Mississippi State University
Scholars Junction

Theses and Dissertations | Theses and Dissertations

1-1-2012

Genetic Algorithms Based Feature Selection and Decision Fusion for Robust Remote Sensing Image Analysis

Minshan Cui

Follow this and additional works at: https://scholarsjunction.msstate.edu/td

Recommended Citation
Cui, Minshan, "Genetic Algorithms Based Feature Selection and Decision Fusion for Robust Remote Sensing Image Analysis" (2012). *Theses and Dissertations*. 2391.
https://scholarsjunction.msstate.edu/td/2391

This Graduate Thesis - Open Access is brought to you for free and open access by the Theses and Dissertations at Scholars Junction. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholars Junction. For more information, please contact scholcomm@msstate.libanswers.com.

GENETIC ALGORITHMS BASED FEATURE SELECTION AND DECISION FUSION FOR ROBUST REMOTE SENSING IMAGE ANALYSIS

By

Minshan Cui

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

Mississippi State, Mississippi

May 2012

Copyright 2012

By

Minshan Cui

GENETIC ALGORITHMS BASED FEATURE SELECTION AND DECISION FUSION FOR ROBUST REMOTE SENSING IMAGE ANALYSIS

By

Minshan Cui

Approved:

Saurabh Prasad Assistant Research Professor GeoSystems Research Institute Adjunct Professor of Electrical and Computer Engineering (Major Advisor)	Lori M. Bruce Associate Dean Bagley College of Engineering Professor of Electrical and Computer Engineering (Co-Major Advisor)
Jenny Q. Du Associate Professor of Electrical and Computer Engineering (Committee Member)	James E. Fowler Professor of Electrical and Computer Engineering (Graduate Coordinator)

Sarah A. Rajala
Dean of the Bagley College of Engineering

Name: Minshan Cui

Date of Degree: May 11, 2012

Institution: Mississippi State University

Major Field: Electrical Engineering

Major Professor: Dr. Saurabh Prasad

Title of Study: GENETIC ALGORITHMS BASED FEATURE SELECTION AND DECISION FUSION FOR ROBUST REMOTE SENSING IMAGE ANALYSIS

Pages in Study: 43

Candidate for Degree of Master of Science

Recent developments in remote sensing technologies have made high resolution remotely sensed data such as hyperspectral and synthetic aperture radar (SAR) data readily available to detect and classify objects on the earth using pattern recognition. However, the dimensionality of such remotely sensed data is often large relative to the number of training samples available. Hence, dimensionality reduction technologies are often adopted to overcome the "curse of dimensionality" phenomenon.

This present thesis focuses on the problem of dimensionality reduction of remote sensing data by proposing two algorithms for robust classification of hyperspectral and SAR data. Specifically, for hyperspectral image analysis, a genetic algorithm based feature selection and linear discriminant analysis based dimensionality reduction method is proposed, and, for SAR data, polarization channel based feature grouping followed by a multi-classifier, decision fusion technique is proposed. The algorithmic framework of the proposed approaches and experimental results will be presented in this thesis.

DEDICATION

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

ACKNOWLEDGEMENTS

I would like to gratefully thank my major advisor, Dr. Saurabh Prasad, whose guidance, encouragement, patience and support from the initial to the final level helped me finish my master's degree at Mississippi State University. I would like thank my co-major advisor Dr. Lori M. Bruce for her guidance and support throughout my research work. Finally, I would also like to thank my committee member Dr. Jenny Q. Du for her support of my research work.

I would like to thank Dr. James V. Aanstoos, Majid and Matthew Lee for providing me the synthetic aperture radar data. I would like to thank fellow graduate students Sathishkumar Samiappan, Wei Li and Chen Chen for their precious advice and support throughout my research.

TABLE OF CONTENTS

Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Remote sensing and its applications

Remote sensing involves the acquisition of information about objects through the analysis of data collected by remote sensors that are not in physical contact with the objects of investigation. Often, these sensors are mounted on aircrafts, satellites and ships to capture images of inaccessible or dangerous areas on the earth. These sensors are either passive or active. Active sensors transmit artificially produced energy to a target and record its reflection on the target. Passive sensors, however, don't transmit energy but detect only energy emanating naturally from an object. Examples of passive sensors include traditional cameras, and radiometers, and those of active sensors include Radio Detection and Ranging (RADAR) and Light Detection and Ranging (LIDAR) sensors. Figure 1.1 shows a remotely sensed image of Mississippi State experimental fields at Brooksville, MS using an airborne hyperspectral sensor named Pro-SpecTIR-VNIR acquired by SpecTIR^{TM} for Mississippi State University. Such collected remotely sensed data has a wide range of applications in remote sensing. Figure 1.2 describes an example of a passive remote sensing setup.

1.1.1 Hyperspectral remote sensing

Hyperspectral remote sensing is a new technology and has provided a variety of applications such as geology (mineral exploration), vegetation studies (species identification) and soil science (type mapping). Recent development of remote sensing

has led the way for the development of hyperspectral sensors such as NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Hyperion Instrument and Analytical Spectral Devices (ASD) handheld spectrometer [1]. These hyperspectral remote sensors measure reflected radiation as continuous and narrow wavelength bands and hence produce images with hundreds or thousands of spectral bands which can provide unique spectral signatures for each image pixel. Figure 1.3 displays the mean spectra of cottonwood, tamarisk and willow species acquired using a HYPERION sensor. From this plot, we can easily notice that each species has a distinct spectral signature, making it unique and identifiable by that spectral signature. Such advantages make hyperspectral images extremely suitable for statistical pattern recognition.

Although hyperspectral data can provide finely resolved details about the spectral properties of features to be identified, it also has some limitations. When dealing with such high-dimensional data, one is faced with the "curse of dimensionality" problem [2]. One popular way to tackle the curse of dimensionality is to employ a feature extraction technique.

Figure 1.1 Optical remote sensing – Hyperspectral image of Mississippi State fields at Brooksville, MS, acquired using an airborne SpecTIR^{TM} inc. sensor.

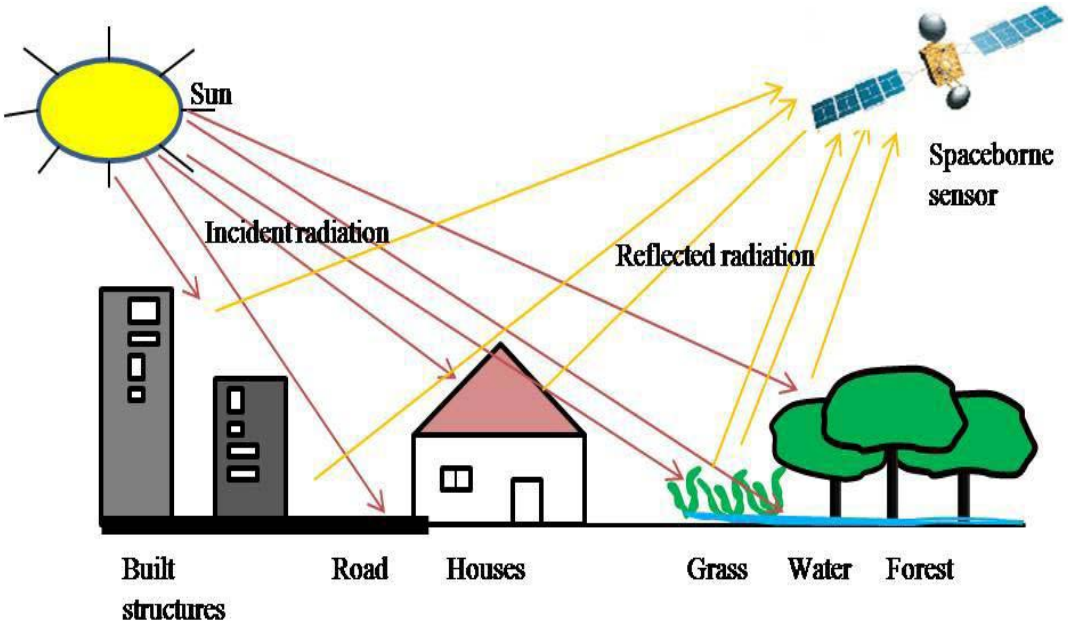

Figure 1.2 An example of a typical optical remote sensing system using a space borne sensor to measure electromagnetic radiation (EMR) reflected or emitted from the terrain.

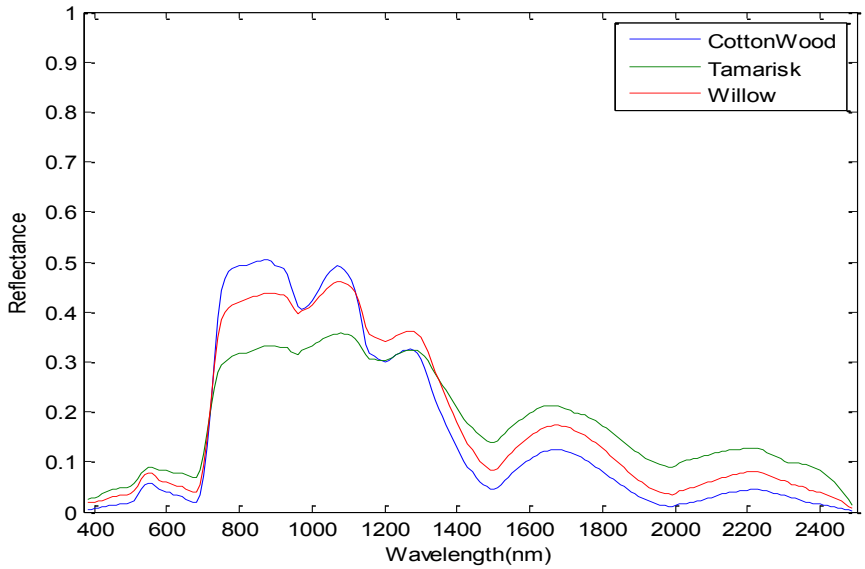

Figure 1.3 Plot of mean spectra of cottonwood, tamarisk and willow species acquired using HYPERION sensor.

1.1.2 Synthetic aperture radar remote sensing

Synthetic aperture radar (SAR) is a form of radar, mounted on satellites or aircraft that utilizes the flight path of the platform to simulate an extremely large antenna or aperture electronically, and that generates high-resolution remotely sensed imagery. Typical SAR satellite systems include RADARSAT, ENVISAT (Environmental Satellite) and TerraSAR-X and examples of airborne systems include UAV (Unpiloted Arial Vehicle) SAR. SAR systems are composed of active sensors that can transmit their own electromagnetic energy, from very long ranges and over wide areas of coverage, toward the terrain and record the backscatter of energy produced by the terrain. Unlike some passive optical sensors, SAR sensors do not depend on daylight and are insensitive to weather conditions. These advantages make SAR images highly attractive for the purpose of environmental monitoring and military surveillance and reconnaissance. SAR images can be created by sending and receiving different types of polarized energy such as VV, HH, VH and HV. VV implies that the radar antenna send out vertically polarized energy and receives only vertically polarized energy. Likewise, HH indicates that on both transmit and receive, the waves are horizontally polarized. If we say that the polarization is VH, then that means the transmitted waves are polarized in the horizontal direction, while the received waves are vertically polarized. Finally, HV means the transmitted radiation is horizontally polarized, and the polarization of the received radiation is restricted to just vertically polarized waves. With the dramatically improved image resolution and wide availability of such imagery, SAR images have been widely used in pattern recognition.

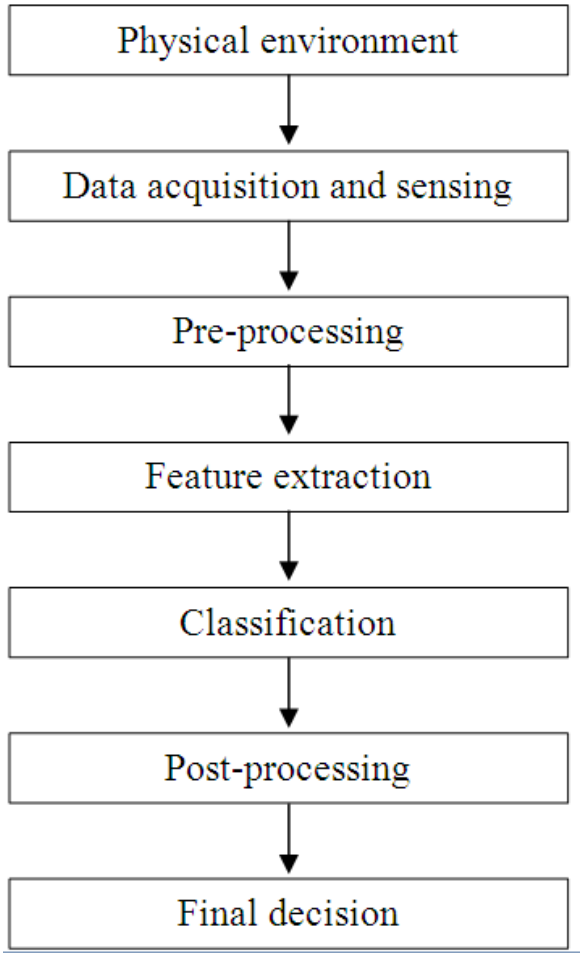

Figure 1.4 Block diagram of pattern classification system.

1.2 Pattern recognition methodologies for remote sensing

The task of pattern recognition system is to assign some input data to a correct category based on the measurements of the input data. A typical system of pattern recognition for remotely sensed data can be described by the block diagram shown in Figure 1.4. Typically, such a system employs a sensor, a pre-processing mechanism, a feature extraction mechanism, a classification scheme and a post-processing mechanism. Sensing refers to measurements of physical variables about the object to be classified using one or more of the above-mentioned sensors. For example, if the input data is an image, then the sensing equipment can be a camera. After data acquisition, the data will

be fed into the pre-processing stage which includes removal of noise in the data, isolation of patterns of interest from the background and some other operations performed on the raw data to make it more suitable for pattern classification. In the feature extraction stage, we extract features that best represent the data for classification. This step is critical to the success of pattern recognition systems, especially when dealing with high-dimensional data such as hyperspectral data. At the classification stage, we use extracted features and learned models to assign a pattern to a category. There are two types of classification techniques – supervised and unsupervised. In supervised classification, we have some training samples that are correctly labeled to learn or estimate the unknown parameters of the model. In unsupervised classification, we are only provided with unlabeled data which is to be grouped into different clusters. In this study, we focus on supervised pattern recognition techniques. Finally, we exploit context (e.g. spatial information) to improve system performance in post-processing stage.

1.3 Thesis organization and contributions

In this thesis, two different pattern recognition techniques are proposed to analyze remotely sensed data. A brief overview of the proposed algorithms is presented below. The overall objective of the proposed algorithms is to provide robust ground-cover classification of high dimensional remotely sensed data, particularly when the number of training samples is very small.

1.3.1 Decision Fusion of Textural Features Derived from Polarimetric Data for Levee Assessment

Texture features derived from Synthetic Aperture Radar (SAR) imagery using grey level co-occurrence matrix (GLCM) can result in very high dimensional feature

spaces. Although this high dimensional texture feature space can potentially provide relevant class-specific information for classification, it often also results in over-dimensionality and ill-conditioned statistical formulations. In this thesis, we propose a polarization channel based feature grouping followed by a multi-classifier decision fusion framework (MCDF) for a levee health monitoring system that seeks to detect landslides in earthen levees. In this system, texture features derived from the SAR imagery are partitioned into three groups according to different polarization channels (HH, HV and VV) contained in SAR data. A multi-classifier system is then applied to each group to perform a local classification. Finally, a decision fusion system is employed to fuse decisions generated by each classifier to make a final decision which is levee or landslide. The resulting system can handle the high dimensionality of the problem very effectively, and only needs a few training samples for training and optimization. This proposed method will be explained in detail in Chapter 2.

1.3.2 Genetic Algorithms and Linear Discriminant Analysis based Dimensionality Reduction for Hyperspectral Data

Remotely sensed data (such as hyperspectral imagery) is typically associated with a large number of features, which makes classification challenging. Feature subset selection is an effective approach to alleviate the curse of dimensionality when the number of features contained in datasets is huge. Considering the merits of genetic algorithms (GA) in solving combinatorial problems, GA is becoming an increasingly popular tool for feature subset selection. Most algorithms presented in the literature using GA for feature subset selection use the training classification accuracy of a specific algorithm as the fitness function to optimize over the space of possible feature subsets. Such algorithms require a large amount of time to search for an optimal feature subset. In

this thesis, we will present a new approach called Genetic Algorithm based Linear Discriminant Analysis (GA-LDA) to extract features in which feature selection and feature extraction are performed simultaneously to alleviate over-dimensionality and result in a useful and robust feature space. Experimental results with classification tasks involving hyperspectral imagery indicate that GA-LDA can result in very low-dimensional feature subspaces yielding high classification accuracies. A description of this proposed method with corresponding experimental results will be given in Chapter 3.

CHAPTER II

DECISION FUSION OF TEXTURAL FEATURES DERIVED FROM POLARIMETRIC DATA FOR LEVEE ASSESSMENT

2.1 Introduction

Grey level co-occurrence matrix (GLCM) based texture measurements have been a popular method for texture extraction in remotely sensed images [3], [4]. The texture features extracted from GLCM can potentially be very high dimensional. Oftentimes, the dimensionality may be much higher than the number of available training data samples. This adversely affects the classifiers ability to model the properties of high dimensional data and therefore, results in poor classification. To efficiently address this problem, feature selection or feature extraction methods are usually employed to reduce the dimensionality of the feature space. Typically, Stepwise Linear Discriminant Analysis (SLDA) [5] is performed to select a subset of features. The key idea behind SLDA is that a preliminary forward selection and backward rejection is employed to discard redundant and less relevant features, and then a Linear Discriminant Analysis (LDA) projection is applied on this reduced subset of features to further reduce the dimensionality of the feature space. One disadvantage of SLDA is that in forward selection (FS) [6], one is unable to reevaluate the features that become irrelevant after adding other features. Similarly, in backward rejection (BR), one is unable to reevaluate the features after they have been discarded. This drawback results in the algorithm discarding potentially useful information that is important to the classification task.

In this work, we propose a decision fusion approach for robust classification of GLCM features derived from SAR images. The system is an extension of the Multi-Classifier and Decision Fusion (MCDF) system developed previously [7]. The proposed approach first uses GLCM to extract texture features from SAR data and then groups the entire feature space into three subspaces. This grouping partitions the high dimensional feature space into many small groups. A multi-classifier system is then employed to perform local classification for each group, and an appropriate decision fusion rule is applied to merge these local class labels generated by each classifier to perform a final levee or landslide classification for the SAR image. Unlike SLDA, this proposed system not only efficiently uses all the available information provided by the dataset but also avoids the over-dimensionality problem by employing MCDF.

2.2 Traditional methods

Dimensionality reduction is an important preprocessing step before classification for high dimensional data. It attempts to overcome the over-dimensionality problem induced by high dimensional feature spaces. In this section, we review two popular dimensionality reduction approaches - LDA and SLDA and a commonly used classifier - the maximum likelihood classifier.

2.2.1 Dimensionality Reduction

LDA is a commonly used technique for dimensionality reduction. The objective of LDA is to perform dimensionality reduction while preserving as much of the class discrimination information as possible. LDA seeks to find a linear transformation w to a reduced dimensional subspace such that the ratio of within-class scatter to between-class scatter, $J(w)$ in this projected subspace provided by Fisher's ratio [8], [9] is maximized:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{2.1}$$

where S_B and S_W are the between-class and within-class scatter matrices.

One limitation of LDA based dimensionality reduction is that S_w is required to be non-singular - if this matrix is singular, LDA fails. To prevent this, a large number of training samples must be provided to make scatter matrices non-singular.

Traditionally, for small-sample-size and high dimensional classification problems, a technique known as SLDA is commonly employed for dimensionality reduction. The procedure of SLDA is described as follows. Before feature selection, we rank features by a certain metric such as Bhattacharyya Distance (BD) [10] or the area under the Receiver Operator Characteristic (ROC) curve. In FS, we start from an empty set and sequentially add features from the highest rank. If an added feature improves some objective function (e.g. BD or area under the ROC curves), we retain the added feature, and otherwise we discard it. This process continues until all the features are evaluated or satisfy a certain criteria. The main drawback of FS is its inability to remove features that become obsolete after adding some other features. In BR, we start from full feature set and sequentially remove features one by one. If removal of some feature led to deterioration in the objective function, we remove that feature from the feature set, otherwise we retain it. This process continues until all the features are examined or satisfy one of stopping criteria. The main disadvantage of BR is that it is unable to reevaluate the usefulness of a feature after it has been removed.

2.2.2 Maximum Likelihood Classifier

The Maximum Likelihood Classifier is a popular classification method for remote sensing tasks. It relies on the class conditional probability density functions to calculate

the likelihood that a given pixel, with the unique mean vector and covariance matrix of each class estimated from the training data, belongs to one of these reference classes. Every pixel is assigned to the class that has the highest probability. The equation for the discriminant function for each class is

$$g_i(x) = p(x|w_i)P(w_i) \tag{2.2}$$

$$P(x|w_i) = \frac{1}{2\pi^{n/2}|\sum_i|^{1/2}} \cdot e^{-(1/2)(x-\mu_i)^T\sum_i^{-1}(x-\mu_i)} \tag{2.3}$$

where n is the number of features, x is data vector, μ_i is the mean vector of class i and $\sum_i$ is the covariance matrix of class i. Given its simplicity in representation and good convergence properties, the maximum likelihood classifier is commonly used in image analysis.

2.3 Proposed system

2.3.1 Grey-level co-occurrence matrix

Before describing the proposed polarization channel based feature grouping and MCDF system, we briefly introduce GLCM feature extraction. GLCM texture measurements have been a popular method for texture extraction in remotely sensed images since they were first introduced by Haralick in the 1970s [11]. GLCM describes texture of images by statistically sampling certain grey-levels in relation to spatially adjacent pixels. For a position operator p, we can define a matrix C_{ij} that counts the number of times a pixel with grey-level i occurs at position p from a pixel with grey-level j. Figure 2.1 shows an image that has three different grey-levels 0, 1, and 2 and its corresponding counts matrix C defined by the position operator p as "lower right". If we normalize the counts matrix C by the total number of pixels and make it symmetrical by

having the same values occur in cells on opposite sides of the diagonal, we get a grey-level co-occurrence matrix, *P*. For each GLCM, among the set of 14 texture features defined in [11], we select the six commonly used features [12] in our analysis which is tabulated in Table 2.1.

0	0	0	1	2
1	1	0	1	1
2	2	1	0	0
1	1	0	2	0
0	0	1	0	1

4	2	1
2	3	2
0	2	0

Figure 2.1 Top: describing an image; Bottom: corresponding counts matrix C.

Table 2.1 Six common features used in GLCM.

Contrast	$\sum_i \sum_j P_{ij} \cdot (i-j)^2$
Entropy	$-\sum_i \sum_j P_{ij} \log P^{ij}$
Correlation	$\sum_i \sum_j \frac{(i-\mu)(j-\mu)P_{ij}}{\sigma^2}$
Energy	$\sum_i \sum_j P_{ij}^2$
Homogeneity	$\sum_i \sum_j \frac{P_{ij}}{1+(i-j)^2}$
Variance	$\sum_i (i-\mu^2) \sum_j P_{ij}$

*$P(i,j)$ represents the value of the element of GLCM at the coordinate(i,j),$\mu = \sum_i \sum_j i \cdot P(i,j)$ and σ is the variance.

2.3.2 Polarization channel based feature grouping and Multi-classifier decision fusion

Figure 2.2 describes the overall block diagram of the proposed system. First, we set up as many groups as are the number of polarization channels. Next, we use GLCM to extract texture features from each polarization channel. For example, if the SAR data has *n* polarization channels, then we have *n* groups, and each group is composed of the same texture features derived from GLCM. After dividing texture features into small groups, LDA based feature extraction would be beneficial for the 'local' classification. This is expected because LDA will further project texture features into subspaces where all classes are well separated from each other. Since each group contains a small number of

features compared to the original feature space, the scatter matrices tend to be non-singular even though they might be singular in the original feature space. After LDA based preprocessing, each group will be fed into a certain type of classifier. The multi-classifier system is essentially a bank of classifiers that is applied to each group to do local classification. Class labels and posterior probabilities derived from each classifier will be merged into a single healthy levee or landslide class label per SAR image pixel according to some decision fusion scheme. This decision fusion scheme can be hard fusion which directly uses class labels from each group to make the final classification decision (e.g. conventional majority voting). It can also be a soft decision fusion scheme which makes use of posterior probabilities from each group to make the final classification decision. In our experiment, we use both schemes to test our proposed system [13].

In hard decision fusion, we make a final classification decision by voting over the class labels produced from each group. Since this decision fusion only relies on individual class label, it is somewhat insensitive to inaccurate posterior probability estimates. Linear Opinion Pool (LOP) is a popular soft decision fusion scheme that uses posterior probabilities produced by each classifier $P_j(w_i|x)$ to estimate class labels. The LOP decision fusion is given by

$$C(w_i|x) = \sum_{j=1}^{n} a_j P_j(w_i|x) \tag{2.4}$$

$$w = \text{argmax}_{i\epsilon\{1,2,\ldots C\}}\, C(w_i|x) \tag{2.5}$$

where a_j is the confidence score / weight (e.g., training accuracies) for the *j*-th classifier, *w* is the class label from one of the *C* possible classes for the test pixel, *j* is the classifier index, *n* is the number of subspaces / classifiers.

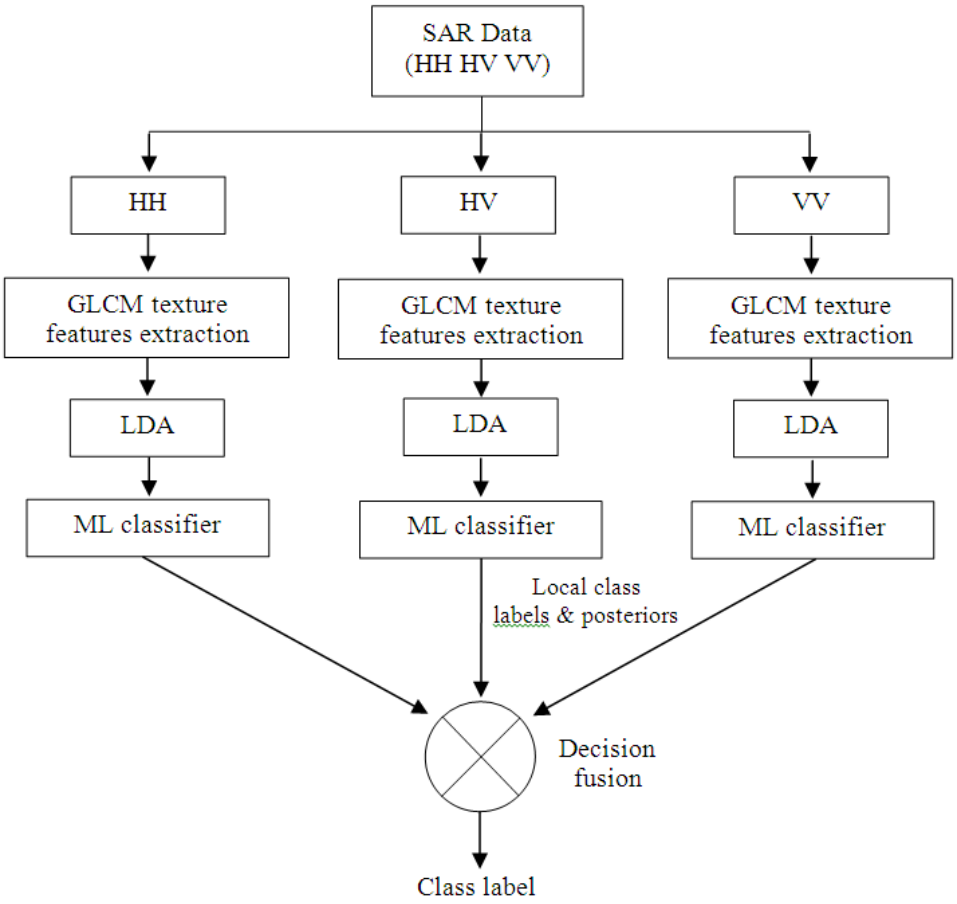

Figure 2.2 Schematic of the proposed system for robust classification of SAR data.

2.4 Experimental SAR data

In this paper our methodology is tested in an application in which SAR is used to detect potential failure zones in earthen levees used for flood control [14]. In particular, this study focuses on slump (or slough) slides. These are slope failures along a levee, which leave areas of the levee vulnerable to seepage failure during high water events. A picture of a typical levee slump slide is shown in Figure 2.3. The roughness and related textural characteristics of the soil in a slide affect the amount and pattern of radar backscatter. Secondary physical characteristics including soil moisture and types and density of vegetation that grows in a slide area differ from the surrounding "healthy" levee areas, and can also be utilized in detecting slides [15]. Early detection of the

occurrence of slump slides can assist levee managers in prioritizing their inspection and repair efforts. A remote sensing based solution for their rapid detection would be more efficient and cost effective than frequent site visits. Our approach relies on the premise that characteristics of the soil and vegetation that can be used to discriminate slump slides from healthy levee are manifested in the backscatter of polarimetric radar due to its response to spatially variant soil moisture and roughness. L band radar is known to penetrate dry soils up to one meter in depth, and has been used to map surface soil moisture [16].

Figure 2.3 Typical levee slump slide

The data used in this study is from the NASA Jet Propulsion Laboratory's UAVSAR (Uninhabited Aerial Vehicle Synthetic Aperture Radar) instrument, a fully polarimetric L-band synthetic aperture radar flown on a Gulfstream-3 research aircraft [17]. The UAVSAR is flown at an altitude of 12.5 km and takes an image swath 20 km wide. For this study we are focusing on a 500 square meter portion of a Mississippi River levee near Vicksburg, MS. This section of levee contains a slump slide which had not yet been repaired at the time of the radar flight. Although the raw ground sample distance is

1.6 by 0.6 meters, our efforts use the multi-look 5 by 7 meter data to minimize speckle effects. The radar backscattering coefficient magnitudes in each of the 3 polarization channels HH, VV, and HV are the input data for this study.

2.5 Experimental settings and results

In this work, three channels (HH, HV and VV) of the SAR data are re-quantized into 512 grey levels to extract texture features using GLCM. The GLCM is computed over a 7 pixels by 7 pixels window of the SAR image, and a pixel distance of 1 and an angle (direction operator) of 0 is used. Figure 2.4 depicts six common features that are used to extract texture features from SAR imagery.

All experimental results reported in this paper are performed using a repeated random sub-sampling validation technique to evaluate the performance of the proposed system. Specifically, random selections of training and test samples are made from different regions of the actual levee segment respectively and classification accuracies are estimated using the proposed system, in addition to two single classifier systems - LDA or SLDA based dimensionality reduction followed by a single ML classifier. Each experiment with random selection of training and test data is repeated 50 times and mean accuracies is reported. We employ SLDA and LDA as baseline experiments to compare against the proposed system. Gaussian maximum likelihood classifier is employed to do local classification in the MCDF system, and both MV and LOP based decision fusion schemes are tested in this work. Note that the random sub-sampling is carried out after GLCM feature extraction.

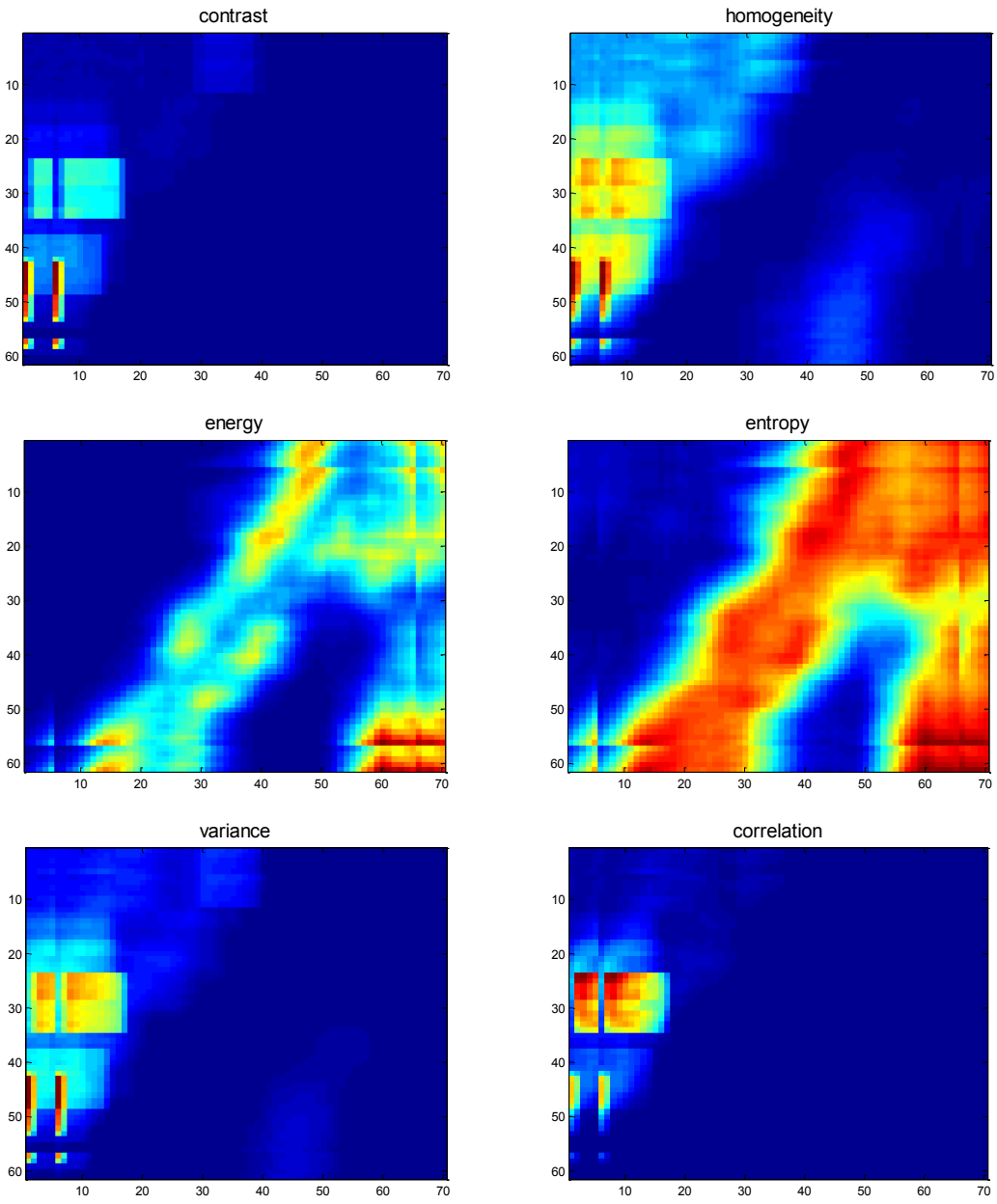

Figure 2.4 Plot of six common texture features used in GLCM.

Figure 2.5 describes the SAR image and Mask of actual levee segment used to test our proposed system. In this image, we only retain the healthy levee and landslide classes while masking out all other classes.

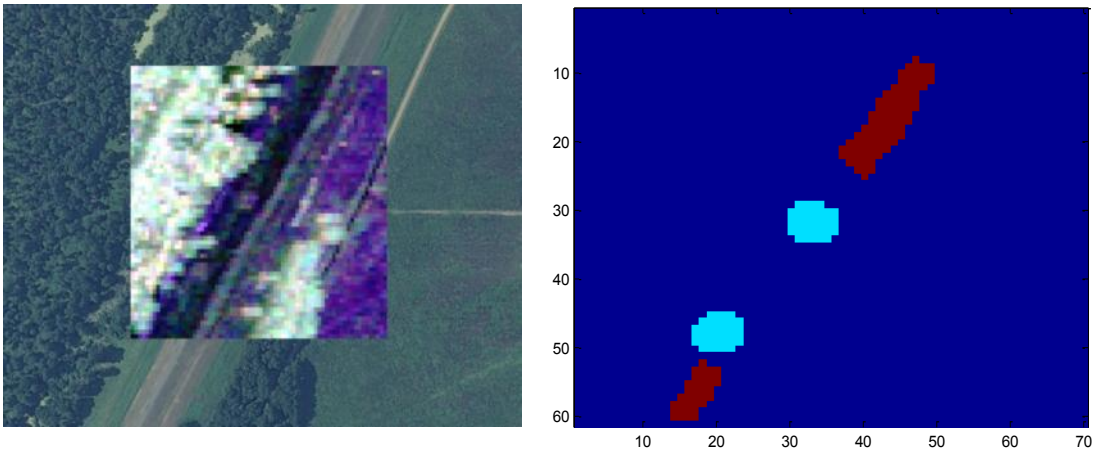

Figure 2.5 Left: Actual levee segment SAR image inset with the optical image; Right: The mask used in this image; Green: landslide; Red: healthy levee.

Figure 2.6 shows the overall classification accuracies obtained as a function of varying number of training samples. From this figure, we can infer that the proposed polarization channel based feature grouping followed by the MCDF framework approach for the SAR data classification problem outperforms traditional SLDA and LDA based classification. As the number of training samples increases, the accuracies of our proposed algorithm improve significantly. However, LDA based projection fails to effectively extract textural features and thus results in very low classification accuracy, and the accuracy improvement of SLDA as the number of training samples increases is much smaller compared to our proposed algorithm. Based on the plots of standard deviations described in Figure 2.7, our proposed algorithms show less diversity than the SLDA and LDA, especially when the number of training sample is small.

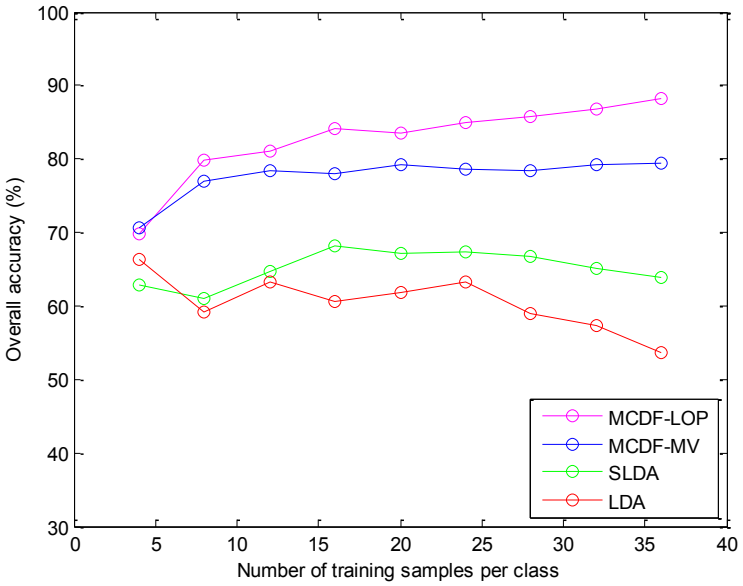

Figure 2.6 Illustrating the benefits of proposed system compared to traditional methods with regards to overall accuracy versus number of training samples. MCDF-LOP: Using LOP as the decision fusion scheme in MCDF; MCDF-MV: Using MV as the decision fusion scheme in MCDF.

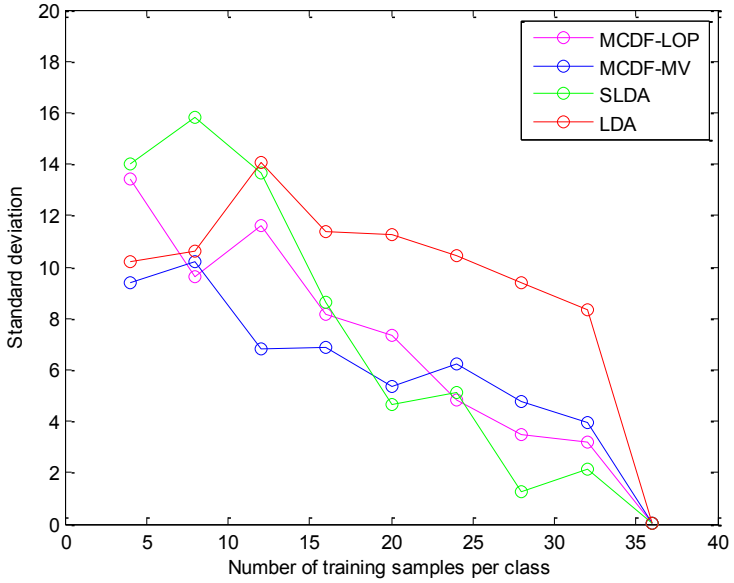

Figure 2.7 Plots of standard deviation versus number of training samples of four different algorithms in 2-class problem.

In an attempt to demonstrate the feasibility of our proposed algorithm in a more complicated scene, we employed a multi-class SAR imagery to test our proposed system. Figure 2.8 depicts the SAR image and the mask of the actual levee segment used to test our proposed system. Five different classes are presented in this image, namely landslide, healthy levee, trees, agriculture fields and roads.

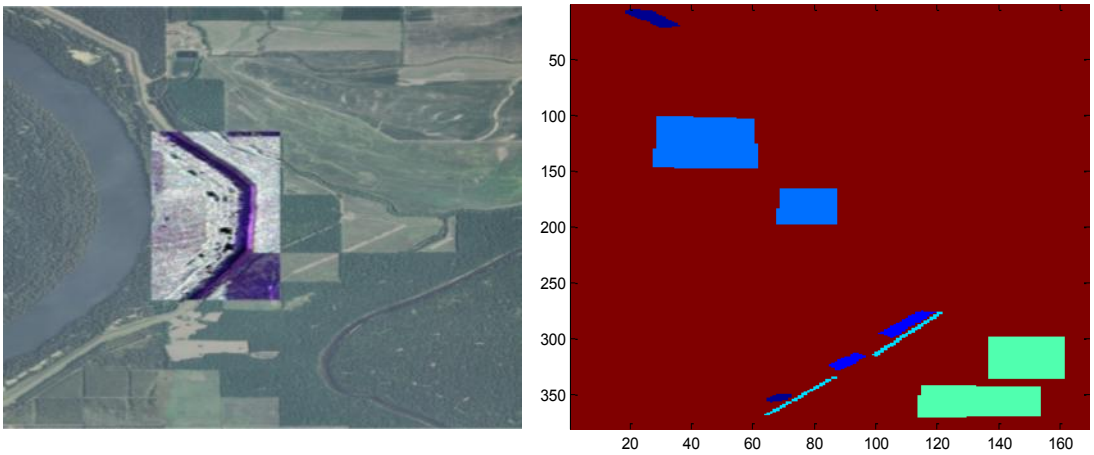

Figure 2.8 Left: Actual levee segment SAR image; Right: Mask used in this image. Dark blue: landslide; Blue: healthy levee; Light blue: tree; Light green: agriculture field; Cyan: road.

The overall classification accuracies of landslide and healthy levee obtained as a function of varying number of training samples is plotted in Figure 2.9. From this figure, we can infer that our proposed algorithms still outperforms traditional SLDA and LDA in a multi-class problem, and one can obtain an overall accuracy of around 70% even when the number of available training sample is only 10. Figure 2.10 shows the standard deviation of the overall accuracy with the four different algorithms. From this figure, it can be seen that the proposed system has a much lower standard deviation than SLDA and a little lower than the LDA. Based on these observations, we can infer that our

proposed methods deliver significantly better performance in terms of both classification accuracy and standard deviation in a complicated dataset.

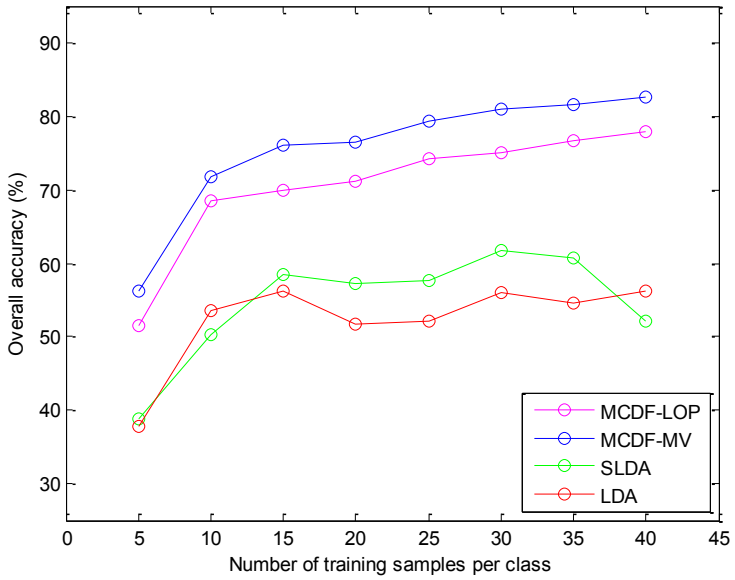

Figure 2.9 llustrating the benefits of the proposed system compared to traditional methods with regards to overall accuracy versus number of training samples in a multi-class problem.

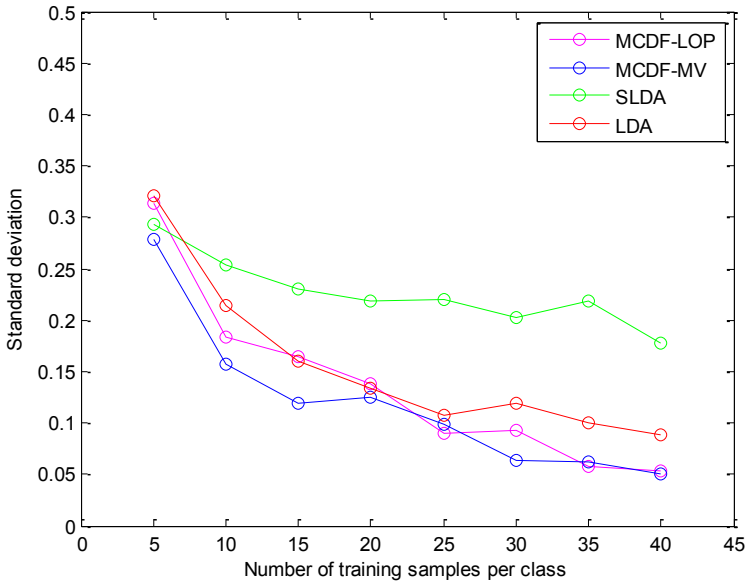

Figure 2.10 Plots of standard deviation versus number of training samples of four different algorithms in multi-class problem.

2.6 Conclusion

Experimental results presented in this chapter indicate that polarization channel based feature grouping followed by a multi-classifier decision fusion framework performs markedly better than traditional SLDA and LDA based dimensionality reduction, followed by a single-classifier. Figure 2.11 illustrates the classification maps generated using the proposed MCDF framework as well as conventional dimensionality reduction and single-classifier approaches (SLDA and LDA). The benefit of using decision fusion is apparent from the reduced salt-and-pepper misclassifications with the MCDF approach. For many applications, such as the Levee health monitoring problem used in this chapter, collecting ground-reference data is very expensive. This robustness to a very small training sample size is hence highly desired.

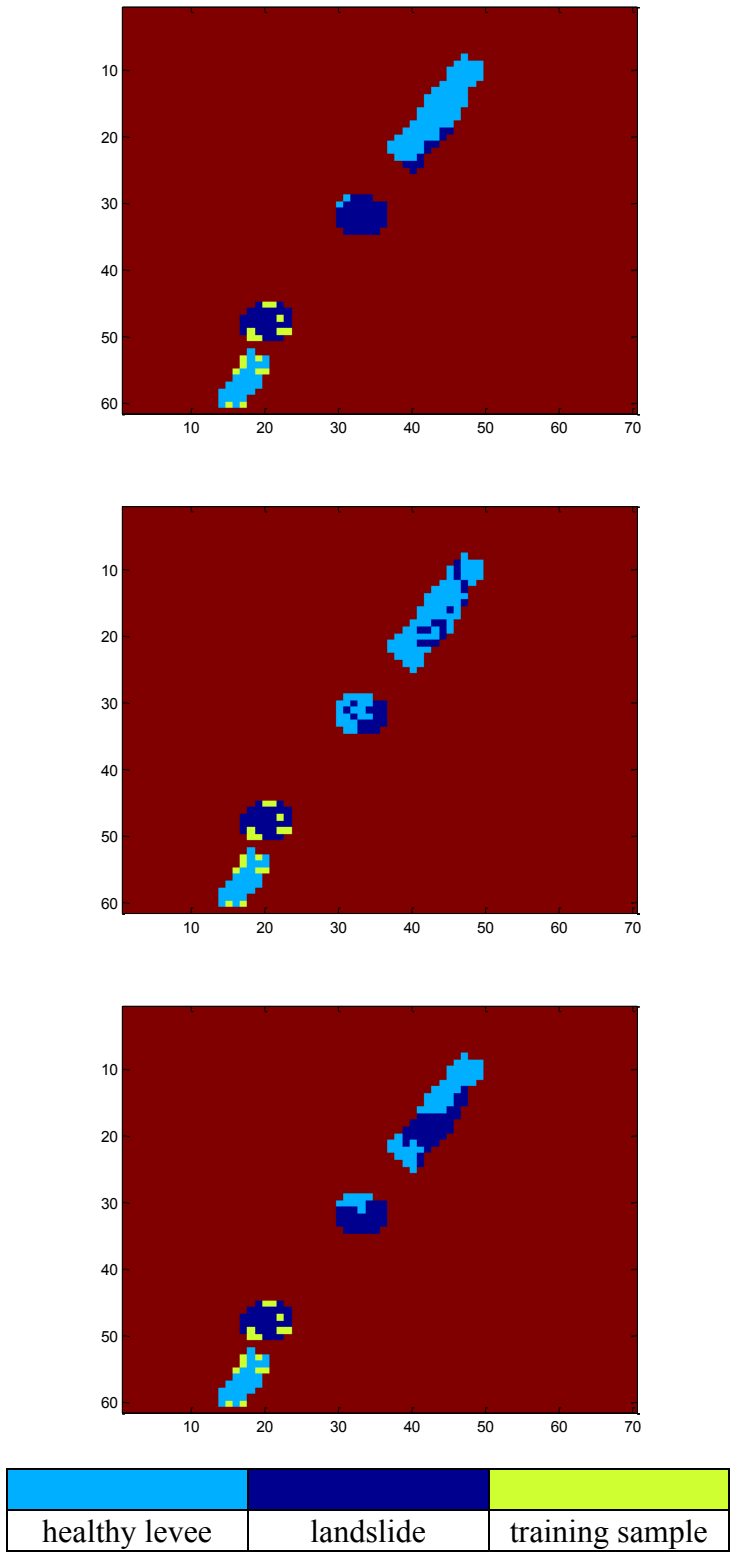

Figure 2.11 Classification map of MCDF [Top], SLDA [Middle] and LDA [Bottom].

CHAPTER III

GENETIC ALGORITHMS AND LINEAR DISCRIMINANT ANALYSIS BASED DIMENSIONALITY REDUCTION FOR REMOTELY SENSED IMAGE ANALYSIS

3.1 Introduction

Hyperspectral Imagery (HSI) capture reflected radiation over a series of contiguous bands covering a wide range of the electromagnetic spectrum for every pixel in the image. Such imagery can provide features pertinent to the classification task at hand. However, analysis methods for such imagery must first reduce the dimensionality of this very high dimensional feature space to make any classification analysis tractable. Although this high dimensional data potentially provides relevant class-specific information for image analysis, it often also results in over-dimensionality and ill-conditioned statistical formulations.

Feature subset selection is hence a useful tool when the analysis involves high dimensional feature spaces. However, identifying and selecting relevant features from a large set of features is not a trivial task. Genetic algorithms (GA) have become popular tools for various feature subset selection problems [18]. However, the most popular GA based feature selection strategy uses the training classification accuracy to optimize and find the "best" feature subset. This greedy search is only suitable for classification tasks that do not operate in high dimensional feature spaces. As the feature-space dimensionality increases, the amount of time required for such a search increases

significantly. Alternately, another class of GA-based feature subset selection algorithms employs a "filter" function that can be thought of as some metric that is optimized during the GA search. For classification problems, a good fitness function effectively measures the class-separation potential of feature subsets, thereby resulting in feature subsets that maximize class-separation.

In this work, we proposed and study two different metrics as potential filter functions for a GA based feature selection of high dimensional remotely sensed data. (1) Bhattacharyya distance (BD) [19], and (2) Fisher's ratio. In particular, we propose an algorithm called Genetic Algorithm based Linear Discriminant Analysis (GA-LDA). The main idea behind GA-LDA is that GA is employed to select an optimal feature subset, and then a LDA projection is applied on this reduced subset of features to further reduce the dimensionality of the feature space. Classification experiments demonstrate that GA-LDA is much more successful at extracting optimal features compared to conventional approaches such as SLDA and LDA.

3.2 Proposed system

Traditional feature selection techniques (such as forward selection and backward rejection) focus on evaluating the merits of each feature at a time and tend to ignore the importance of the relationship between features. The main advantage of a GA search compared with forward selection and backward rejection is that GA can take into account relationships between features. In forward selection (FS), one is unable to reevaluate the features that become irrelevant after adding some other features. Similarly, in backward rejection (BR), one is unable to reevaluate the features after they have been discarded. On

the contrary, a GA search always attempts to evaluate the merits of combinations of features and their contribution to a fitness function.

3.2.1 Genetic algorithm

Genetic algorithms are a class of optimization techniques that search for the global minimum of a fitness function. This typically involves four steps – *evaluation*, *reproduction*, *recombination*, and *mutation* which are briefly explained below. The reader is referred to [20] for a detailed description.

- *Evaluation*: In this step, a random initial set of individuals will be selected, and each individual will be evaluated by a fitness function and will be assigned a fitness value. Then, all individuals will be ranked on the basis of the fitness values.
- *Reproduction*: During this step, a number of individuals with the best fitness values in the current generation will be copied to the next generation. These individuals are called elite children.
- *Recombination*: In this step, some individuals with high fitness values other than elite children will be combined to produce new individuals. This step tries to extract best genes from different individuals and recombine them into potentially superior children.
- *Mutation*: In this step, small portions of individuals undergo mutation according to some mutations rules. This step not only prevents the algorithm from getting trapped in a local minimum but increases the likelihood that the algorithm will generate individuals with better fitness values.

After executing the evaluation step once, GA recursively goes through the reproduction, recombination and mutation steps to produce new generation until one of the stopping criteria is satisfied. Figure 3.1 illustrates the flowchart of the genetic algorithm.

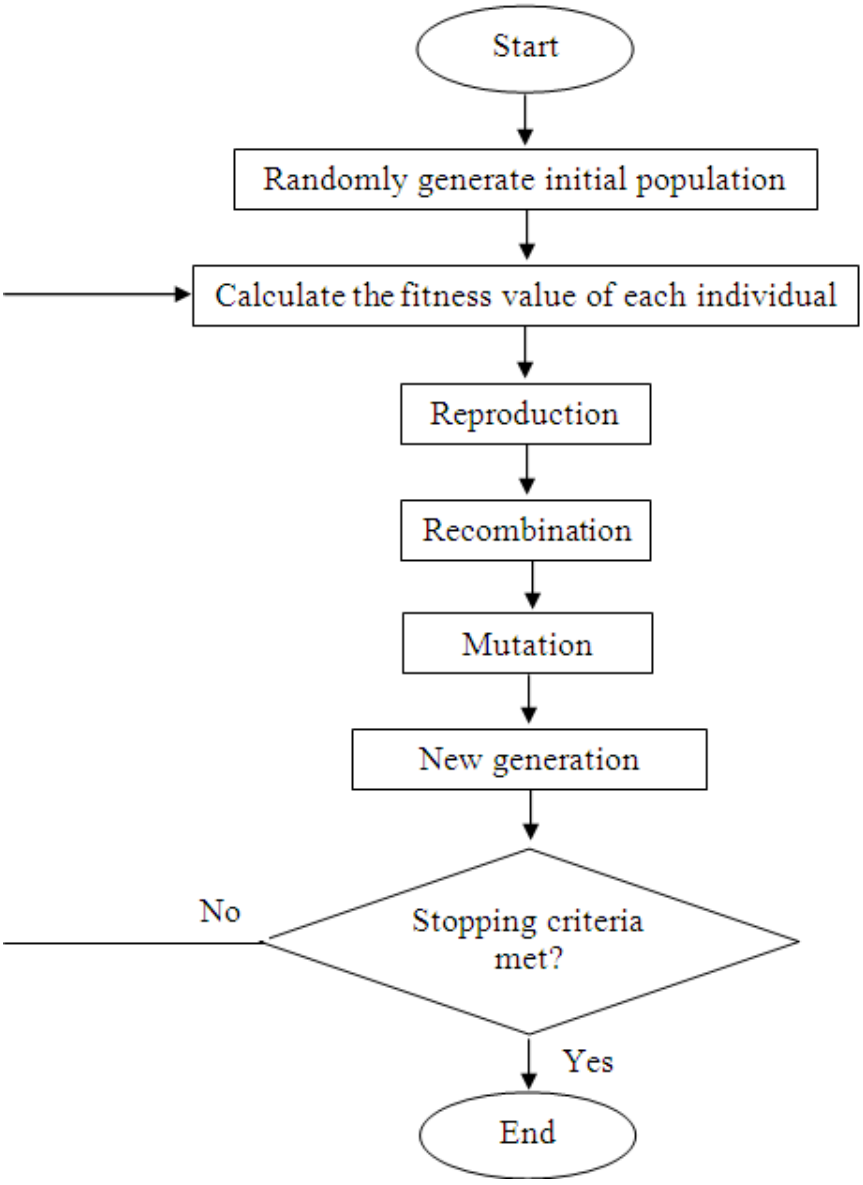

Figure 3.1 Flow chart of the genetic algorithm.

3.2.2 Fitness functions used in genetic algorithm

In this proposed algorithm, we will study two different metrics as filter functions for a GA-LDA based feature selection and dimensionality reduction: Bhattacharyya distance and Fisher's ratio.

3.2.2.1 Bhattacharyya distance

Bhattacharyya distance (BD) [20] uses the first and second order statistics to measure the separation between two probability distribution functions. For two normally distributed classes, BD is defined as

$$BD = \frac{1}{8}(\mu_2 - \mu_1)^T \left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1} (\mu_2 - \mu_1) + \frac{1}{2} ln \frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}} \quad (3.1)$$

where μ_i and $\sum_i$ are the mean vector and covariance matrix of class *i*, respectively. When BD is used as a fitness function in GA, GA will search for a feature subset that maximizes the BD value. Feature subsets producing a higher BD value are hence likely to be useful for the classification task at hand.

3.2.2.2 Fisher's ratio

Another metric we studied in this work is Fisher's ratio. LDA seeks to find a linear transformation w to a reduced dimensional subspace such that the ratio of within-class scatter to between class scatter, $J(w)$ in this projected subspace (provided by Fisher's ratio) is maximized:

$$J(w) = \frac{w^T S_B W}{w^T S_W W}, \quad (3.2)$$

where S_B and S_W are the between-class and within-class scatter matrices. When Fisher's ratio is used as a fitness function in GA, it will search for features that maximize the

Fisher's ratio, selecting a subset of features that yield the highest Fisher's ratio (and hence class-separation) after the LDA projection *w* is applied on them.

3.2.3 Genetic algorithm based linear discriminant analysis

After GA based feature selection (using BD or Fisher's ratio), we apply an LDA projection to further project the subset of features on a reduced dimensional subspace optimized for classification. This GA-LDA approach is very similar to conventional SLDA, where a forward selection/backward rejection prunes redundant and less-useful features following which an LDA projection is carried out. Such algorithms are particularly useful when LDA cannot be directly applied on the input feature space owing to its very high dimensionality. We will demonstrate with our experimental results that GA is much more efficient at such a pruning compared to traditional stepwise selection approaches. The overall block diagram of our proposed system is presented in Figure 3.2.

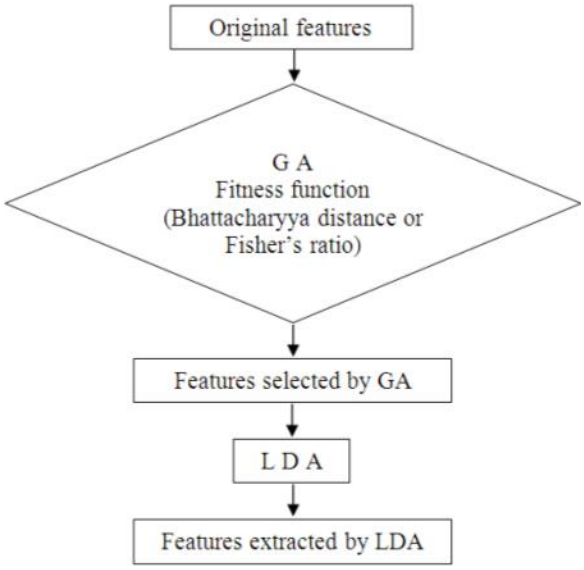

Figure 3.2 Block diagram of genetic algorithm based linear discriminant analysis used in this study.

3.3 Experimental data

The dataset employed in this study is acquired using NASA's AVIRIS sensor [21] and collected over northwest Indiana's Indian Pine test site in June 1992. The image represents a vegetation-classification scenario with 145x145 pixels (20m spatial resolution) and 220 spectral bands in the 400 to 2450 nm region of the visible and infrared spectrum. Figure 3.3 depicts the spectral signatures for the eight classes extracted from this imagery.

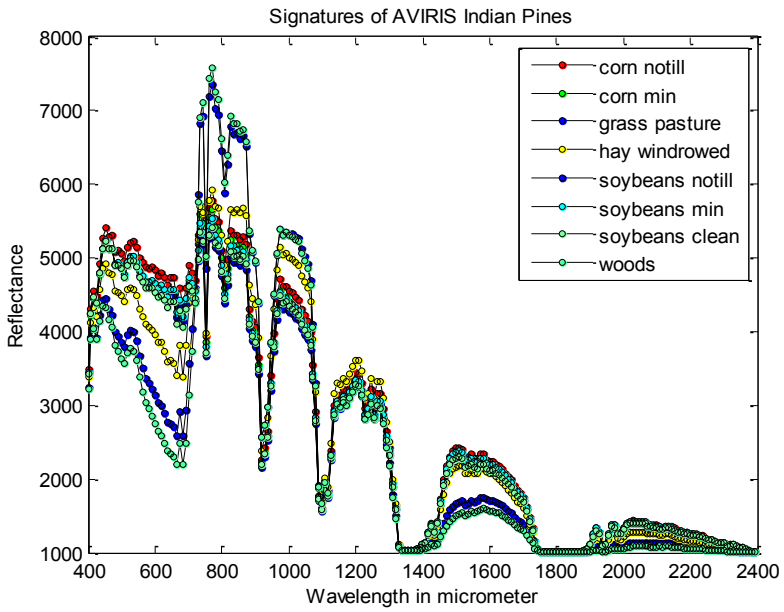

Figure 3.3 A plot of reflectance versus wavelength for eight classes of spectral signatures from AVIRIS Indian Pines data.

3.4 Experimental settings and results

All experiments reported in this study are performed using a repeated random sub-sampling validation technique. The data is split randomly into a fixed number of samples for training and other samples for testing. Such a data split is repeated 10 times and the overall recognition accuracy expressed in percentage is reported.

System parameters for the GA search are as follows. The population size is set to 100. Population size specifies how many individuals there are in each generation. The stall time limit is set to 30 seconds. The stall time limit specifies the time in seconds after which the algorithm stops if there is no improvement in the best fitness value.

Two different baseline experiments using LDA and SLDA are reported to compare the performance of the proposed system. All classifiers employed in this study are maximum-likelihood classifiers.

Figure 3.4 shows the classification accuracy versus number of training samples for the Indian Pines and corn crop datasets. From this figure, it is clear that the proposed GA-LDA based feature extraction technique outperforms the other two "baseline" approaches to classification, especially when the number of available training samples is small. In other words, the GA-LDA approach is very effective at identifying the most relevant features compared to SLDA and LDA.

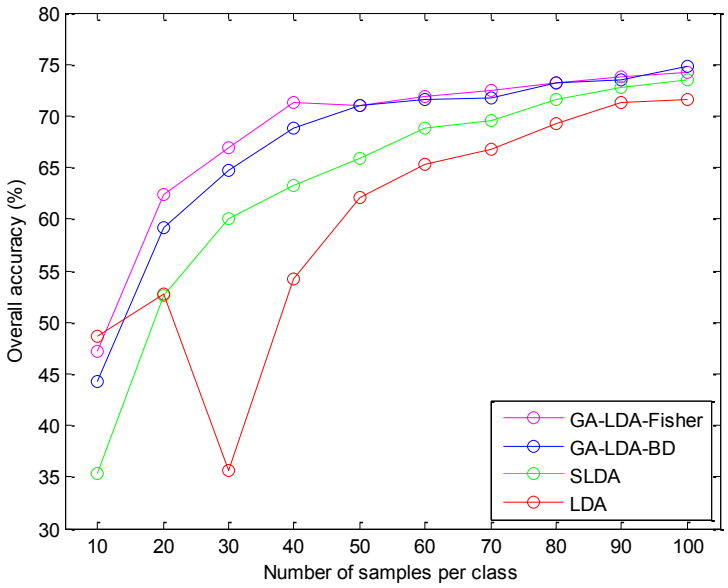

Figure 3.4 Illustrating the benefits of the proposed GA-LDA feature reduction approach for the Indian pines dataset. GA-LDA-Fisher: Using fisher's ratio as the fitness function in GA-LDA; GA-LDA-BD: Using BD as the fitness function in the GA-LDA.

In Figure 3.5, the standard deviation and the execution time of the experiment versus the number of training samples of the four algorithms are plotted. From this figure, it can be seen that the four algorithms show similar variability with regards to different sizes of training samples, and even GA based feature selection using BD as a fitness function is computationally a little more expensive than others. However, using Fisher's ratio as a fitness function in the GA search is not computationally expensive compared to the traditional feature selection method SLDA.

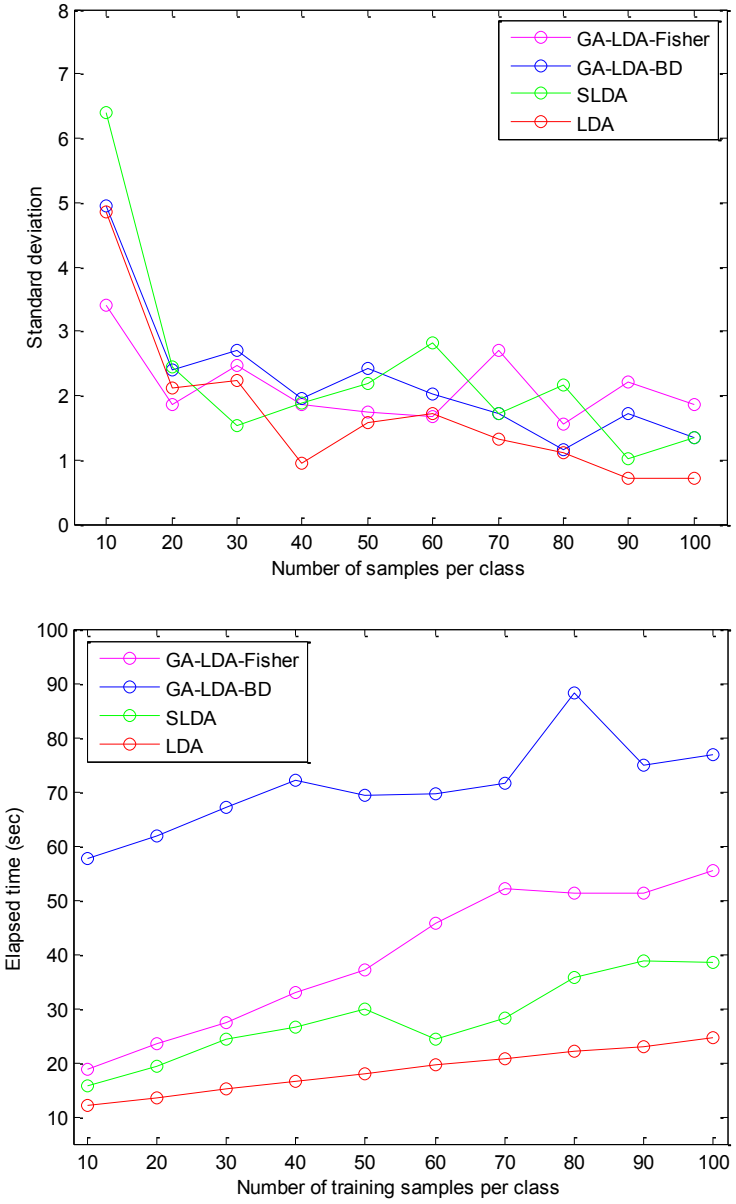

Figure 3.5 Plots of standard deviation [Top] and experiment elapsed time [Bottom] versus number of training samples of four different algorithms.

In order to test the ability of the proposed system to select pertinent features in much higher dimensional feature space, this dataset is re-quantized into 256 grey levels to extract texture features using GLCM described in Chapter 2. Six common texture features used in GLCM are described in Table 2.1. GLCM is computed over a 9 pixels by 9 pixels window, and a pixel distance of 1 and an angle of 0 is used in GLCM. After

feature extraction using GLCM, the number of features of the Indian Pines data is enlarged to1320.

Figure 3.6 shows the classification accuracy versus number of training samples for both of the datasets that including texture features only. From this figure, we can infer that the GA-LDA approach for the extremely high dimensional data is much better than the traditional SLDA and LDA approaches, and the improvement is specifically more pronounced when the available amount of training sample-size is small. The standard deviation and the execution time of the experiment versus the number of training samples of the four algorithms are plotted in Figure 3.7. In a much higher dimensional feature space, our proposed methods show lower standard deviations compared to SLDA and LDA when the training sample size is small, and the experimental computation time is significantly less than SLDA.

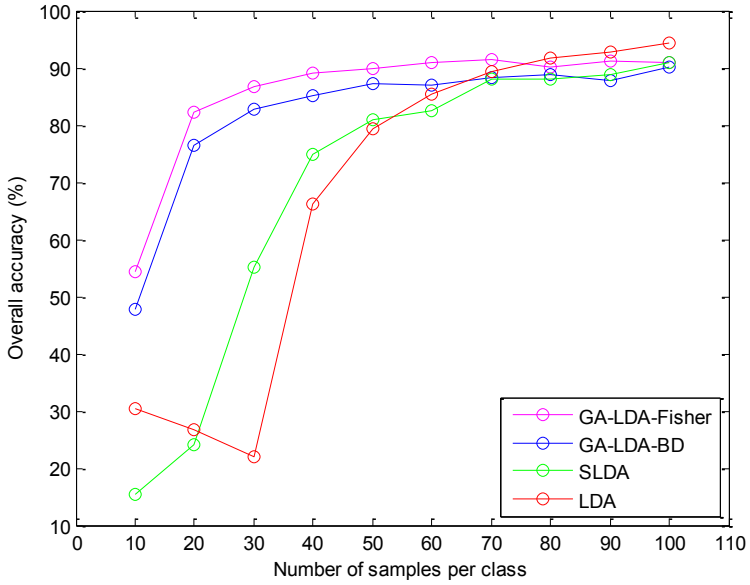

Figure 3.6 Illustrating the benefits of the proposed GA-LDA feature reduction approach for Indian pines dataset using texture features only.

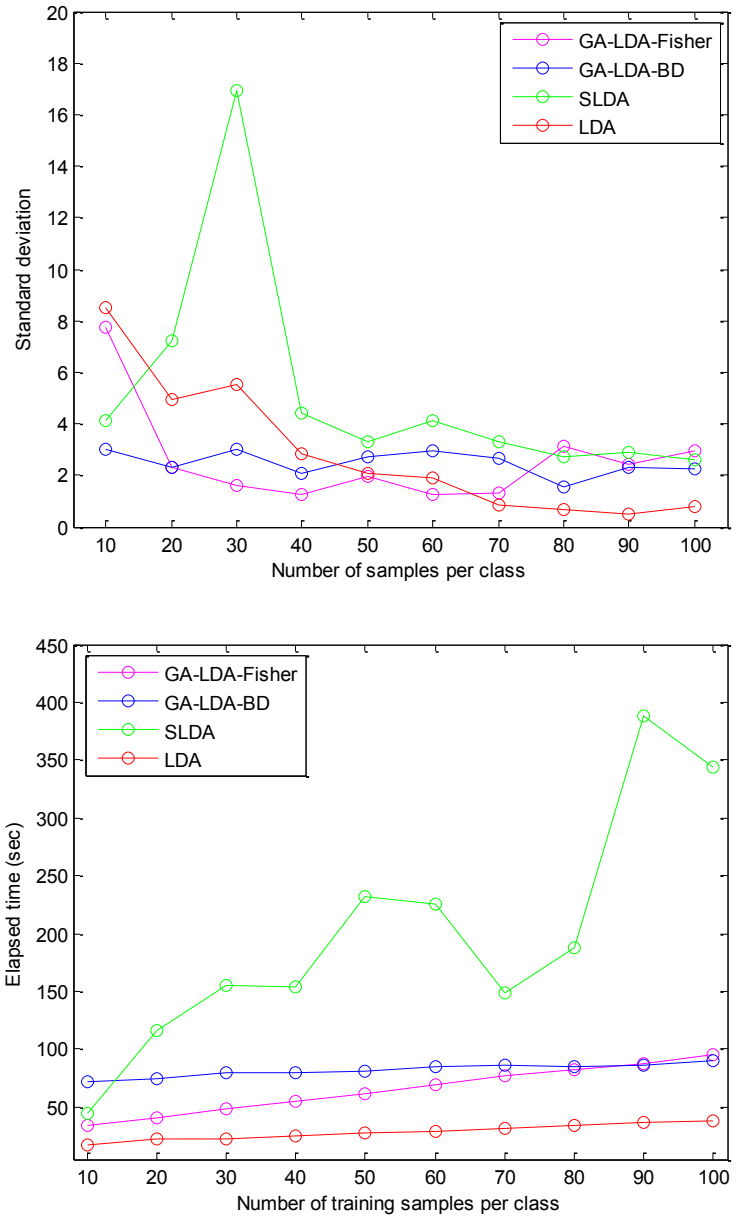

Figure 3.7 Plots of standard deviation [Top] and experiment elapsed time [Bottom] versus the number of training samples of four different algorithms.

3.5 Conclusions

Experimental results presented in this paper indicate that a GA search is very effective at selecting the most pertinent features, while pruning out the most redundant features for classification tasks when an appropriate fitness function is employed. Akin to

the conventional stepwise-LDA approach, we proposed a GA-LDA approach where GA first identifies a smaller subset of features upon which LDA is applied for final dimensionality reduction. Given a moderate feature space dimensionality and sufficient training samples, LDA is a good projection based dimensionality reduction strategy. However, as the number of features increases and the training-sample-size decreases, methods such as GA-LDA can assist by providing a robust intermediate step of pruning away redundant and less useful features. Consistent improvements in classification performance when using GA-LDA can be noted in our results. Finally, although the Fisher's ratio and BD provide similar information (by quantifying the class-separation ability of features), since LDA optimizes the Fisher's ratio, we note that GA-LDA-Fisher slightly outperforms GA-LDA-BD. This is expected because when the GA uses Fisher's ratio to perform its search, the final subset of features it identifies is already optimizing Fisher's ratio, resulting in a slightly better performance when LDA is applied on these features.

CHAPTER IV

CONCLUSIONS

In this thesis, the problem of information-extraction from remotely sensed imageries is addressed. The two proposed algorithms aim at analyzing two different types of remotely sensed imagery, namely hyperspectral and synthetic aperture radar (SAR) imagery. The growing availability of such remotely sensed imagery provides a robust characterization of the earth's surface, but requires improved image processing techniques. Our work involves the use of genetic algorithm and multi-classifier decision fusion techniques for designing robust pattern recognition algorithms for hyperspectral and synthetic aperture radar imagery operating under small sample size conditions. The main conclusion of each algorithm is drawn below.

The problem of classification of healthy levee versus landslide using SAR data was addressed in this thesis. Owing to its high resolution and growing availability, SAR image has a wide area of application in remote sensing. In order to exploit more class discriminative information contained in the SAR backscatter, one of the most frequently used texture feature extraction technique - grey level co-occurrence matrix (GLCM) has been utilized in this thesis. Texture features derived from GLCM not only provide valuable information for classification, but sometimes also result in very high dimensional feature spaces which makes model parameter estimation ill-conditioned. To efficiently address this problem, polarization channel based feature grouping followed by multi-classifier decision fusion technique has been proposed and studied in this thesis.

Experiment results show that our proposed method can robustly classify landslides between healthy levee segments using a small number of training samples.

The primary advantage of hyperspectral (HSI) imagery is that, it has hundred or thousand of spectral features which can potentially provide unique spectral signatures of each category / class. However, the major hurdle that many researchers have had to face is finding or extracting useful information from this high-dimensional feature space where relevant and irrelevant features coexist. Thus, we employ genetic algorithm combined with Bhattacharyya distance or Fisher's ratio as a fitness functions to search the most relevant features followed by linear discriminant analysis (LDA) to classify HSI data. From the experimental results show in Chapter 3, we conclude that our proposed algorithm is very efficient at selecting the most pertinent features among these hundreds or thousands of spectral features. Moreover, we noticed that the usefulness of our proposed method is more pronounced when the number of training samples decreases.

REFERENCES

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

[2] S. Prasad and L. M. Bruce, "Overcoming the Small Sample Size Problem in Hyperspectral Classification and Detection Tasks," *Proceedings of the IEEE International Geoscience.and Remote Sensing Symposium*, Boston, MA, July 2008, pp. 381- 384.

[3] D. R. Paudyal, A. Eiumnoh and J. Aschbacher, "Textural Information in SAR Images for Land-Cover Applications," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Toronto, Canada, August 2002, pp. 1020-1022.

[4] T. M. Kuplich, P. J. Curran and P. M. Atkinson, "Relating SAR image texture and backscatter to tropical forest biomass," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Toulouse, France, May 2004, pp. 2872-2874.

[5] S. Kumar, J. Ghosh and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, July 2001, pp. 1368-1379.

[6] S. Nakariyakul, D. P. Casasent, "Improved forward floating selection algorithm for feature subset selection," *Proceedings of the IEEE International Conference on Wavelet Analysis and Pattern Recognition*, Hong Kong, China, August 2008, pp. 793-798.

[7] S. Prasad and L. M. Bruce, "A multi-classifier and decision fusion framework for robust classification of mammographic masses," *Proceedings of the 30th Annual IEEE EMBS International Conference,* Vancouver, Canada, October 2008, pp. 3048-3051.

[8] R. O. Duda, P. E. Hart and D. G. Stock, Pattern Classification, 2 edition, John Wiley and Sons, Inc., New York, NY, 2001.

[9] Tran Huy Dat and Cuntai Guan, "Feature selection based on fisher ratio and mutual information analysis for robust brain computer interface," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, June 2007, pp. 337-340.

[10] Chulhee Lee and Daesik Hong, "Feature Extraction Using the Bhattacharyya Distance," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, Orlando, Florida, October 1997, pp. 2147-2150.

[11] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural Features for Image Classification," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, Newport Beach, CA, November 2007, pp. 610-621.

[12] A. Baraldi and F. Parmiggiani, "An investigation of the textural characteristics associated with the gray level cooccurrence matrix statistical parameters," *IEEE Transactions on Geoscience and Remote Sensing*, August 2002, pp. 293-304.

[13] M. Petrakos, J. A. Benediktsson, and I. Kanellopoulos, "The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion," *Transactions on Geoscience and Remote Sensing*, November 2001, pp. 2539-2546.

[14] J. V. Aanstoos, K. Hasan, C. G. O'Hara, S. Prasad, L. Dabbiru, R. Nobrega, L. Matthew and B. Shrestha, "Use of Remote Sensing to Screen Earthen Levees," *Proceedings of the 39th IEEE Applied Imagery Pattern Recognition Workshop*, Washington, DC, April 2011, pp. 1-6.

[15] A. K. M. A. Hossain, G. Easson and K. Hasan, "Detection of Levee Slides Using Commercially Available Remotely Sensed Data", *Environmental and Engineering Geoscience*, vol. 12, no. 3, August, 2006, pp. 235-246.

[16] Jiancheng Shi, J. Wang, A. Hsu, P. O'Neill and E.T. Engman, "Estimation of bare soil moisture and surface roughness parameter using L-band SAR image data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 5, August 2002, pp. 1254-1266.

[17] Kevin Wheeler, Scott Hensley, Yunling Lou, Tim Miller and Jim Hoffman, "An L-band SAR for repeat pass deformation measurements on a UAV platform," *Proceedings of the IEEE International Conference on Radar,* Philadelphia, PA, April 2004, pp. 317-322.

[18] Ho-Duck Kim, Chang-Hyun Park, Hyun-Chang Yang and Kwee-Bo Sim, "Genetic Algorithm Based Feature Selection Method Development for Pattern Recognition," *Proceedings of the SICE-ICASE International Joint Conference*, Busan, Korea, February 2007, pp. 1020-1025.

[19] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathmatical Society*, vol 35, 1943, pp. 99-109.

[20] K.S. Tang, K.F. Man, S. Kwong and Q. He, "Genetic algorithms and their applications," *IEEE Signal Processing Magazine*, vol. 13, no. 6, August 2002, pp. 22-37.

[21] NASA Jet Propulsion Laboratory Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) web page http://aviris.jpl.nasa.gov