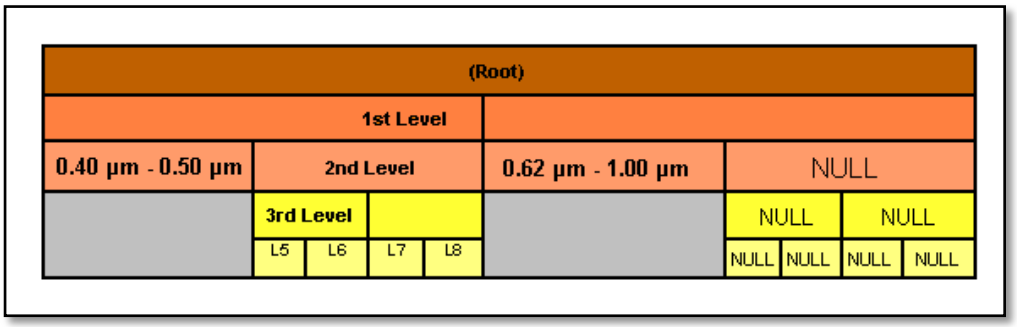

Figure 6.1 – The spectral bands between 0.40 µm – 0.50 µm and 0.62 µm – 1.00 µm are pruned (The last four remaining null nodes are ignored).

In the second step of the proposed algorithm, Daubechies 8 tap filter was used in order to form the two level wavelet decomposition quad trees. A two level wavelet quad tree consists of 21 subbands images. For the FLC1 12-band multispectral airborne dataset, total of 252 spatial-frequency patterns were placed on the spatial-frequency decomposition quad tree before the spatial axis pruning process.

The most discriminative features can be observed after completing both spectral and spatial-frequency pruning processes. In order to select the best discriminative features for a given class, two different preparations were made when assigning the fields as test and training sets. When preparing the test and training labels, two main rules are applied. The

first rule is that for the two different preparations, half of the labeled fields containing half of the total number of pixels are assigned as the training set and the other half is assigned as the test set. The second rule is that the test and training fields are completely different from each other. It means that none of the training labels are used for testing. Two types of pixel sampling methods were proposed for the formation of the test and training sets.

In the first formation of the test and training sets, every main labeled vegetation field which is extracted from the original ground truth image is divided into smaller sub fields. For example, a labeled rectangular corn field is divided into smaller rectangles. The training and test fields are selected randomly from these sub labels.

In the second formation of test and training set, all of the samples belonging to that class are divided into two parts without dividing the labels into sub labels. After that, one half (50% percentages of all samples) is assigned as the training set and the remaining half (remaining 50% percentage of all samples) is assigned as the testing set. The process is repeated two times until every single part is assigned as a testing set and the other three parts are assigned as the training set.

Training and testing set assignment procedure is applied to all other classes. The classification accuracy is the average from all results obtained from each binary classification. At the end, one feature map is

built for each training set of a given class. A resulting feature map of oat-red clover discriminative features after applying the first and second formation of the test and training sets are given in Figure 6.2 and Figure 6.3 respectively.

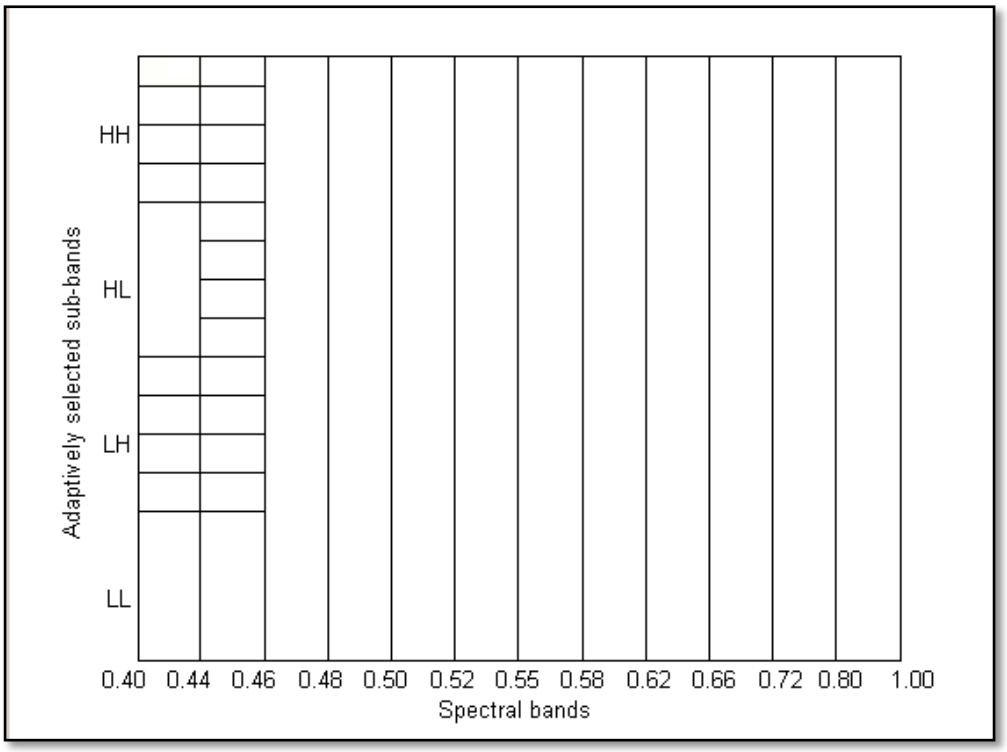

Figure 6.2 – An example “Feature Map” of oat - red clover discrimination features when using the first formation of the test and training set. The map shows the location of spectral bands and decomposed wavelet components.

The maximum number of candidate feature dimension for the lowest (fourth) level for a 12-band multispectral data is 192 (12 spectral bands x 16 frequency sub-bands). For the given oat-red clover feature map, the number of spectral-frequency feature dimension reduced from 192 to 33. The percentage of this dimension reduction gain is 82.81%. Although none of the spectral bands are pruned along spectral axis, every spatial frequency features except 0.40 μm - 0.44 μm and 0.44 μm - 0.46 μm are merged in spatial-frequency axis. The 0.40 μm - 0.44 μm spectral interval is decomposed into 12 spatial-frequency sub components and the 0.40 μm - 0.44 μm spectral interval is decomposed into 16 spatial frequency sub components. Another feature map of oat-red clover discriminative features is given in Figure 6.3.

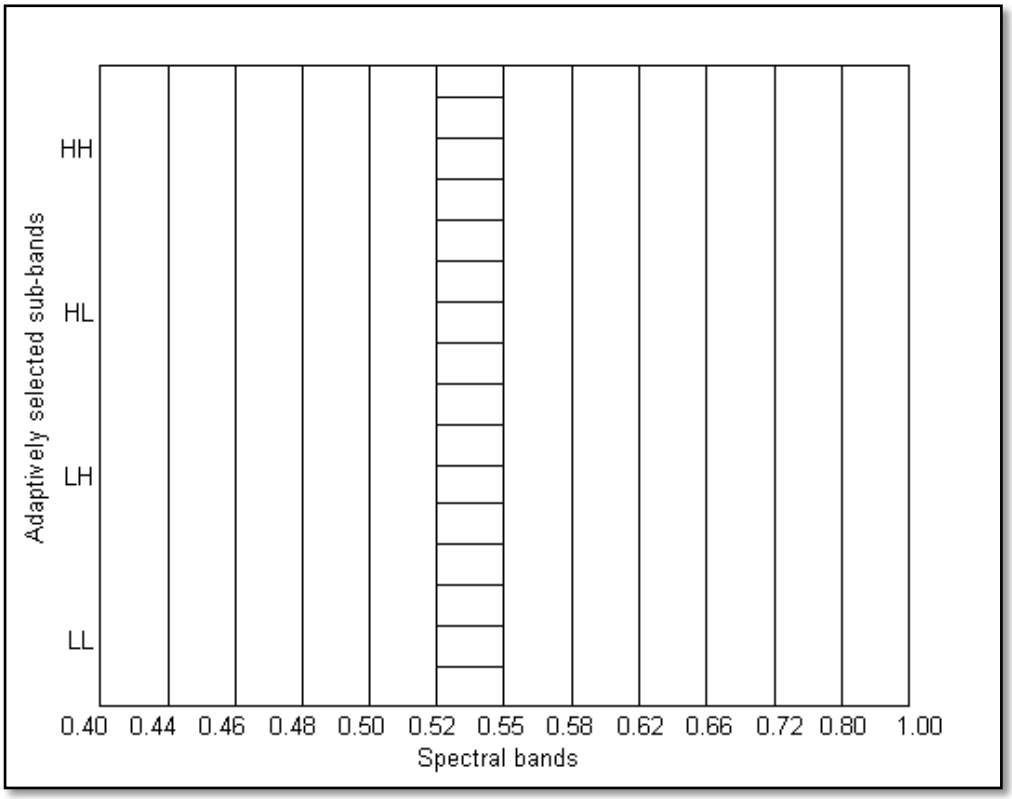

Figure 6.3 – An example “Feature Map” of oat - red clover discrimination features when using the second formation of the test and training set.

It can be observed from the Figure 6.3 that after the pruning operations, the number of spectral-frequency feature dimension reduced from 192 to 27. The percentage of this dimension reduction gain is 85.93% for the given oat-red clover feature map. Similar to the feature map in Figure 6.2, none of the spectral bands are pruned along spectral axis in this feature map and every spatial-frequency sub band except 0.52 μm - 0.55 μm wavelet band’s spatial-frequency components are merged in

spatial-frequency axis. The 0.52 μm - 0.55 μm spectral interval is decomposed into 16 spatial-frequency components.

From different training sets and different pairs of classes, the feature extraction phase delivers different numbers and types of discriminative features to the classification step. The implementation of the proposed algorithm generates a total number of 10 feature maps for the $\binom{5}{2}$ pairwise combinations where 5 is the selected number of vegetative classes from the data set. Some of these pairwise combinations are corn-soybeans, corn-wheat, soybeans-oat, wheat-red clover etc.

Each feature has a feature-score based on its discrimination powers. For the given feature maps of oat-red clover discriminative features in Figure 6.2 and Figure 6.3, the extracted 33 and 32 features are ranked using the Fisher Distance Based Feature Selection (FDB) algorithm. After the ranking process, these 33 and 32 features are sorted by the feature selection algorithm based on their discriminative powers and fed into the linear classifier. The process is repeated for all of the 10 pairwise class combinations. After applying the FDS algorithm, the features are given a color and a number. Both the darkness of the feature color and the number on each rectangular area in the feature tree show the feature's relevance. Figure 6.4 and Figure 6.5 show the spectral spatial-frequency feature map of oat-red clover and oat-red clover discriminative features given in Figure 6.2 and Figure 6.3 with their colors and their numbers on them ranked by FDS. All of the

generated and ranked spectral spatial-frequency feature maps with feature colors and feature numbers are given in Appendix A.

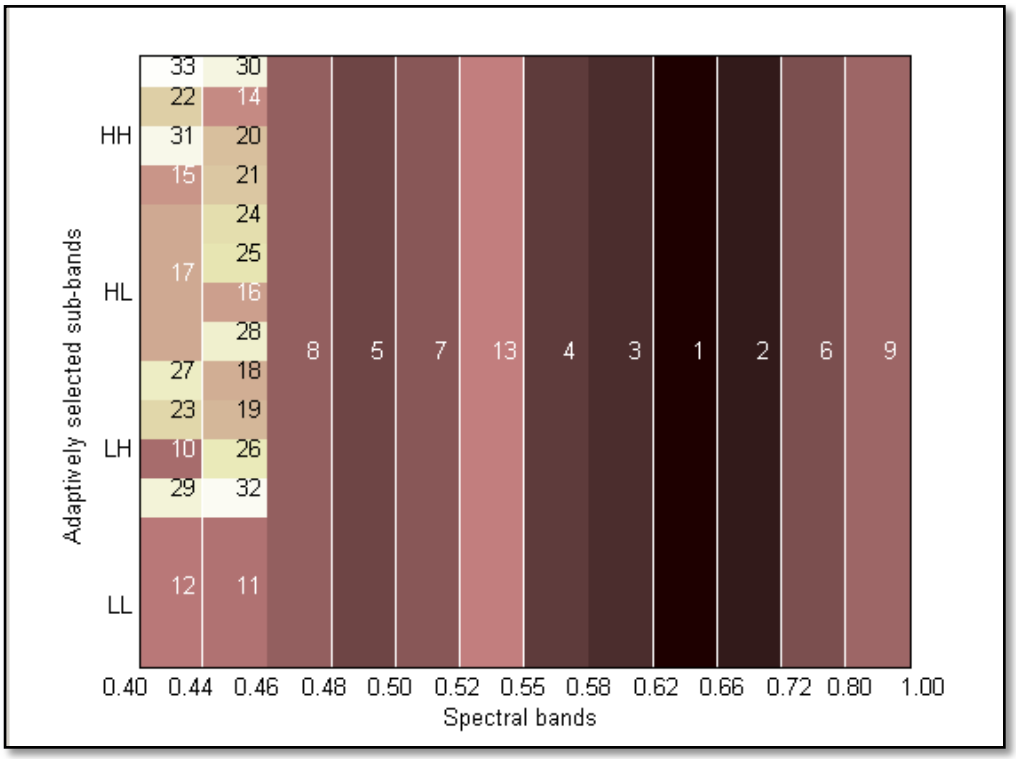

Figure 6.4 - Spectral Spatial-Frequency feature map of oat-red clover features ranked by the FDS algorithm when using the first type of test and training sets.

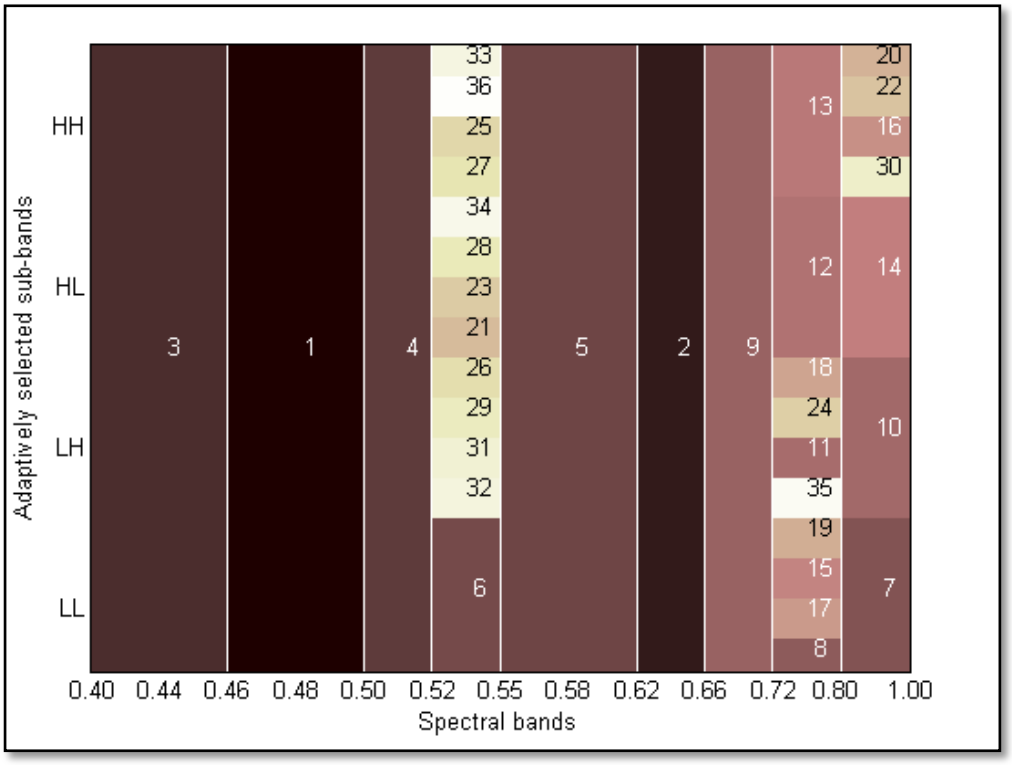

Figure 6.5 - Spectral Spatial-Frequency feature map of oat-red clover features ranked by the FDS algorithm when using the second type of test and training set.

In order to find the optimal number of features for the classification process, the spectral spatial-frequency features which are ranked by FDS are fed into the linear classifier one by one. This process is repeated for all 10 generated feature maps. After calculating the classification error percentage for each class, the mean error is calculated. The mean classification error curve is given in Figure 6.6.

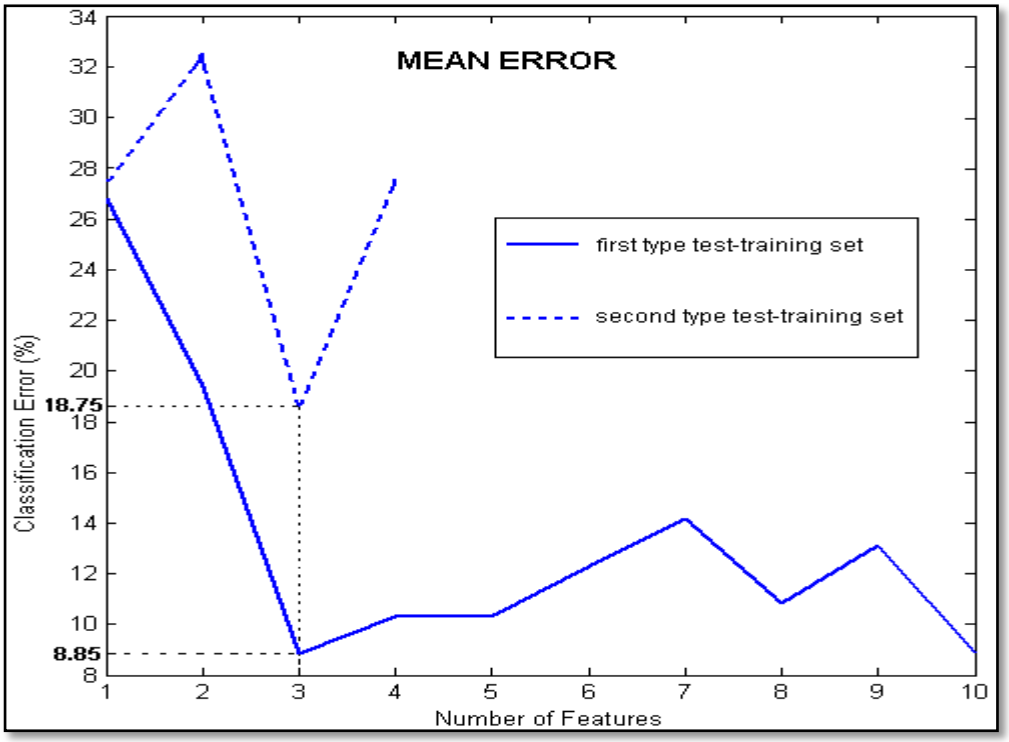

Figure 6.6 – The mean classification error curves of in percentage when using the first and second format of test and training sets. The graph shows the average of all pairwise classification results.

It can be seen from the Figure 6.6 that the minimum classification error with minimum number of features is 8.85%. This error rate is obtained only with 3 features using the first type of test and training data. The minimum classification error with minimum number of features is 18.75% when using the second type of test-training set. This error rate is also obtained with 3 features. The results are given in Table 6.1 below.

Table 6.1 - Mean errors obtained by using 3 features.

	Number of Features	**Classification Error (%)**
Proposed Algorithm (with first type of test-training set)	3	**8.85**
Proposed Algorithm (with second type of test-training set)	3	**18.75**

In order to compare the accuracy of the feature extraction algorithm with a well-known algorithm, the PCA algorithm is also implemented as an alternative feature extraction method and tested on the same data set. To do the comparison appropriately, the remaining LDA classifier part is completely identical with the LDB algorithm's implementation. The mean classification error provided by both feature extraction methods when using the first test-training formation is given in Figure 6.7.

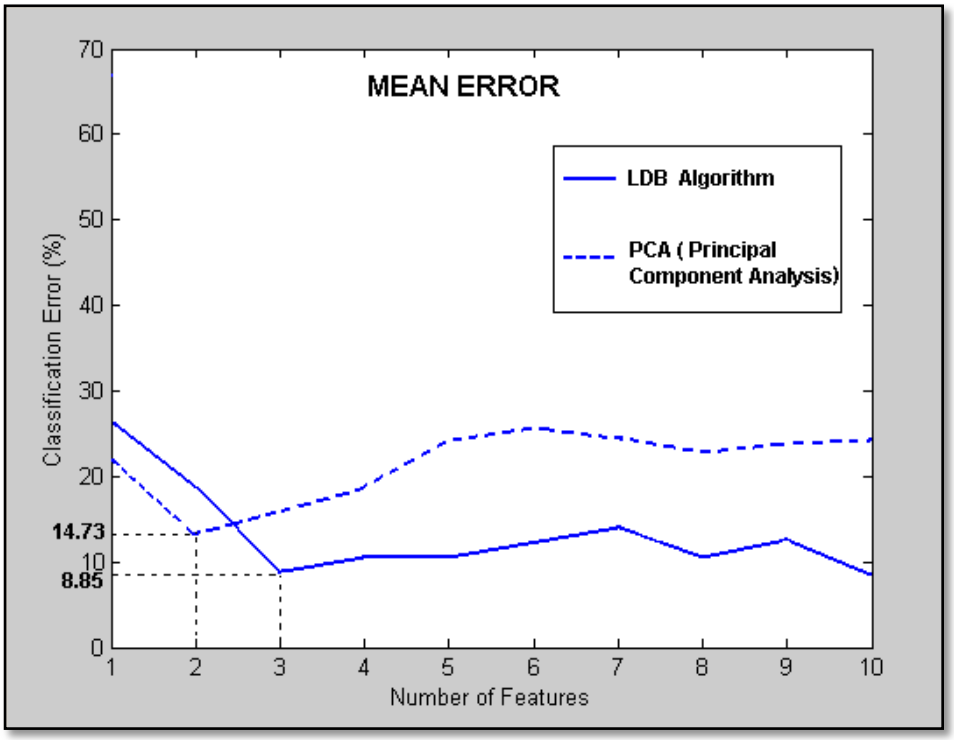

Figure 6.7 - The general mean classification error curves of in percentage when using the LDB algorithm for feature extraction and PCA method for feature extraction.

A total number of 66534 pixels are labeled as corn, soybeans, oat, wheat and red clover based on the ground truth images given in Figure 5.4 and Figure 5.5 in Chapter 5. The distribution of all the pixels to each class and all the wrong decisions made during the classification process when using the first and second test-training sets are given in two different decision matrices in Table 6.2 and Table 6.3.

Table 6.2 – Decision Matrix for the FLC1 data set when using 3 features with the first type of test-training set combination.

	True Class				
Classified As	Corn	Soybeans	Oat	Wheat	Red Clover
Corn	**11536**	**0**	**285**	**0**	**0**
Soybeans	**713**	**23193**	**208**	**0**	**0**
Oat	**182**	**0**	**7119**	**0**	**0**
Wheat	**0**	**0**	**190**	**11174**	**0**
Red Clover	**0**	**0**	**0**	**0**	**13512**

Number of Labeled Pixels Belong to that Class
Wrong Corn Decisions
Wrong Soybeans Decisions
Wrong Oat Decisions
Wrong Wheat Decisions
Wrong Red Clover Decisions

***All of the given numbers are representing the number of pixels for that decision.**

Table 6.3 – Decision Matrix for the FLC1 data set when using 3 features with the second type of test-training set combination.

	True Class				
Classified As	Corn	Soybeans	Oat	Wheat	Red Clover
Corn	**11536**	**0**	**493**	**1813**	**528**
Soybeans	**300**	**23193**	**585**	**0**	**1323**
Oat	**128**	**3308**	**7119**	**0**	**0**
Wheat	**0**	**0**	**0**	**11174**	**0**
Red Clover	**0**	**1768**	**0**	**2160**	**13512**

Number of Labeled Pixels Belong to that Class
Wrong Corn Decisions
Wrong Soybeans Decisions
Wrong Oat Decisions
Wrong Wheat Decisions
Wrong Red Clover Decisions

***All of the given numbers are representing the number of pixels for that decision.**

The decision matrix in Table 6.2 shows that approximately 1578 pixels out of 66534 pixels are classified incorrectly by the proposed algorithm by using 3 features for classification. The results are given for 3 features because the minimum mean error is obtained by using 3 features. The results show that approximately 713 pixels from corn class are misclassified as soybean, approximately 182 pixels from corn

class are misclassified as oat, approximately 502 pixels from oat class are misclassified as wheat, 285 pixels from oat class are misclassified as corn and 208 pixels from oat class are misclassified as soybeans. The proposed algorithm did not misclassify almost any of the pixels from soybeans wheat and red clover classes.

Also it can be seen from the given decision matrix in Table 6.3 that approximately 12406 pixels out of 66534 pixels are classified incorrectly by the proposed algorithm by using 4 features for classification when using the second type of test-training set formation. The results for the second type of test-train set are also given for 3 features because the minimum mean error is obtained using 3 features. The results show that total number of 3436 pixels from soybeans and corn classes are misclassified as oat. 3928 soybeans and wheat pixels are misclassified as red clover, 2834 oats, wheat and red clover pixels are misclassified as corn and 2151 corn, oat and red clover pixels are misclassified as soybeans. The proposed algorithm did not misclassify almost any of the wheat pixels. The classification maps which show the comparison of original ground truth maps and the classification maps based on the decision matrices in Tables 6.2 and 6.3 are given in Figure 6.8 and Figure 6.9 respectively.

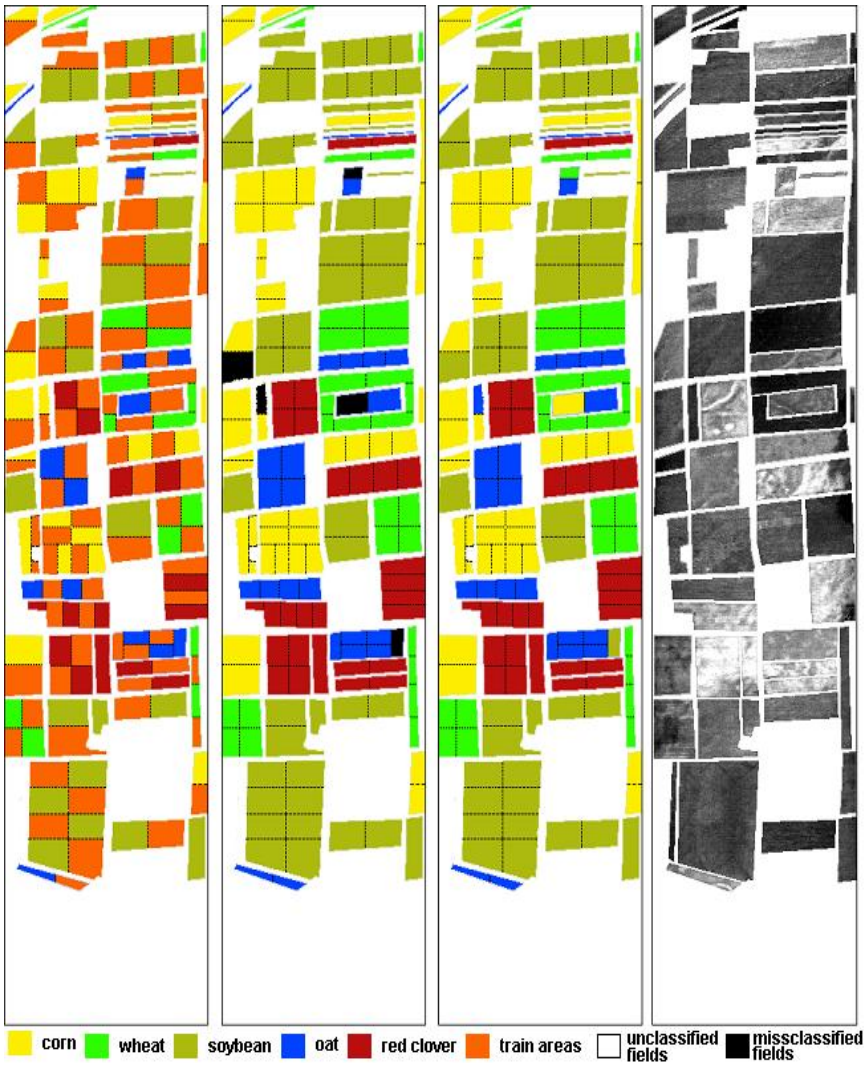

Figure 6.8- The comparison of original ground truth map belonging to the 12th spectral band and the classification maps based on these results is given decision matrix in Table 6.2. First image shows the training fields. Misclassified areas are shown in black color in the second image. The third image shows the misclassified areas with the misclassified vegetation class color on them.

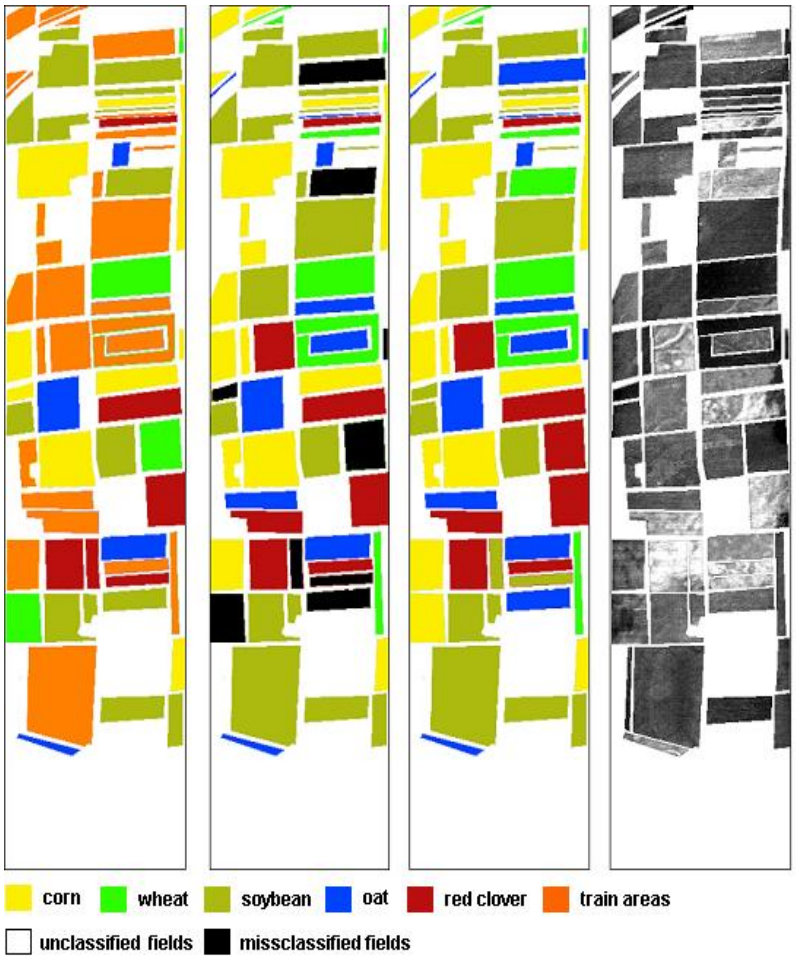

Figure 6.9- The comparison of original ground truth map belongs to the 12th spectral band and the classification maps based on these results is given decision matrix in Table 6.3. First image shows the training fields. Misclassified areas are shown in black color in the second image. The third image shows the misclassified areas with the misclassified vegetation class color on them.

Obtaining the best spatial spectral-frequency decomposition of the data and bands to be retained for classification is a computationally demanding step. In order to classify M-ary data, this binary step has to be carried out for $\binom{M}{2}$ times. When the first 3 features providing the best classiffication rate is employed, all 178 samples need to be processed. In this case, the computations for feature extraction and classification require 2102.5 msec on a dedicated Intel Core Duo 1.66GHz processor. It can be concluded that a single labelled sample is processed in approximately 11.3 msec. The computation times are calculated by using the “cputime” function of Matlab 7.1 software.

CHAPTER 7

CONCLUSIONS

The 2D Local Discriminant Bases (LDB) algorithm used to classify remotely sensed data. The LDB algorithm extracts the relevant features by removing the irrelevant ones and/or combining the ones together which do not contain supplemental information on their own. The aim of our study is reaching the minimum mean error with fewer numbers of features. In order to test this goal, the 2D LDA classifier is modified to M-ary form by combining the Majority Voting Principle and the linear distance parameters to classify more than two classes appropriately in the remotely sensed data. Better classification accuracies are predicted to be obtained by using more complex classifiers like other non-linear classifiers, Support Vector Machines or Neural Networks with extracted features. Also one can try to develop his own adaptive [61] or estimation based classifiers [68] for this kind of problems where limited numbers of training samples are available.

The proposed method is implemented on a multispectral airborne data set, also known as Flight Line C1 (FLC1), from Tippecanoe County,

Indiana to evaluate its performance. The spectral and spatial-frequency features are extracted from the image and used for classifying corn, soybeans, red clover, wheat, and oat in the data set. The motivation behind this study is enabling a low cost but highly accurate remote sensing system which uses fewer numbers of bands with fewer amounts of data.

During the experimental study, in order to select the best discriminative features for a given class, two different preparations were made when assigning the fields as test and training sets. We used half of the labeled fields that contain the half of the total number of pixels are assigned as training set and the other half is assigned as test set for each preparation. Also, we completely separate the test and training fields from each other. The feature extraction step generates a total number of 20 feature maps where 10 of the feature maps resulted from using the first type of training set and the other 10 feature maps are belong to the second type of training set.

In the best case scenario, our proposed algorithm reduced the number of spectral spatial-frequency features from 192 to 10 when generating the feature map for soybeans and red clover using the fist type of training set. The result means that both of the classes can be easily separated from each other using fewer numbers of features. Additionally, it can be understood from the generated feature map of soybeans and red clover that because the feature map was not divided

into any spatial-frequency component, both classes can be separated from each other only by using spectral features.

In the first formation of the test and training set, every labeled vegetation field extracted from the original ground truth image is divided into smaller sub fields. This kind of test and train data combination gives the advantage of testing our algorithm with characteristically similar pixels included in the training set. Therefore, this train-test set formation provided a small classification mean error which is 8.85%. On the other hand, the second formation of the test-training set only contains the main labeled fields. The fields were not divided into any smaller sub fields. The correlation between the characteristics for this test and train set is a lot smaller than the previously prepared test-training set. This situation leads to a bigger mean error for classification which is 18.75%.

When using the first training set formation, total number 1578 pixels out of 66534 are misclassified. Expectedly, when the second training set is used for classification, the number of misclassified pixels is increased to 12406. Although both of the training sets reach the minimum classification mean error by using only 3 features, the reason behind this 10% difference between two mean errors is the correlation of pixel characteristics between the test and train sets. An almost similar study was proposed in [74], using a similar five-class data set approach. In this study, using Binary Tree Edge Sensing Demosaicking (BTES) and Binary Tree Bilinear Interpolation (BTBI) methods, the obtained mean errors for classification when using 3 features were 21.57% and 22.21%

respectively. Although their classification methods are different, the results show that our mean errors when using the same amount of test and training pixels and same number of features is smaller. This shows that instead of using only spectral features, combining the information both along the spectral and spatial-frequency axes and selecting the most discriminative ones can provide better performance for classification. Also, one can try to use the BTES and BTBI methods as a classifier after using 2D LDB algorithm and compare the results.

The PCA algorithm is also implemented as a feature extraction method to compare the classification results with the LDB algorithm. To do the comparison appropriately, the LDA classifier part remains completely identical with the LDB algorithm's implementation. Because the PCA method converts the correlated variables into a small size of uncorrelated variables by using Karhunen-Loeve transformation and only first a few components which represent the whole data set are selected and used for classification, it is expected that until the 3rd feature where minimum mean error is reached for LDB, the PCA can provide smaller classification mean error. However, when the 3rd feature is added to the LDA classifier, the LDB provides smaller mean error than PCA. The results show that the mean error of LDB is 8.85% whereas the PCA algorithm gives an error of 14.73% when both of them are used for feature extraction.

The feature extraction algorithm can be applied to different kinds of hyperspectral/multispectral data with different characteristics to observe the accuracy of the algorithm in different fields of study. During this study, the proposed methods are also tried using the hyperspectral HYDICE Washington DC Mall data set which is obtained by .an airborne flight line over Washington DC Mall. This data is provided with the permission of Spectral Information Technology Application Center of Virginia who was responsible for its collection. The useful part of the data, however, was very limited because the visual bands are corrupted by noise. In addition; only a small portion of the data is annotated. We attempted to annotate the remaining parts by hand but possible errors at this stage have the potential of misleading the train stage. Yet promising results are obtained with the limited trials. Reliable prior information is needed before these results can be reported.

7.1 Future Works

A single LDA classifier is selected for the analysis of this multispectral remotely sensed data set because the main focus of this study is extracting the most discriminative features for classification purposes and emphasizing the power of this feature extraction algorithm by using with a basic classifier. On the other hand, the overall performance of the proposed algorithm can be increased by using more complex classifiers like SVM, Neural Networks, and other non-linear classifiers.

In our study, the correlation between features is not taken into consideration when extracting the discriminative features. From that perspective, the feature extraction algorithm can be modified to get the most discriminative and also independent features by considering the correlation between them. Furthermore, only filtering technique is considered here, but the wrapper technique could also be used.

Different feature selection algorithms can be used or modified specifically for this problem in order to find the best results and prove the robustness of the algorithm by comparing the accuracy of each method.

For the wavelet decomposition step, Daubechies 8 tap filter was used. One can also use different filters for decomposing the spectral bands into the wavelet subbands.

The feature extraction algorithm can be applied to different kinds of hyperspectral/multispectral data with different characteristics to observe the accuracy of the algorithm in different fields of study. Obtaining reliable hyperspectral data with accurate classification of the spatial regions in the data set is necessary. Unfortunately, such data is rarely publicly available.

REFERENCES

[1] Toomarian, N. B., Kohen, H., and Gat, N., 1998. Neural Network for Analysis of Hyperspectral Imagery, *International Conference on Neural Networks and their Applications, Marseilles, France.*

[2] Martinez-Uso, A., Pla, F., Sotoca, J., and Garcia-Sevilla, P., 2008. From Narrow to Broad Band Design and Selection in Hyperspectral Images. *Lecture Notes in Computer Science, Springer-Verlag,* 5112 LNCS*,* 1091-1100.

[3] Bellman, R., 1961. Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton, New Jersey.

[4] Hughes, G.F., 1968. On mean accuracy of statistical pattern recognizers, *IEEE Transactions on Information Theory*, 14 (1), 55–63.

[5] Hsu, P., 2004. Feature Extraction of HyperSpectral Images Using Matching Pursuit, *Proceedings of the XXth ISPRS Congress, Istanbul*.

[6] Landgrebe, D.A., 2003. Signal Theory Methods in Multispectral Remote Sensing, *John Wiley & Sons.*

[7] Jain, A., Duin, R., and Mao, J., 2000. Statistical Pattern Recognition: A Review, *IEEE Trans. Pattern Analysis and Machine Intelligence,* 22(1), 4-37.

[8] Hsieh, P.F., Wang, D.S., and Hsu, C.W., 2006. Pattern Classification - A Linear Feature Extraction for Multiclass Classification Problems Based on Class Mean and Covariance Discriminant Information, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 28 (2), 223.

[9] Lee, H.M., Chen, C.M., Chen, J.M., and Lou, Y.L., 2001. An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy, *IEEE Transactions on Systems Man and Cybernetics Part B Cybernetics,* 31 (3), 426-432.

[10] Thawonmas, R., and Abe, S., 1997. A Novel Approach to Feature Selection Based on Analysis of Class Regions, *IEEE Transactions on Systems Man and Cybernetics Part B Cybernetics,* 27 (2), 196-207.

[11] Kalkan, H., 2008. Feature Extraction from Acoustic and Hyperspectral Data by 2D Local Discriminant Bases Search, Unpublished Doctoral Dissertation, Middle East Technical University, Ankara.

[12] Fauvel, M., Chanussot, J., and Benediktsson, J.A., 2009. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas, *Eurasip Journal on Advances in Signal Processing,* 2009, 1-14.

[13] Kwon, H. and Nasrabadi, N.M., 2007. A comparative analysis of kernel subspace target detectors for hyperspectral imagery, *Eurasip Journal on Applied Signal Processing,* 2007, 13.

[14] Hsieh, P.F. and Landgrebe, D., 1998. Linear Feature Extraction for Multiclass Problems, *International geoscience and remote sensing symposium,* 4, 2050-2052.

[15] Fisher, R. A., 1936. The use of multiple measurements in taxonomic problems, *Annals of Eugenics,* 7, 179-188

.

[16] Rao, C.R., 1948. The Utilization of Multiple Measurements in Problems of Biological Classification*, Journal of Royal Statistical Society, Series B,* 10, 159-203.

[17] Fukunaga, K., 1990. Introduction to statistical pattern recognition, *Computer science and scientific computing, Boston: Academic Press.*

[18] Gokce, I. and Peng, J., 2002. Comparing Linear Discriminant Analysis and Support Vector Machines, *Springer-Verlag,* 2457 LNCS*,* 104-113.

[19] Du, Q., 2007. Modified Fisher's Linear Discriminant Analysis for Hyperspectral Imagery, *IEEE geoscience and remote sensing letters,* 4 (4), 503-507.

[20] Liang, S., 1998. Land Cover Classification Methods for Multiyear AVHRR Data, *International geoscience and remote sensing symposium,* 5, 2521-2524.

[21] Feng, G., Hu, D., Li, M. and Zhou, Z., 2005. A Novel LDA Approach for High-Dimensional Data, *Lecture Notes in Computer Science, Springer-Verlag,* 3610, 209-212.

[22] Jolliffe, I., 2002. Principal Component Analysis, second ed, *Springer.*

[23] Jia, X. and Richards, J., 1999. Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification, *IEEE Transactions on Geoscience and Remote Sensing,* 37(1), 538-542.

[24] Hsu P. H. and Tseng, Y. H., 1999. Feature extraction for hyperspectral image, *Proc. 20th ACRS*, 1, 405-410.

[25] Hsu, P. and Tseng, Y., 2002. Feature Extraction of Hyperspectral Data Using the Best Wavelet Packet Basis, *Proceedings of the IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, Toronto.

[26] Bruce, L. M. Koger, C., H., and Li, J., 2002. Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction, *IEEE Transactions on Geoscience and Remote Sensing,* 40, 2331-2338.

[27] Mallat S.G., 1989. A theory for multi-resolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674-693.

[28] Michalak, K. and Kwasnicka, H., 2006. Correlation based feature selection strategy in classification problems*, International Journal of Applied Mathematics and Computer Science,* 16(4), 503-511.

[29] Das, S., 2001. Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection, *Machine Learning – International Workshop Then Conference,* 18, 74-81.

[30] Vega, A. and Manian, V., 2001. Comparison of PCT and Fisher Discriminant Analysis for Texture Feature Selection, *Computing Research Conference*, *Mayaguez, Puerto Rico.*

[31] Hall, M. A., 1999. Correlation-based Feature Selection for Machine Learning, *Unpublished Ph.D. Dissertation, The University of Waikato, Hamilton, New Zeland*.

[32] Duch, W., 2006. Filter Methods, *Studies in Fuzziness and Soft Computing*, 207, 89-118.

[33] Coifman, R. and Wickerhauser, M.V., 1992. Entropy-based Algorithms for Best Basis Selection, *IEEE Trans. Inform. Theory*, 38, 713-719.

[34] Capobianco, E., 2001. Independent multiresolution component analysis and matching pursuit, *Report PNA, 11, Amsterdam: CWI.*

[35] Kalkan, H. and Yardimci Y., 2008. Extraction of Discriminative Features from Hyperspectral Data, *IEEE International Conference on Data Mining Workshops 2008,* 414-419.

[36] Kumar, S., Ghosh, J. and Crawford, M., 2001. Best-Bases Feature Extraction Algorithms for Classification of Hyperspectral Data, *IEEE Trans Geoscience and Remote Sensing,* 39, 1368-1379.

[37] Saito, N. and Coifman, R.R., 1994. Local Discriminant Bases, in Mathematical Imaging: Wavelet Applications in Signal and Image Processing II, *Proceedings of the SPIE*, 2003, 2-14.

[38] He, P., 2005. Classification methods and applications to mass spectral data, *Unpublished Doctoral Dissertation, Hong Kong Baptist University, Kowloon Tong, Hong Kong*.

[39] Lee, C. H. and Yang, H. C., 2009. Construction of supervised and unsupervised learning systems for multilingual text categorization, *Expert Systems With Applications,* 36(2), 2400-2410.

[40] Duda, R. and Hart, P., 1973. Pattern Classification and Scene Analysis, *John Wiley & Sons, Inc*.

[41] Manandhar, R., 2008. Land Cover Classification from Remote Sensing Imagery: Revisiting and Reevaluating Classification Accuracy, *Proceedings of the 11th Annual Map India Conference.*

[42] Lu, D. and Q. Weng., 2007. A survey of image classification methods and techniques for improving classification performance, *International Journal of Remote Sensing,* 28(5), 823-870.

[43] Wilkinson, G.G., 2005. Results of Implications of a Study of Fiteen Years of Satellite Classification Experiments, *IEEE Transaction on Geoscience and Remote Sensing,* 43(3), 433-440.

[44] Foody, G.M., 2002. Status of Land Cover Classification Accuracy Assessment, *Remote Sensing of Environment,* 80, 185-201.

[45] Bischof, H., Schneider W. and Pinz, A. J., 1992. Multispectral classification of Landsat-images using neural networks, *IEEE Transactions on Geoscience and Remote Sensing,* 30(3), 482-490.

[46] Carmel, Y. and Kadmon, R., 1998. Computerized classification of Mediterranean vegetation using panchromatic aerial photographs, *Journal of Vegetation Science,* 9 (3), 445-454.

[47] Tsai, F., 2002. A derivative-aided hyperspectral image analysis system for land-cover classification, *IEEE Transactions on Geoscience and Remote Sensing,* 40 (2), 416-425.

[48] Fung, T. and Chan, K., 1994. Spatial composition of spectral classes: a structural approach for image analysis of heterogeneous land-use and land-cover types, *Photogrammetric* Engineering *and* Remote Sensing*,* 60 (2), 173-180.

[49] Govender, M. , Chetty, K. and Bulcock, H., 2007. A review of hyperspectral remote sensing and its application in vegetation and water resource studies, *Water SA,* 33(2), 145–151.

[50] Smith, R.B., 2001. Introduction to remote sensing of the environment, *retrieved december 12, 2009 from* http://www.microimages.com/getstart/pdf/introrse.pdf.

[51] Landgrebe D., 1999. Some Fundamentals and methods for hyperspectral image Data Analysis, *Proceedings of SPIE Int. Symp. On Biomedical Optics (Photonics West), San Jose CA,* 3603, 104-113.

[52] Lee, K.S., Cohen, W. B., Kennedy, R. E., Maiersperger, T. K. and Gower, S. T., 2004. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes, *Remote Sensing of Environment,* 91 (3), 508.

[53] Kumar, S., Ghosh, J. and Crawford, M., 2000. A Hierarchical Multiclassifier System for Hyperspectral Data Analysis, *Lecture Notes in Computer Science,* 1857, 270-279.

[54] Hsu, P.H., Yi-Hsing, T. and Gong, P., 2002. Dimension reduction of hyperspectral images for classification applications, *Geographic Information Sciences*, 8(1), 1–8.

[55] Su, J. and Ning, S. 2008. A Dimensionality Reduction Algorithm of HyperSpectral Image Based on Fract Analysis, *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences,* 37 (1), 297-302.

[56] Jimenez, L. and Landgrebe, D., 1999. Hyperspectral Data Analysis and Feature Reduction Via Projection Pursuit, *IEEE Transactions on Geo- science and Remote Sensing,* 37(6), 2653-2667.

[57] Pal, M., 2009. Margin-based feature selection for hyperspectral data, *International Journal of Applied Earth Observation and Geoinformation,* 11 (3), 212-220.

[58] Sandhya Borra, B.E., 2003. Optimized Band Elimination and Dimensionality Reduction of Hyperspectral Images, Unpublished Doctoral Dissertation, *Texas Tech University, Texas.*

[59] Tian, Y., Guo, P. and Lyu, M.R., 2005. Comparative Studies on Feature Extraction Methods for Multispectral Remote Sensing Image Classification, *IEEE International Conference on Systems, Man and Cybernetics,* 2, 1275-1279.

[60] Aksoy**,** S., 2006. Spatial Techniques for Image Classification, in C.H. Chen, ed., *Signal and Image Processing for Remote Sensing*, 491-513, Taylor & Francis.

[61] Jackson, Q. and Landgrebe, D. A., 2001. An Adaptive Classifier Design for High-Dimensional Data Analysis With a Limited Training Data Set, *IEEE Transactions on Geoscience and Remote Sensing,* 39, 2664-2679.

[62] Karakahya, H., Yazgan, B. and Ersoy, O. K., 2003. A Spectral-Spatial Classification Algorithm for Multispectral Remote Sensing Data, *Lecture Notes in Computer Science,* 2714, 1011-1017.

[63] Haralick, R. M. and Shapiro, L. G., 1992. *Computer and robot vision, Addison-Wesley Pub.*

[64] Evans, C., Jones, R., Svalbe, I. and Berman, M., 2002. Segmenting Multispectral Landsat TM Images Into Field Units, *IEEE Transactions on Geoscience and Remote Sensing,* 40, 1054-1064.

[65] Sarkar, A., Biswas, M. K., Kartikeyan, B., Kumar, V., Majumder, K. L., and Pal, D. K., 2002. A MRF Model-Based Segmentation Approach to Classification for Multispectral Imagery, *IEEE Transactions on Geoscience and Remote Sensing,* 40(5), 1102-1113.

[66] T. Blaschke., 2003. Object-based contextual image classification built on image segmentation, i*n Proceedings of IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data,Washington, DC,* 113–119.

[67] Brodley C. E., Friedl M. A. and Strahler A., 1992. New Approaches to Classification in Remote Sensing Using Homogeneous and Hybrid Decision Trees to Map Land Cover, *Proc. IGARSS,* 532-534.

[68] Dundar, M. M. and Landgrebe, D. A., 2004. Toward an Optimal Supervised Classifier for the Analysis of Hyperspectral Data, *IEEE Transactions on Geoscience and Remote Sensing,* 42, 271-277.

[69] Cetin, A., Pearson, T. and Tewfik, A., 2004. Classification of Closed And Open Shell Pistachio Nuts Using Voice Recognition Technology, Transaction of the ASAE, 47(2), 659-664.

[70] Yang, H., Vuuren S.V. and Hermansky, H., 1999. Relevance of Time-Frequency Features for Phonetic Classification Measured by Mutual Information, *Proceeding of IEEE ICASSP.*

[71] Dharwarkar, G., 2005. Using Temporal Evidence and Fusion of Time-Frequency Features for Brain-Computer Interfacing, *Ms Thesis, Waterloo,Ontario, Canada.*

[72] Ghassemian H. and Landgrebe D.A., 1988. On-Line Object Feature Extraction for Multispectral Scene Representation, *TR-EE 88-34., Purdue University, West Lafayatte.*

[73] Swain, P.H. and Davis, S.M., 1978. Remote Sensing: The Quantitative Approach, *McGraw Hill, Chapter 2.*

[74] Miao, L., Qi, H., Ramanath, R. and Snyder, W.E., 2006. Binary tree-based generic demosaicking algorithm for multispectral filter arrays, *IEEE Transactions on Image Processing : a Publication of the IEEE Signal Processing Society,* 15 (11), 3550-3558.

APPENDIX

Appendix A Generated Spectral Spatial-Frequency Feature Maps

All of the generated and ranked spectral spatial-frequency feature maps with feature colors and feature numbers are given below for the first type of test-training set.

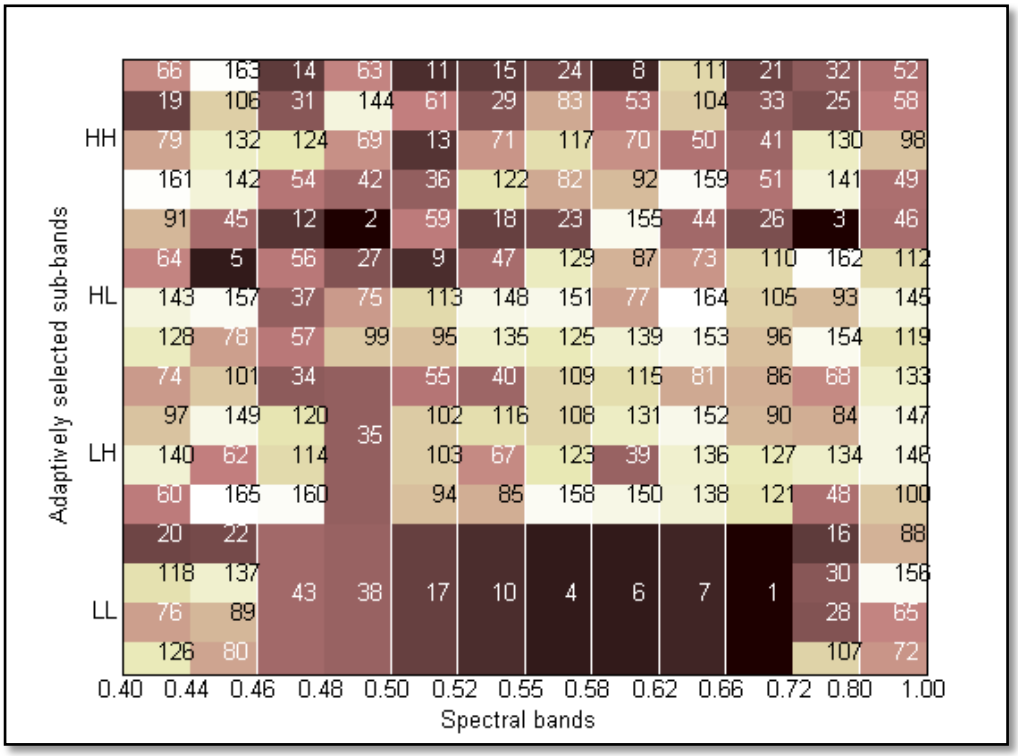

Figure 1 – 165 Spectral Spatial-Frequency Features located in the corn-soybeans feature map ranked by FDS algorithm.

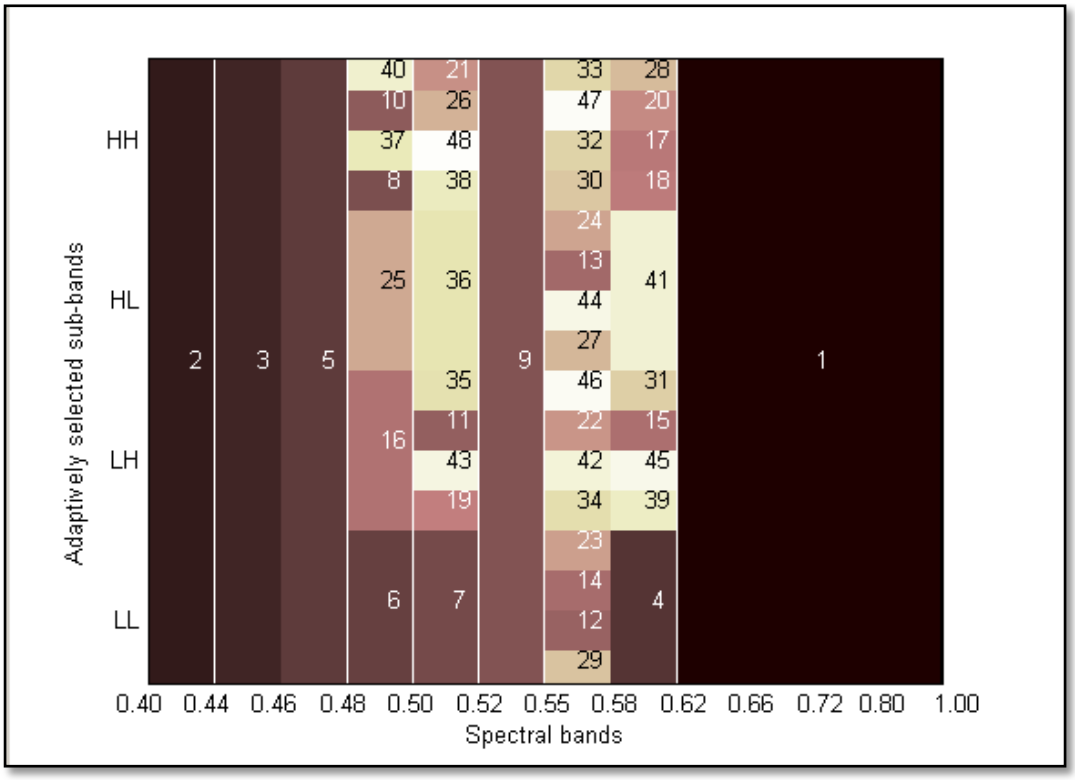

Figure 2 – 48 Spectral Spatial-Frequency Features located in the corn-oat feature map ranked by FDS algorithm.

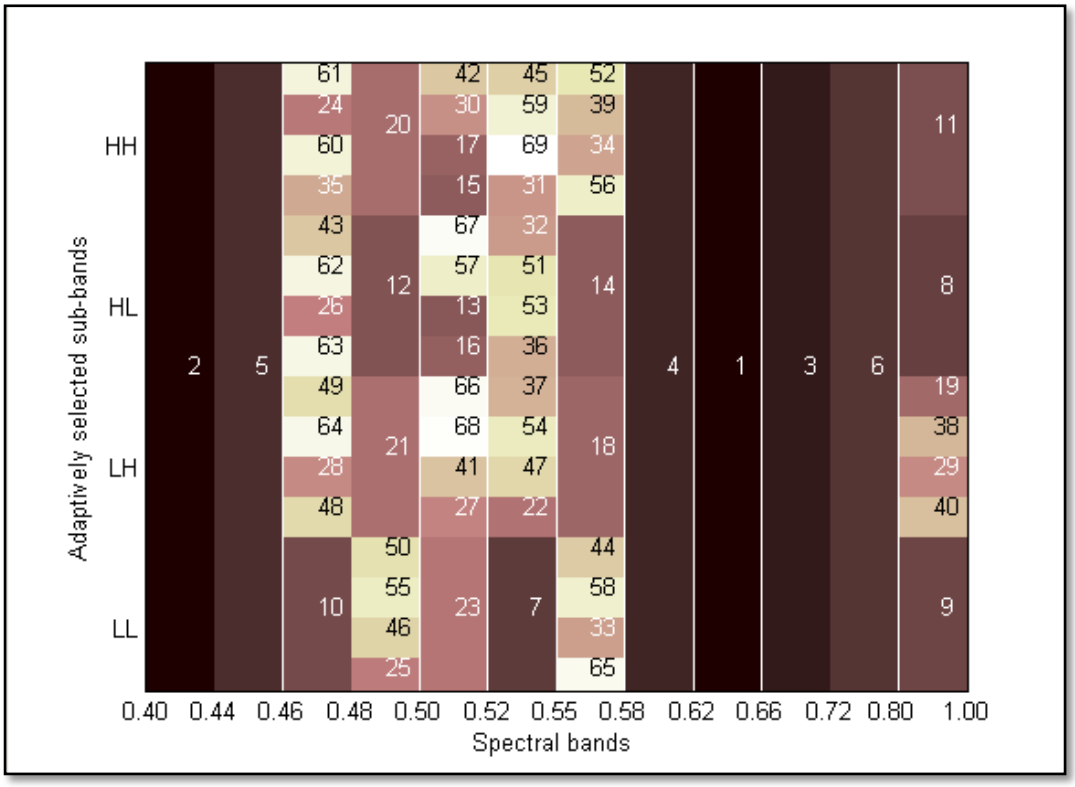

Figure 3 – 69 Spectral Spatial-Frequency Features located in the corn-wheat feature map ranked by FDS algorithm.

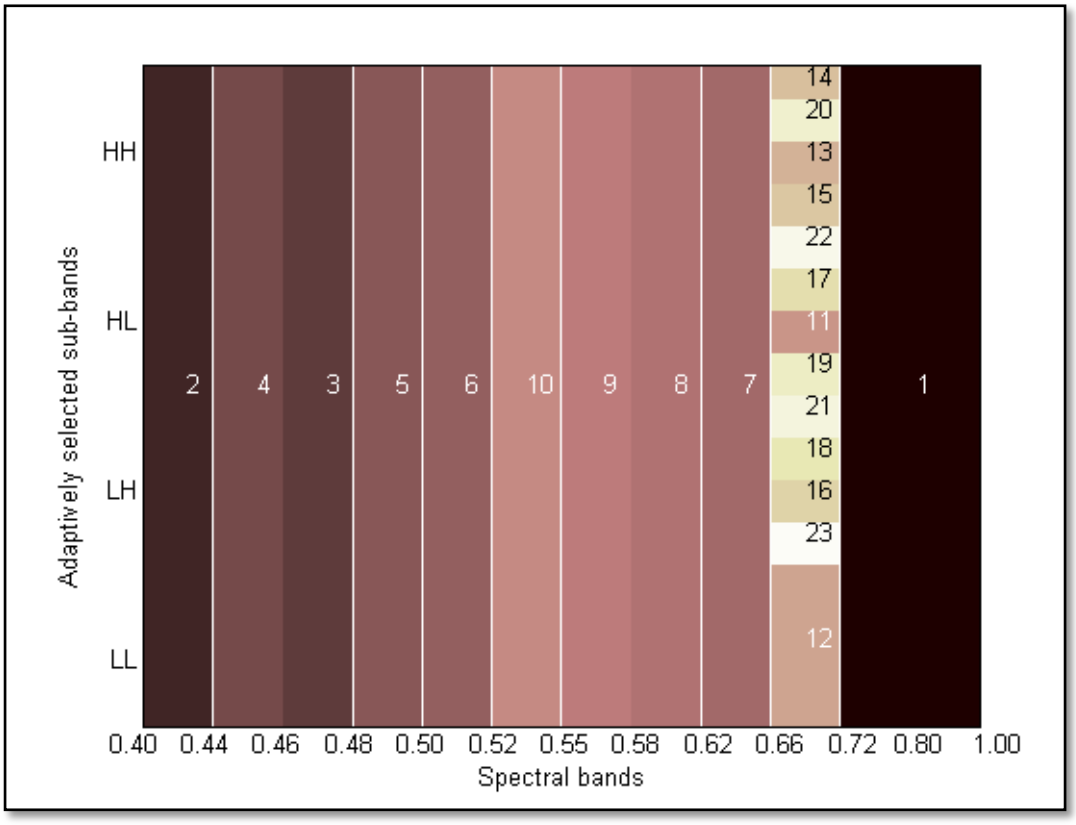

Figure 4 – 23 Spectral Spatial-Frequency Features located in the feature map of corn-red clover features ranked by FDS algorithm.

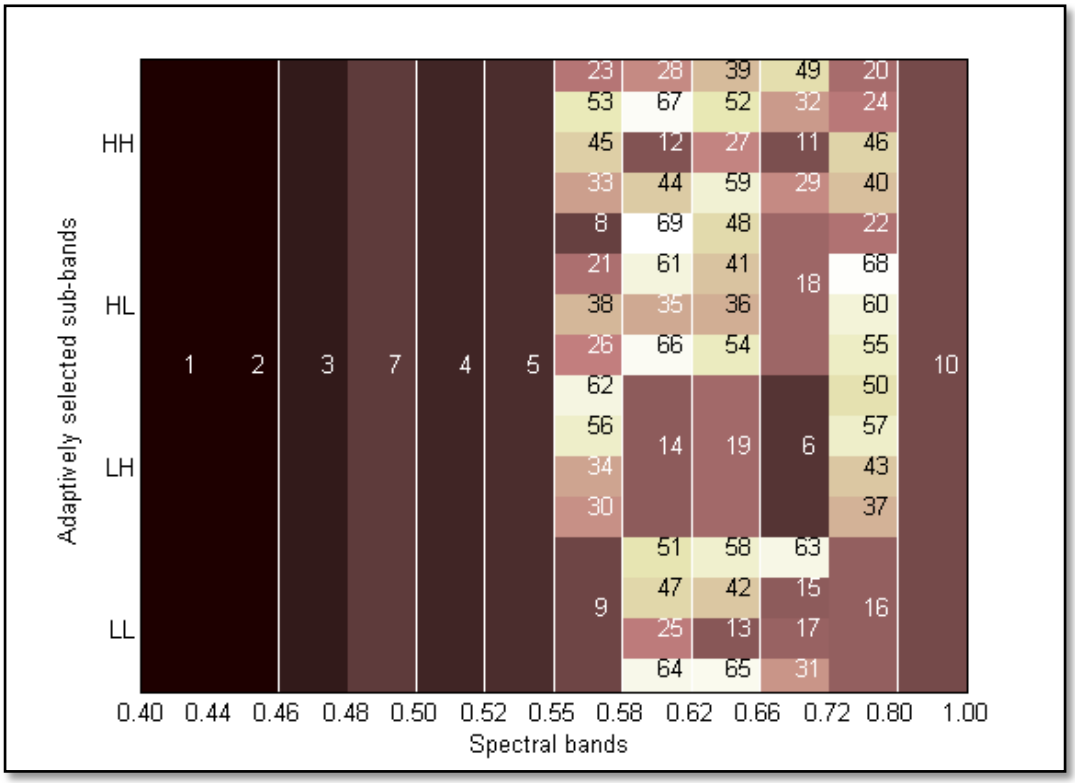

Figure 5 – 69 Spectral Spatial-Frequency Features located in the soybeans-oat feature map ranked by FDS algorithm.

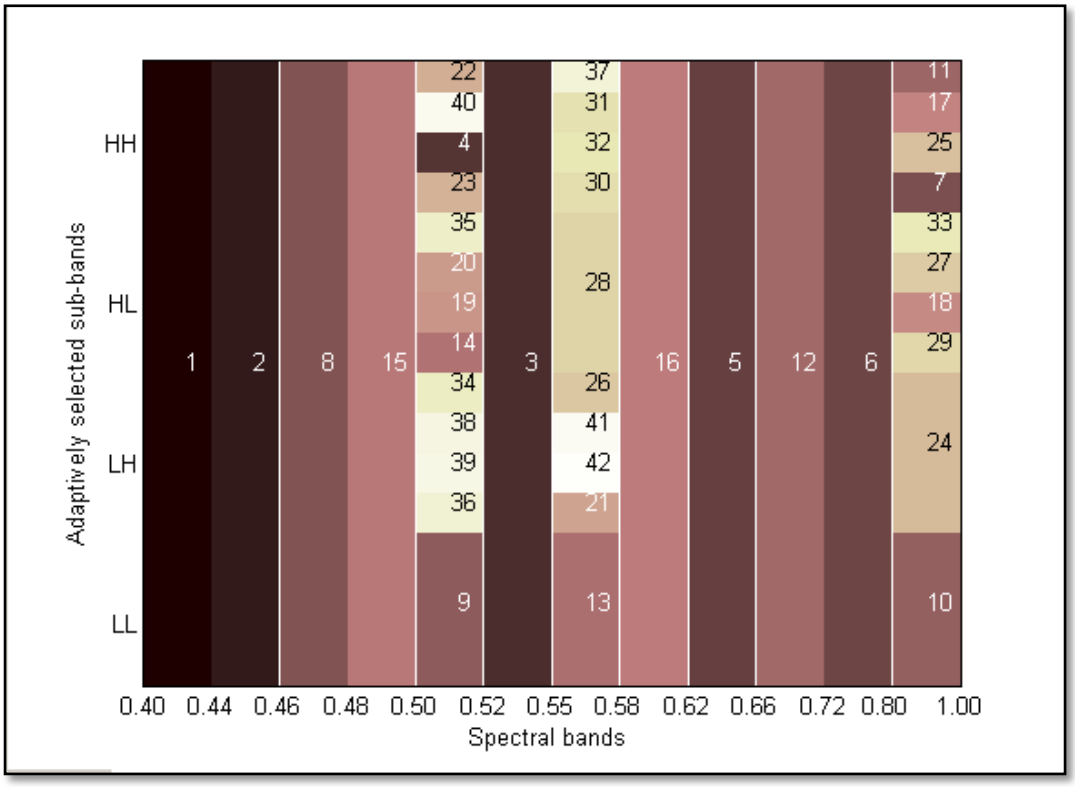

Figure 6 – 42 Spectral Spatial-Frequency Features located in the soybeans-wheat feature map ranked by FDS algorithm

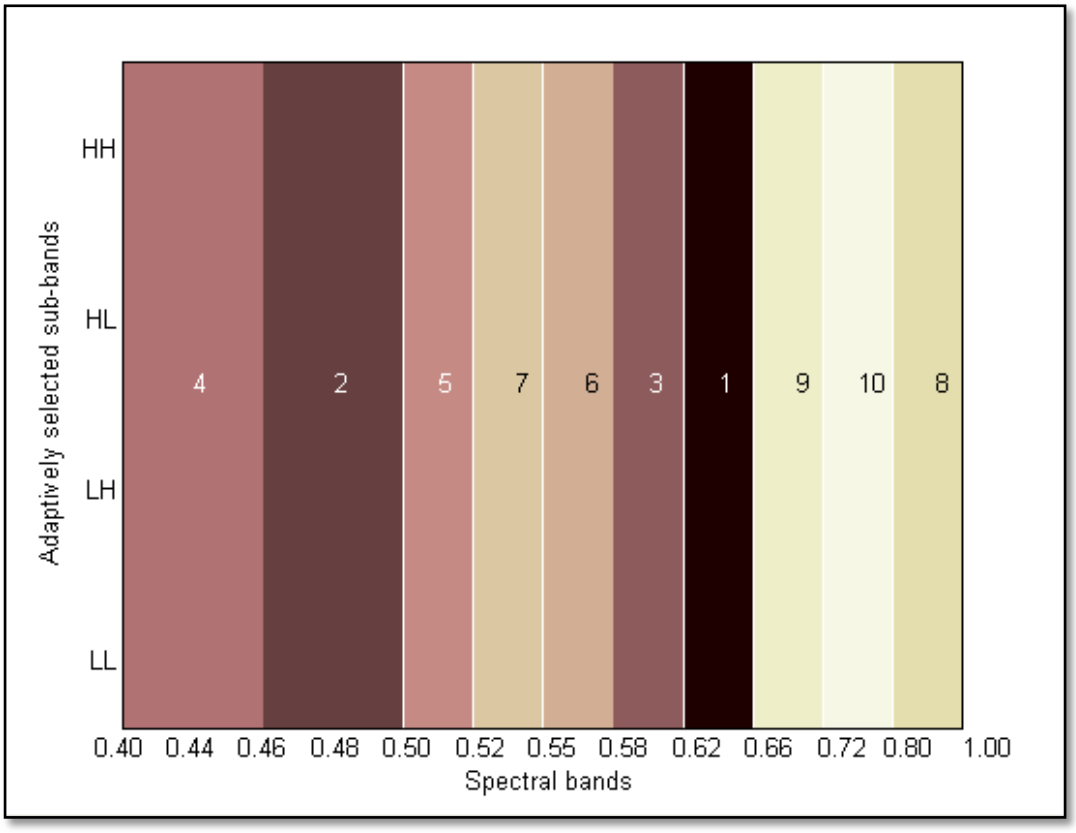

Figure 7 – 10 Spectral Spatial-Frequency Features located in the soybeans-red clover feature map ranked by FDS algorithm.

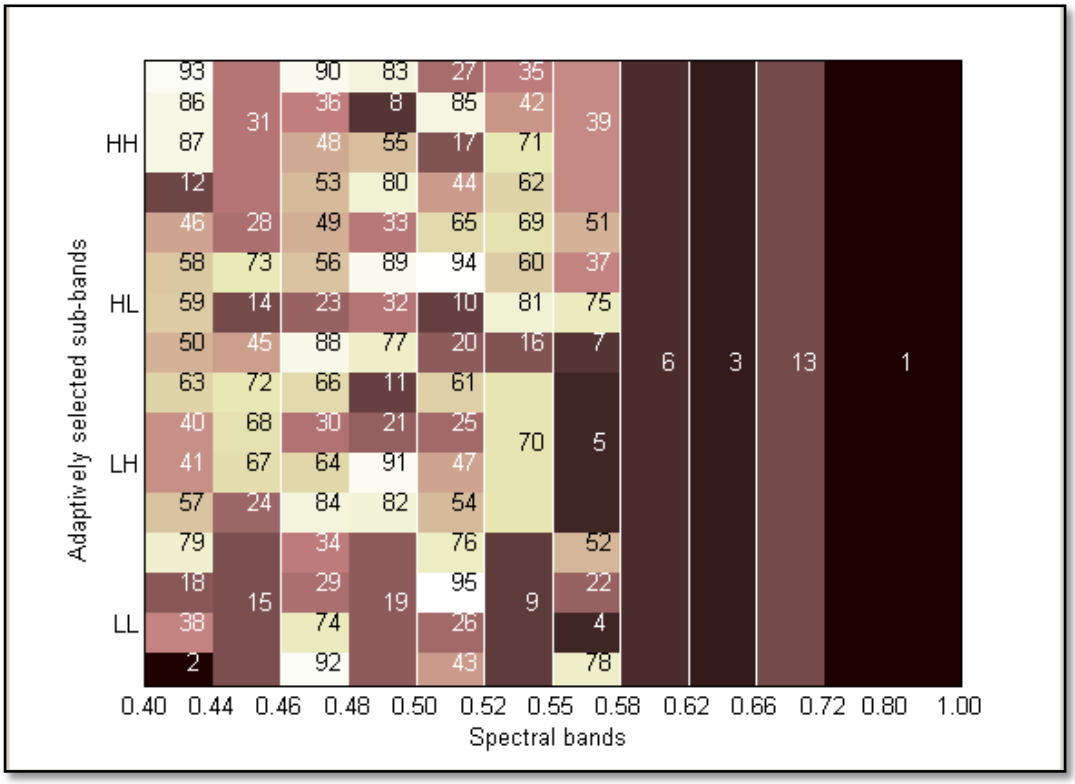

Figure 8 – 95 Spectral Spatial-Frequency Features located in the oat-wheat feature map ranked by FDS algorithm.

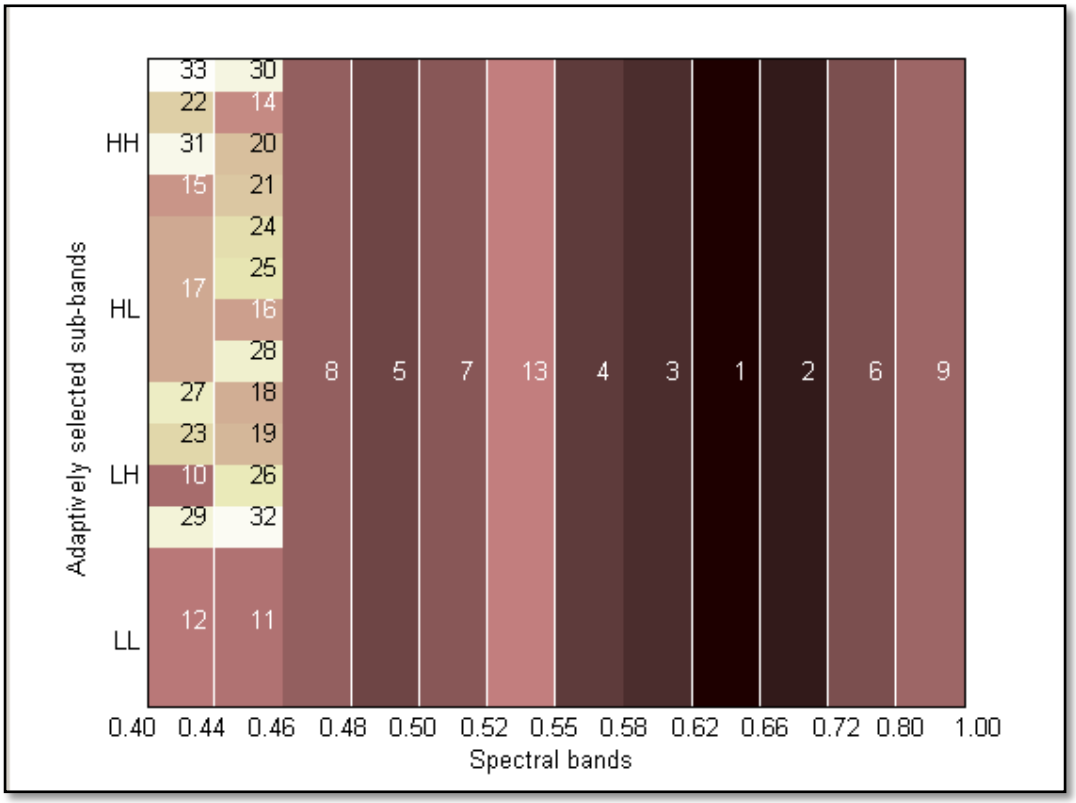

Figure 9 – 33 Spectral Spatial-Frequency Features located in the oat-red clover feature map ranked by FDS algorithm.

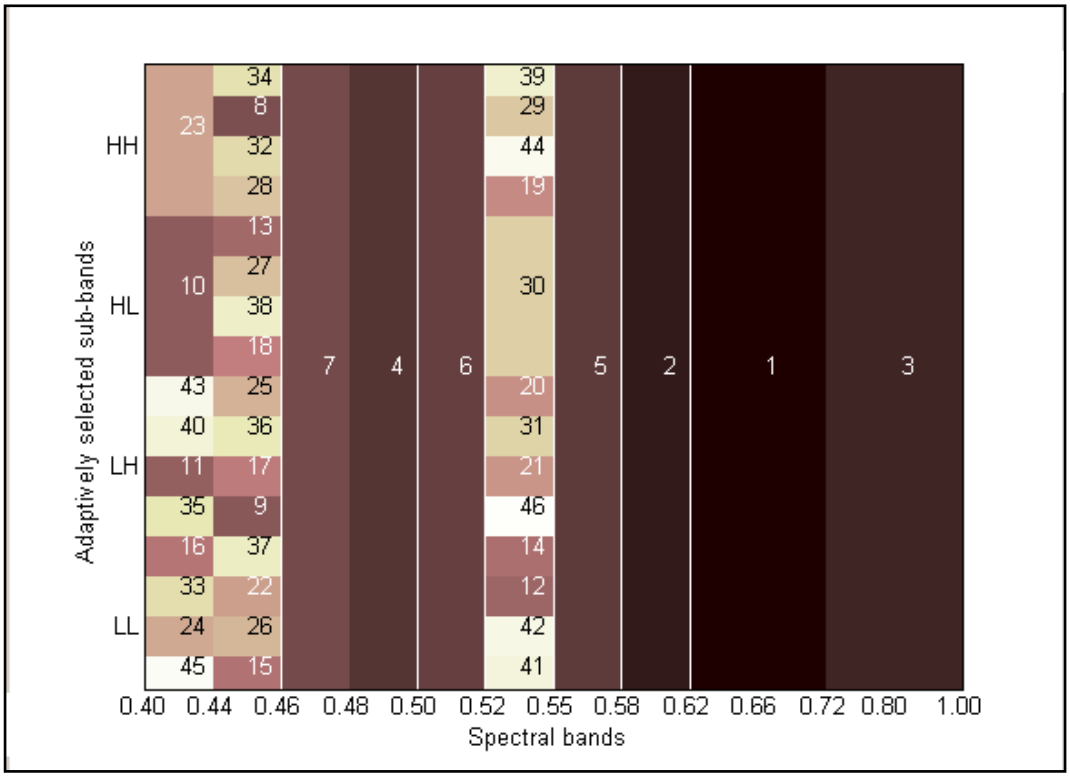

Figure 10 - 46 Spectral Spatial-Frequency Features located in the wheat-red clover feature map ranked by FDS algorithm.

All of the generated and ranked spectral spatial-frequency feature maps with feature colors and feature numbers are given below for the second type of test-training set.

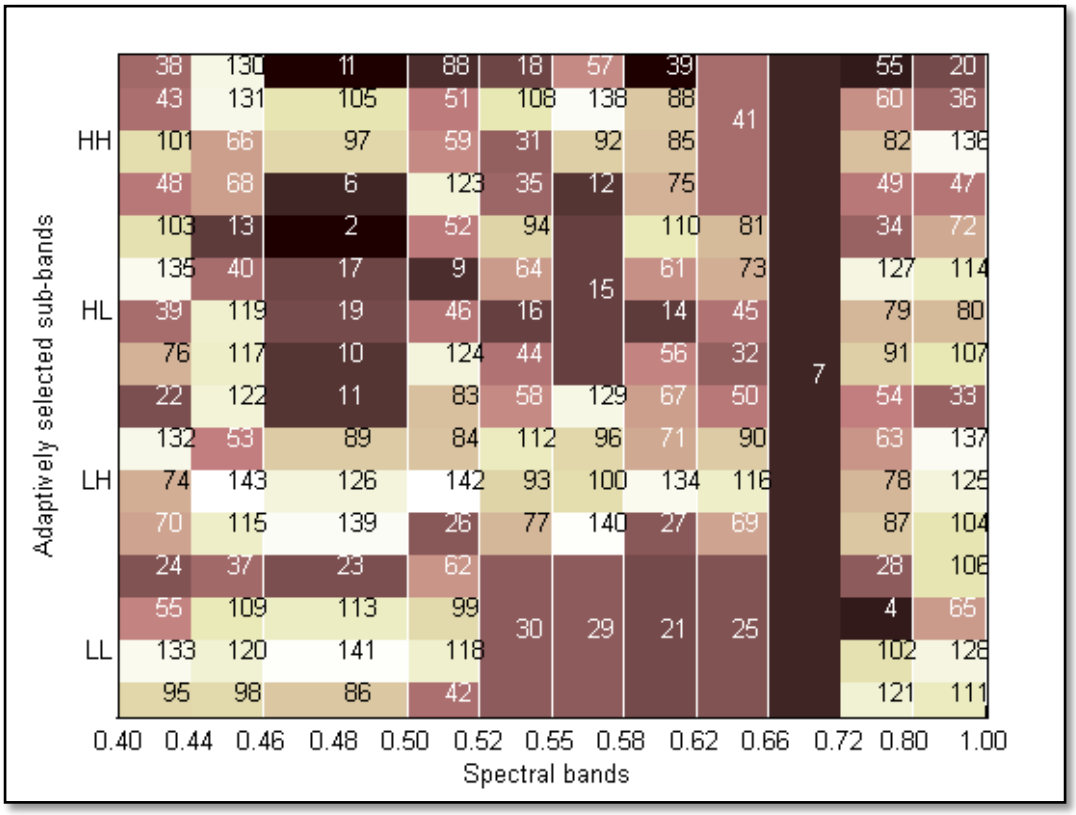

Figure 11 – 143 Spectral Spatial-Frequency Features located in the corn-soybeans feature map ranked by FDS algorithm.

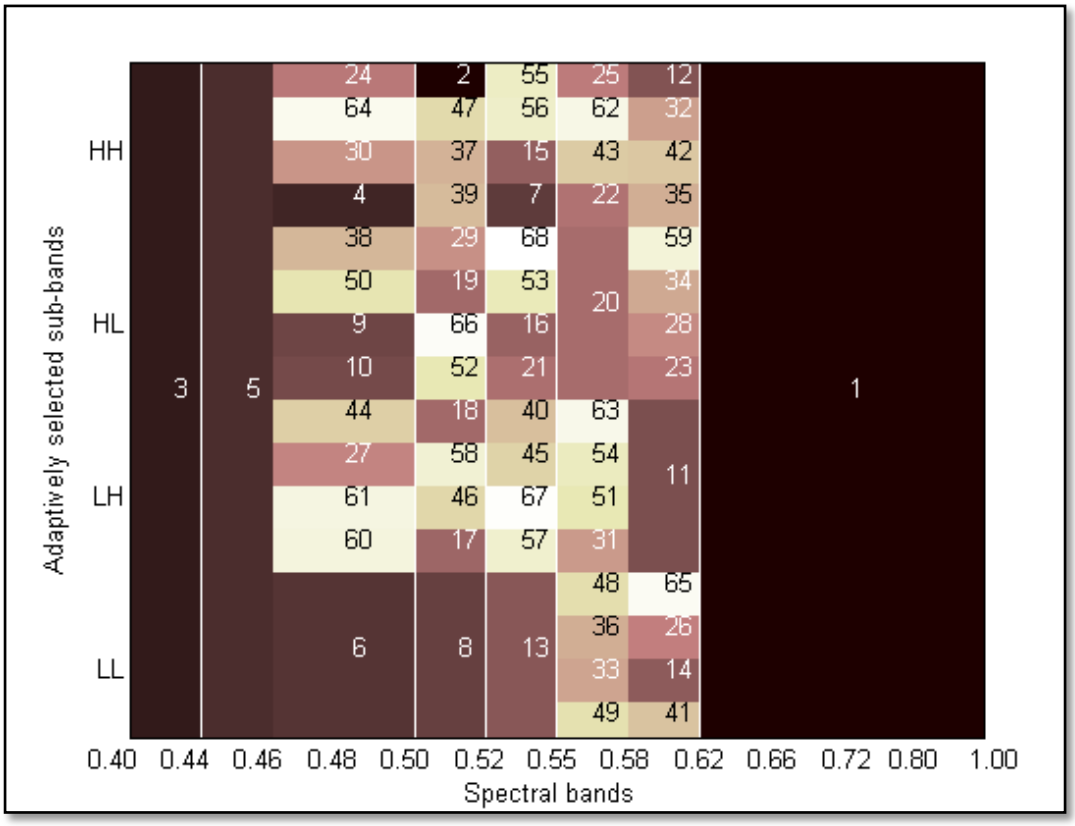

Figure 12 – 68 Spectral Spatial-Frequency Features located in the corn-oat feature map ranked by FDS algorithm.

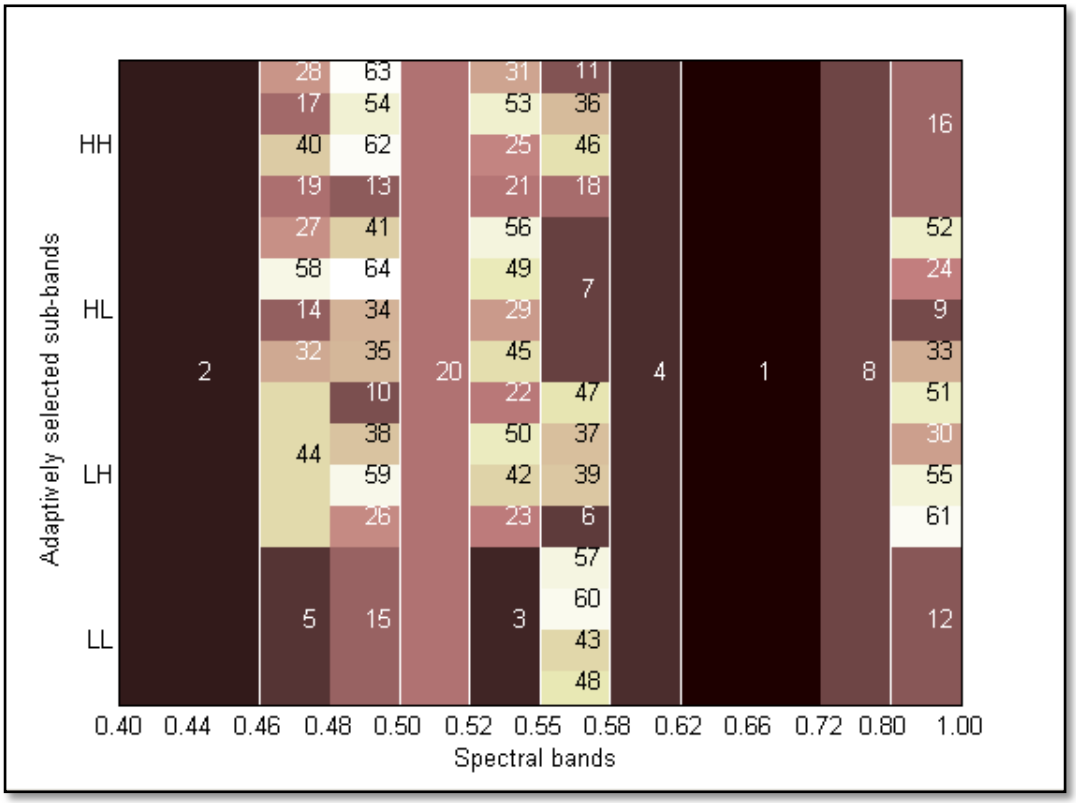

Figure 13 – 64 Spectral Spatial-Frequency Features located in the corn-wheat feature map ranked by FDS algorithm.

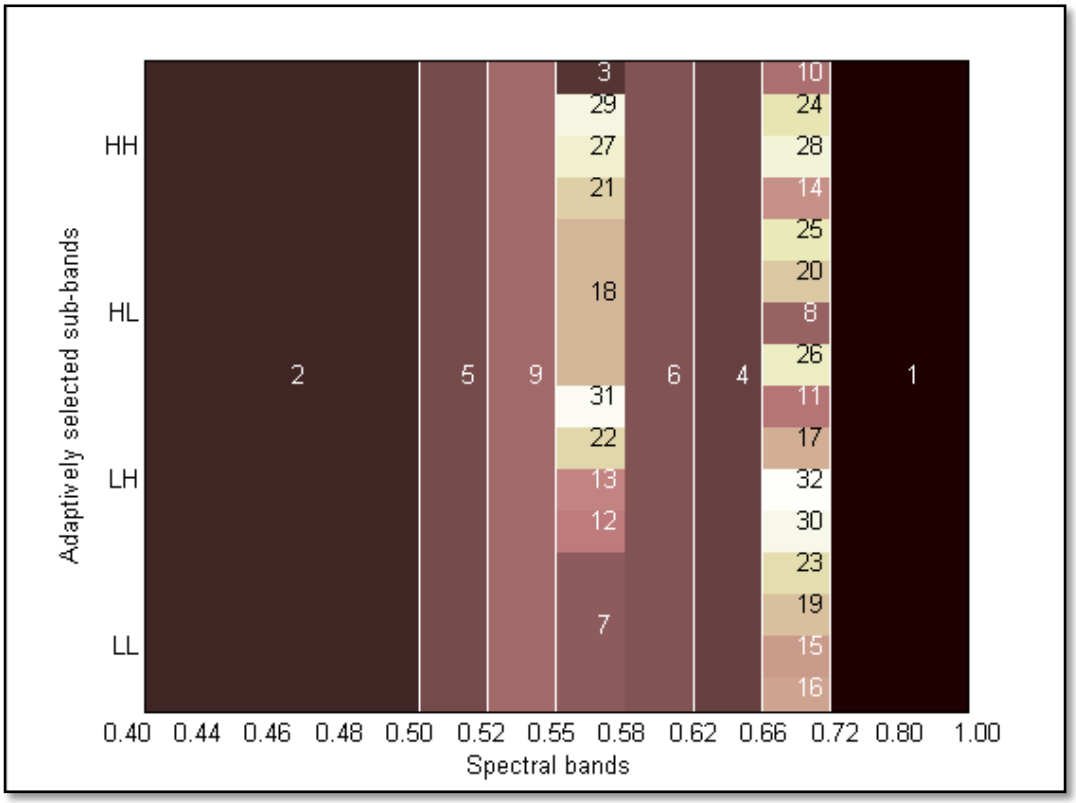

Figure 14 – 32 Spectral Spatial-Frequency Features located in the feature map of corn-red clover features ranked by FDS algorithm.

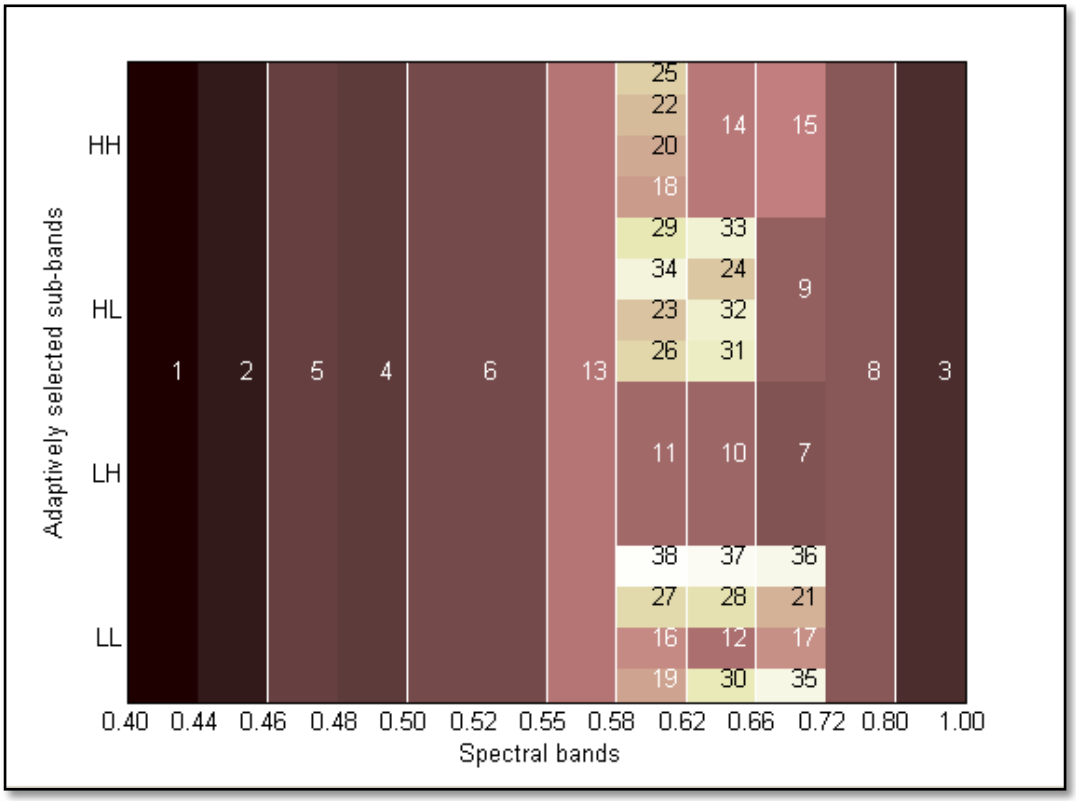

Figure 15 – 38 Spectral Spatial-Frequency Features located in the soybeans-oat feature map ranked by FDS algorithm.

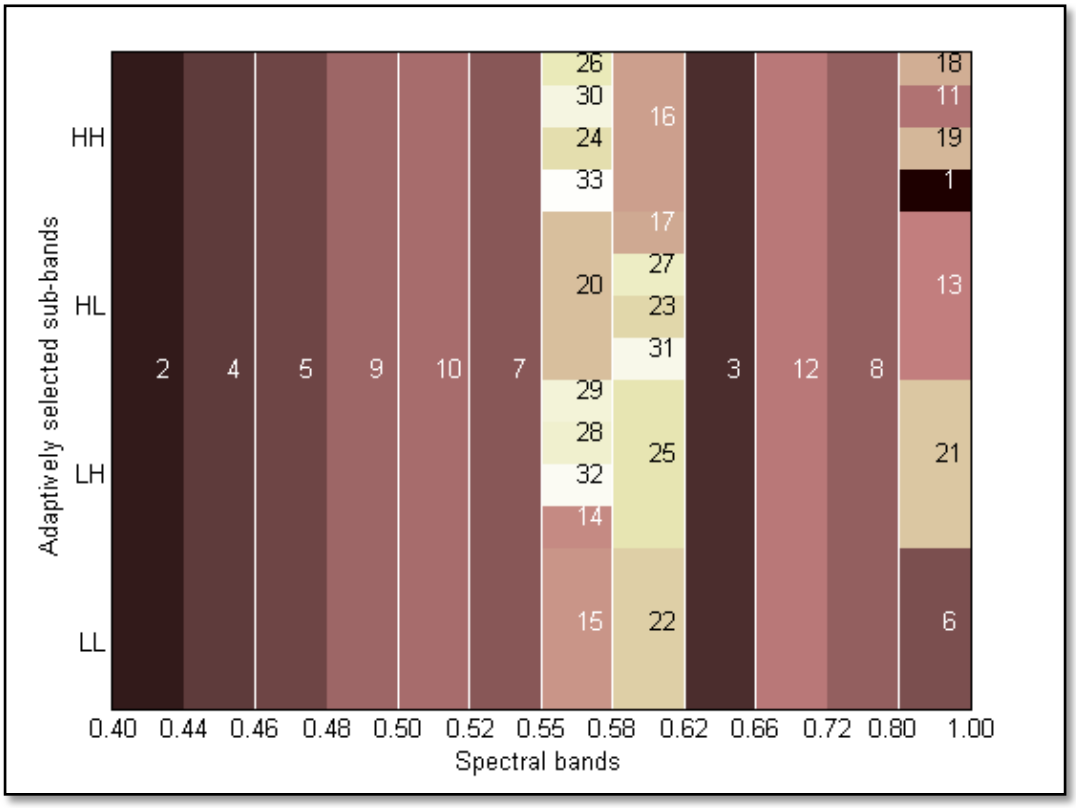

Figure 16 – 33 Spectral Spatial-Frequency Features located in the soybeans-wheat feature map ranked by FDS algorithm.

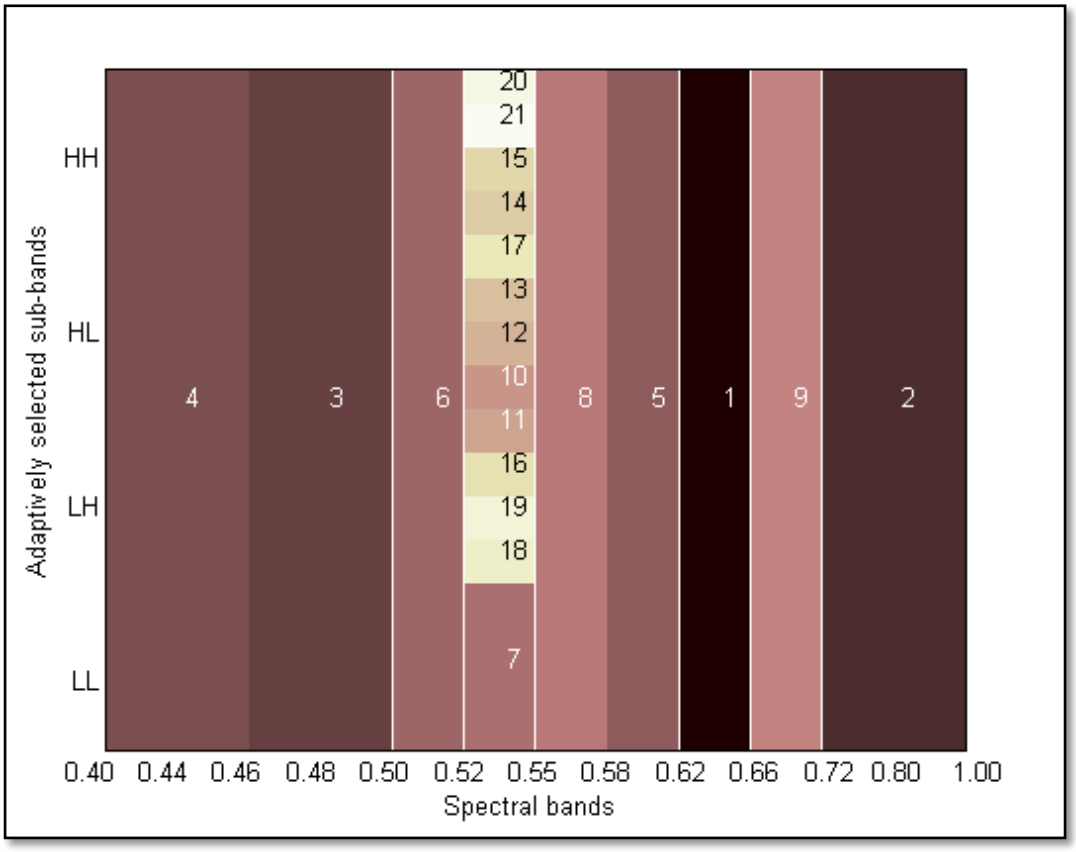

Figure 17 – 21 Spectral Spatial-Frequency Features located in the soybeans-red clover feature map ranked by FDS algorithm.

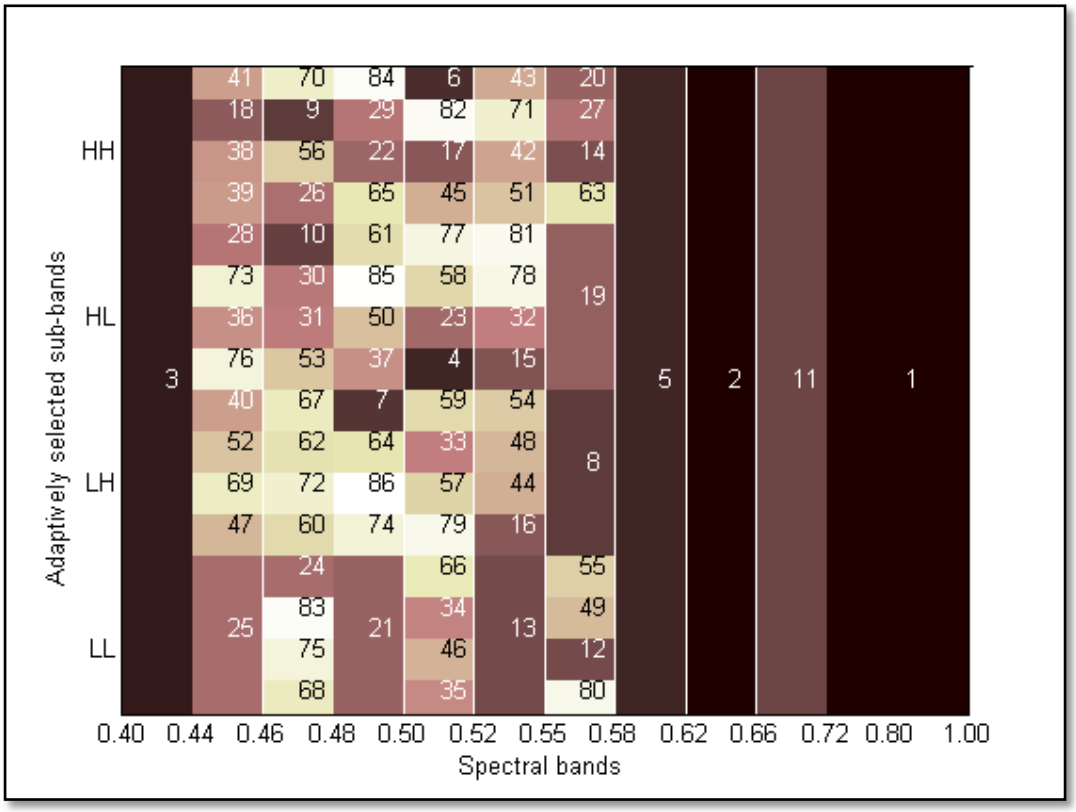

Figure 18 – 86 Spectral Spatial-Frequency Features located in the oat-wheat feature map ranked by FDS algorithm.

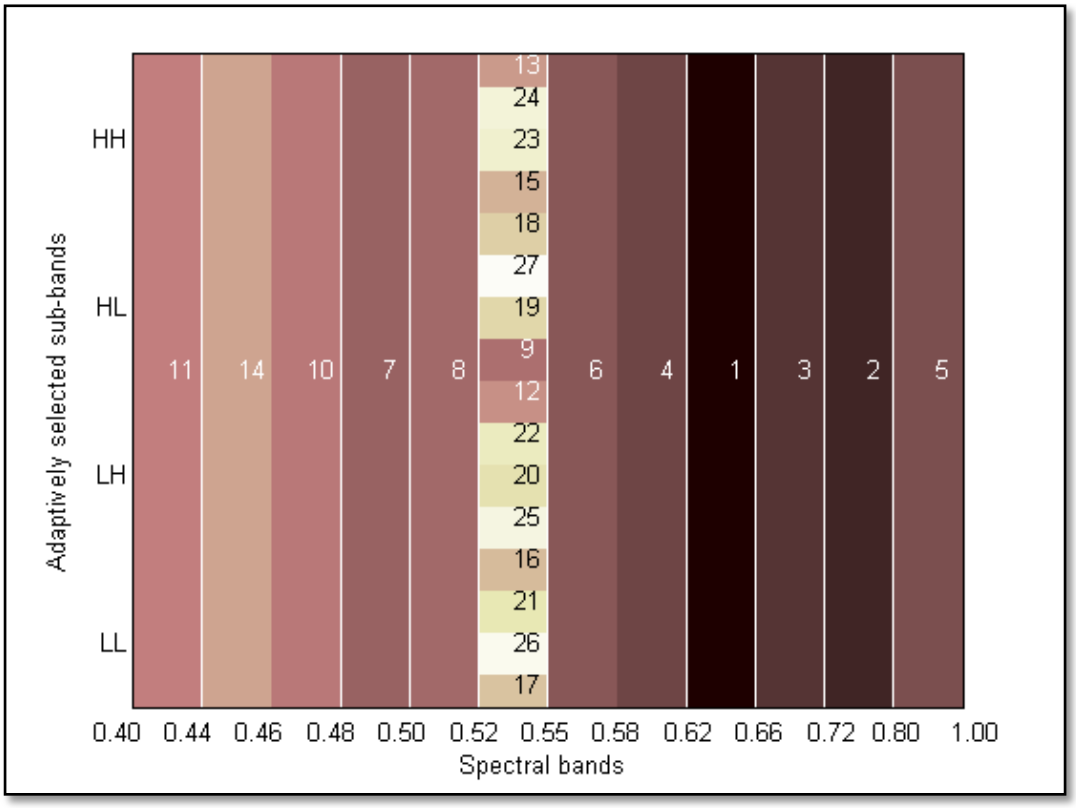

Figure 19 – 27 Spectral Spatial-Frequency Features located in the oat-red clover feature map ranked by FDS algorithm.

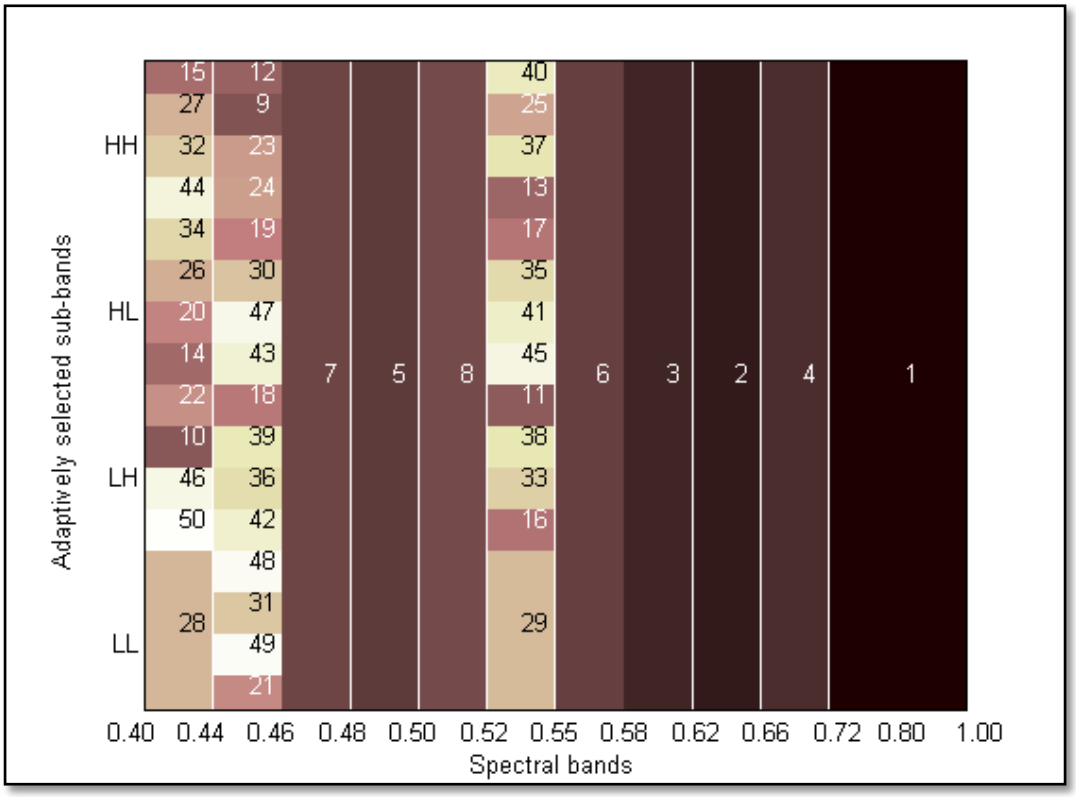

Figure 20 - 50 Spectral Spatial-Frequency Features located in the wheat-red clover feature map ranked by FDS algorithm.