Heaven's Light is Our Guide

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Rajshahi University of Engineering & Technology, Bangladesh

Target Class Oriented Feature Selection for Effective Hyperspectral Image Classification

Author

Md. Tanvir Ahmed

Roll No. 123086

Department of Computer Science & Engineering

Rajshahi University of Engineering & Technology

Supervised by

Dr. Md. Ali Hossain

Assistant Professor

Department of Computer Science & Engineering

Rajshahi University of Engineering & Technology

ACKNOWLEDGEMENT

I would like to express my special appreciation and thanks to my supervisor **Dr. Md. Ali Hossain**, Assistant Professor, Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, you have been a tremendous mentor for me. Again I would like to thank you for encouraging this research. Your advice on both research as well as on my career have been priceless.

I would also like to express my sincere appreciation & deepest sense of gratitude to my honorable teacher **Dr. Md. Rabiul Islam**, Head of the Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology for his email support regarding this research. Finally, thanks to all of my honorable teachers, friends & well-wishers for their great role to do complete this research.

Date: November 21, 2017 Md. Tanvir Ahmed
RUET, Rajshahi

Heaven's Light is Our Guide

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Rajshahi University of Engineering & Technology, Bangladesh

CERTIFICATE

With immense pleasure, it is hereby certified that the thesis **Target Class Oriented Feature Selection for Effective Hyperspectral Image Classification**, is prepared by **Md. Tanvir Ahmed**. Roll. 123086, has been carried out under my supervision. The thesis has been prepared in partial fulfillment of requirements for degree of Bachelor of Science in Computer Science and Engineering.

Supervisor

____

Dr. Md. Ali Hossain
Assistant Professor
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204

External Examiner

____

Dr. X khan
Assistant Professor
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204

Abstract

Hyperspectral sensors are devices that acquire images over hundreds of spectral bands. It has wide range of applications in ground object detection, which makes hyperspectral image analysis an important field of research. Most informative features selection for effective hyperspectral image classification is a difficult task. Different approaches has been proposed to select only relevant features from those large correlated data set. The problem can be address using feature extraction and feature selection. Principal component analysis (PCA) is one of the most popular feature extraction technique, though its components is not always suitable for better classification accuracy. This research proposed a target class oriented features selection method which uses normalized mutual information (nMI) measure with two constraints to maximize general relevance and minimize redundancy on the components obtained via PCA. In this research the proposed feature mining approach is combined with kernel support vector machine (SVM) classifier for the effective classification object. Target class oriented features selection approach shows significant improvement in terms of classification accuracy 97.43% of real hyperspectral data. A comparison among relevant and recent feature selection techniques in terms of their classification accuracy is provided using hyperspectral image.

Contents

Acknowledgement **i**

Certificate **ii**

Abstract **iii**

List of Figures **v**

List of Tables **vi**

1 Introduction **1**

1.1 Overview 1

1.2 Remote Sensing 2

1.3 Challenges of Remote Sensing Data 2

1.4 Motivation 3

1.5 Objectives 3

1.6 Organization 4

2 Background **5**

2.1 Introduction 5

2.2 Spectral Image Basics 5

2.3 Hyperspectral Data 8

2.4 Sensors 9

2.5 Application of Hyperspectral Image Analysis 10

2.6 Problems of Hyperspectral Data 12

3 Literature Review **14**

3.1 Introduction 14

3.2 Feature Mining 15
3.3 Related Works 17
3.4 Feature Relevancy and Redundancy 19
3.5 General Approach for Feature Selection 19
3.6 Feature Mining for Hyperspectral Image 20
3.7 Classification Techniques 21
3.7.1 Unsupervised Classification 22
3.7.2 Supervised Classification 23

4 Feature Selection and Classification **24**
4.1 Feature Extraction Based on Principal Component Analysis 24
4.2 Feature Selection Based on Normalized Mutual Information 25
4.2.1 Original Dataset Plus nMI Approach 26
4.2.2 PCA Dataset Plus nMI Approach 26
4.3 Proposed Method 26
4.4 Implementation 27
4.4.1 Implementation of PCA Based Feature Extraction 27
4.4.2 Implementation of nMI Based Feature Selection 27
4.4.3 Implementation of Target Class Oriented Feature Selection 27
4.5 Experimental Results 27
4.5.1 Indian pines (92AV3C) Dataset 27
4.5.2 Feature Extraction Result 27
4.5.3 nMI Based Feature Selection Result 27
4.5.4 Target Class Oriented Feature Selection Result 27
4.6 Classification Results 27

5 Conclusion and Future Works **28**
5.1 Conclusion 28
5.2 Future Works 28

References **29**

List of Figures

1.1 Remote Sensing 2
1.2 Curse of dimensionality 3

2.1 The electromagnetic spectrum 6
2.2 Reflectance spectra for several common Earth surface materials 6
2.3 Reflectance spectra for important minerals 7
2.4 The concept of hyperspectral imagery. Image measurements are made at many narrow contiguous wavelength bands, resulting in a complete spectrum for each pixel. 7
2.5 Reflectance spectra of the three materials as they would appear to the hyperspectral AVIRIS sensor. The gaps in the spectra are wavelength ranges at which the atmosphere absorbs so much light that no reliable signal is received from the surface. 8

3.1 Four key steps for the feature selection process 16
3.2 Filter Method for feature selection 19
3.3 Wrapper Method for Feature selection 19
3.4 Embedde Method for Feature selection 20
3.5 Possible operations for feature mining (only one path is used for any specific approach). 21
3.6 Unsupervised Classification 22
3.7 Supervised Classificationn 23

List of Tables

2.1 Current and Recent Hyperspectral Sensors and Data Providers 10

Chapter 1

Introduction

1.1 Overview

Hyperspectral sensors simultaneously measure hundreds of continuous spectral bands with a fine resolution to form a three dimensional hyperspectral image data cube. For instance, the AVIRIS sensor simultaneously measures 224 bands with a fine resolution of 0.01μm. This high data volume presents many challenges which creates opportunity for research. The data captured are highly correlated and contains a significant amount of redundant data. All the image bands are not equally important for specific application. Also, as the feature space dimension increases, if the size of the training data does not grow correspondingly, a reduction in the classification accuracy of the testing data is observed due to poor parameter estimation of the supervised classifier. This effect is known as the Hughes phenomenon. So it is required to extract only relevant features from the input dataset. Therefore an effective and efficient technique to find this relevant features is a major interest in current literature. Principal component analysis (PCA) is one of the most popular feature extraction technique, though its components is not always suitable for better classification accuracy. It is also not sensitive to input classes and consider only the global variance of the dataset. There are few techniques for finding relevant features in current literature. For example, mutual information based feature selection and in combination of principal component analysis and normalized mutual information based feature selection. A target class oriented feature selection technique is proposed as an alternative for the effective subspace detection in collaboration with kernel support vector machine (SVM) to achieve better classification accuracy.

1.2 Remote Sensing

Remote sensing is the acquisition of information about an object or phenomenon without making physical contact with the object. Remote sensing is used in numerous fields, including geography, land surveying and most Earth Science disciplines (for example, hydrology, ecology, oceanography, glaciology, geology); it also has military, intelligence, commercial, economic, planning, and humanitarian applications.

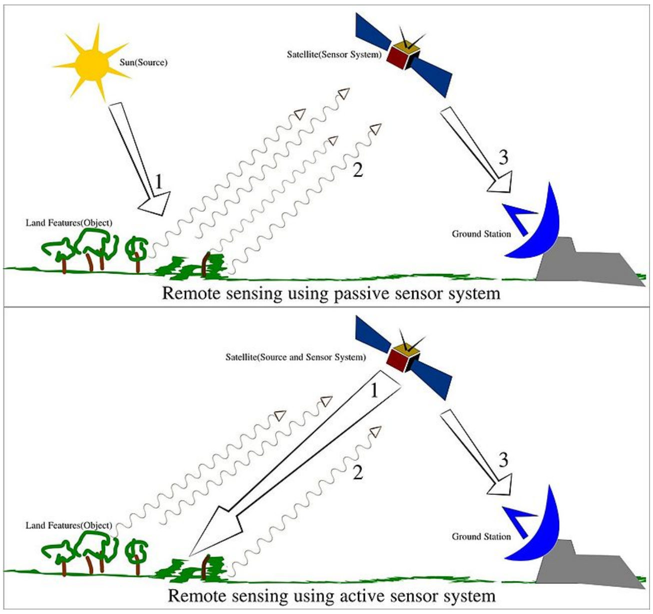

Figure 1.1: Remote Sensing

In current usage, the term "remote sensing" generally refers to the use of satellite- or aircraft-based sensor technologies to detect and classify objects on Earth, including on the surface and in the atmosphere and oceans, based on propagated signals (e.g. electromagnetic radiation). It may be split into "active" remote sensing (i.e., when a signal is emitted by a satellite or aircraft and its reflection by the object is detected by the sensor) and "passive" remote sensing (i.e., when the reflection of sunlight is detected by the sensor).

1.3 Challenges of Remote Sensing Data

Hyperspectral imaging is a popular field of remote sensing. The main challenge of hyperspectral image is its highly correlated huge data set. On the other hand, this large number of spectral bands has a direct impact on the required computational cost for classification. Also, as the feature space dimension increases, if the size of the training data does not grow correspondingly, a reduction in the classification accuracy of the testing data is observed due to poor parameter estimation of the supervised classifier. This effect is known as the Hughes phenomenon [1].

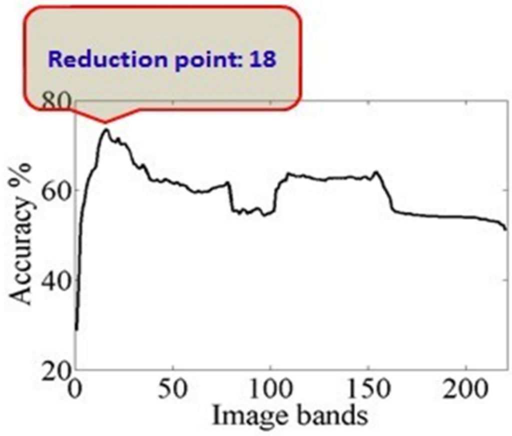

Figure 1.2: Curse of dimensionality

1.4 Motivation

Hyperspectral image processing is fast growing research field because of its various application. Observed area of hyperspectral images are classified into different groups of object by classification of the image. Feature extraction and selecting only relevant features before classification is an important task to achieve high classification accuracy [2]. So relevant feature selection is very important for classification of hyperspectral images.

1.5 Objectives

The main goal of this research is to illustrate a target class oriented feature mining method which is a combination of feature extraction and feature selection which gives a better classification accuracy.

Objective of this research will be:

- Create a method of feature extraction by removing correlation among bands.

- Create an approach to select features after feature extraction.
- Create a method with collaboration of feature mining and classification.
- Improving classification accuracy.

1.6 Organization

This report is organized in 5 chapters discussing all related topics that may be helpful in reproducing a feature selection method for hyperspectral image classification.

Chapter 1

A short overview of the whole research field and topic is discussed in this chapters.

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 2

Background

2.1 Introduction

The most significant recent breakthrough in remote sensing has been the development of hyperspectral sensors and software to analyze the resulting image data. Fifteen years ago only spectral remote sensing experts had access to hyperspectral images or software tools to take advantage of such images. Over the past decade hyperspectral image analysis has matured into one of the most powerful and fastest growing technologies in the field of remote sensing. The "hyper" in hyperspectral means "over" as in "too many" and refers to the large number of measured wavelength bands. Hyperspectral images are spectrally overdetermined, which means that they provide ample spectral information to identify and distinguish spectrally unique materials. Hyperspectral imagery provides the potential for more accurate and detailed information extraction than possible with any other type of remotely sensed data.

2.2 Spectral Image Basics

To understand the advantages of hyperspectral imagery, it may help to first review some basic spectral remote sensing concepts. You may recall that each photon of light has a wavelength determined by its energy level. Light and other forms of electromagnetic radiation are commonly described in terms of their wavelengths. For example, visible light has wavelengths between 0.4 and 0.7 microns, while radio waves have wavelengths greater than about 30 cm (Figure. 2.1).

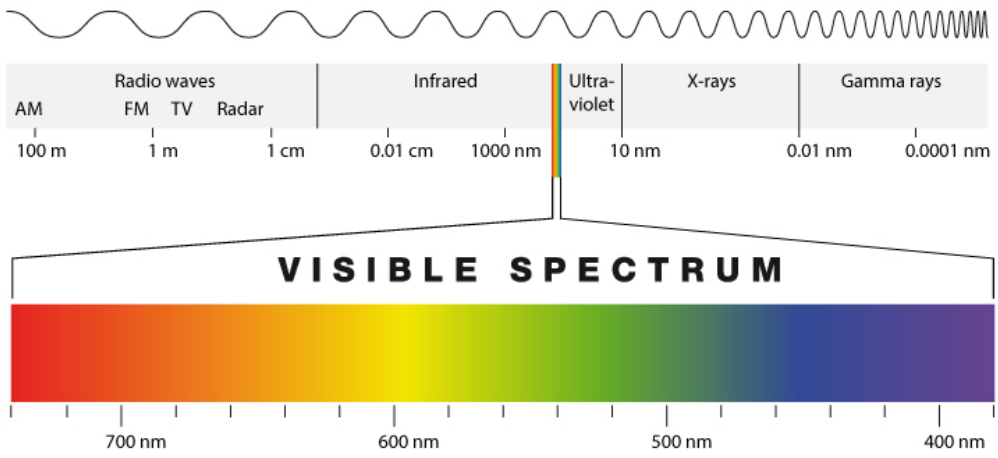

Figure 2.1: The electromagnetic spectrum

Reflectance is the percentage of the light hitting a material that is then reflected by that material (as opposed to being absorbed or transmitted). A reflectance spectrum shows the reflectance of a material measured across a range of wavelengths (Fig. 2.2). Some materials will reflect certain wavelengths of light, while other materials will absorb the same wavelengths. These patterns of reflectance and absorption across wavelengths can uniquely identify certain materials.

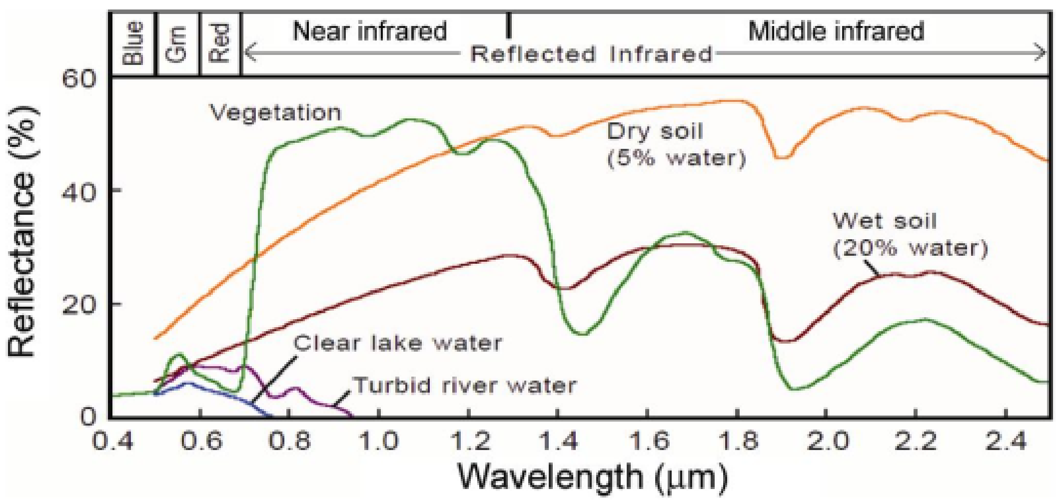

Figure 2.2: Reflectance spectra for several common Earth surface materials

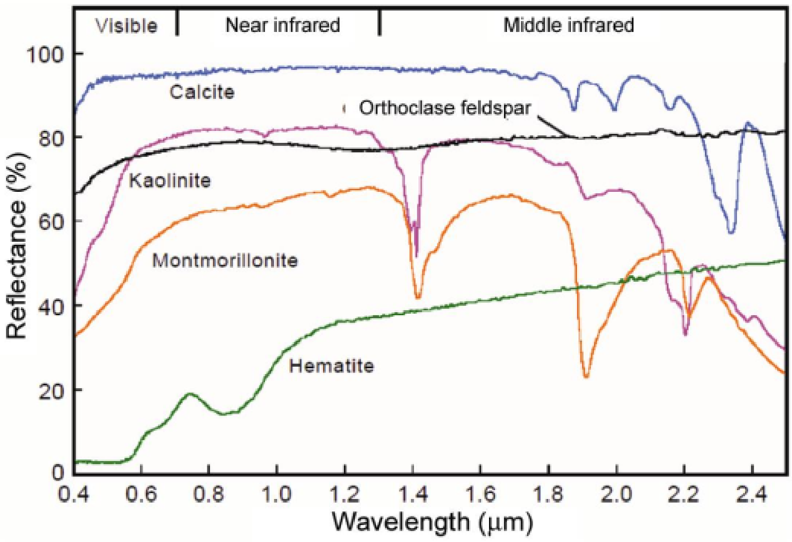

Figure 2.3: Reflectance spectra for important minerals

Field and laboratory spectrometers usually measure reflectance at many narrow, closely spaced wavelength bands, so that the resulting spectra appear to be continuous curves (Fig. 2.2). When a spectrometer is used in an imaging sensor, the resulting images record a reflectance spectrum for each pixel in the image (Fig. 2.3).

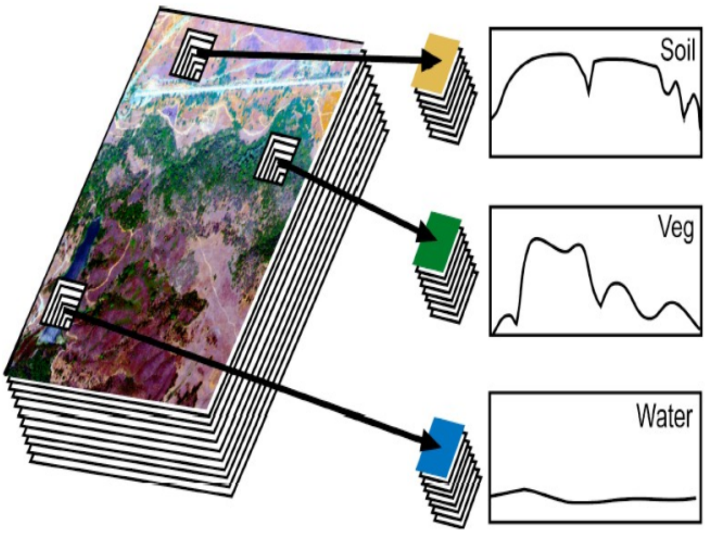

Figure 2.4: The concept of hyperspectral imagery. Image measurements are made at many narrow contiguous wavelength bands, resulting in a complete spectrum for each pixel.

2.3 Hyperspectral Data

Most multispectral imagers (e.g., Landsat, SPOT, AVHRR) measure radiation reflected from a surface at a few wide, separated wavelength bands (Fig. 4). Most hyperspectral imagers (Table 1), on the other hand, measure reflected radiation at a series of narrow and contiguous wavelength bands. When we look at a spectrum for one pixel in a hyperspectral image, it looks very much like a spectrum that would be measured in a spectroscopy laboratory (Fig. 5). This type of detailed pixel spectrum can provide much more information about the surface than a multispectral pixel spectrum.

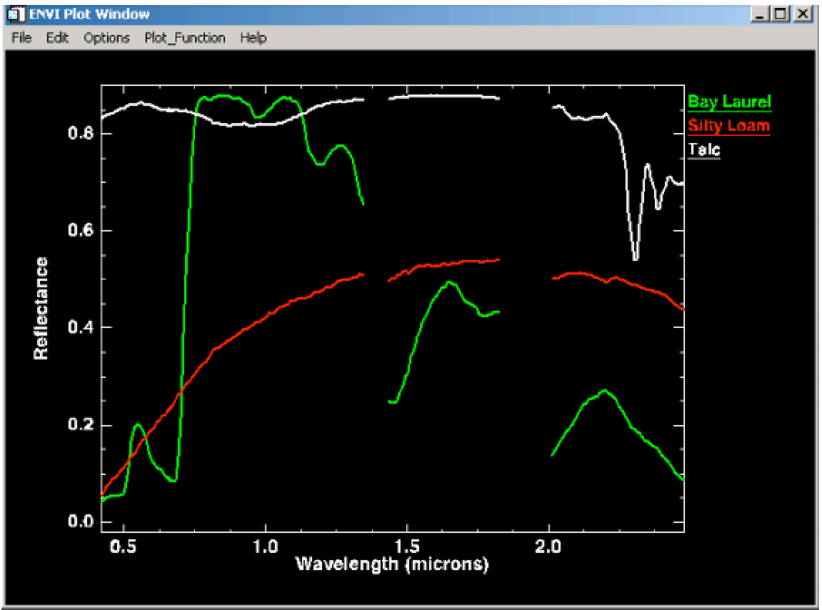

Figure 2.5: Reflectance spectra of the three materials as they would appear to the hyperspectral AVIRIS sensor. The gaps in the spectra are wavelength ranges at which the atmosphere absorbs so much light that no reliable signal is received from the surface.

Although most hyperspectral sensors measure hundreds of wavelengths, it is not the number of measured wavelengths that defines a sensor as hyperspectral. Rather it is the narrowness and contiguous nature of the measurements. For example, a sensor that measured only 20 bands could be considered hyperspectral if those bands were contiguous and, say, 10 nm wide. If a sensor measured 20 wavelength bands that were, say, 100 nm wide, or that were separated by non-measured wavelength ranges, the sensor would no longer be considered hyperspectral.

Standard multispectral image classification techniques were generally developed to classify multispectral images into broad categories. Hyperspectral imagery provides an opportunity for more detailed image analysis. For example, using hyperspectral data, spectrally similar materials can be distinguished, and sub-pixel scale information can be extracted. To fulfill this potential, new image processing techniques have been developed.

2.4 Sensors

Most past and current hyperspectral sensors have been airborne (Table 1), with two recent exceptions: NASA's Hyperion sensor on the EO-1 satellite, and the U.S. Air Force Research Lab's FTHSI sensor on the MightySat II satellite. Several new spacebased hyperspectral sensors have been proposed recently (Table 2). Unlike airborne sensors, space-based sensors are able to provide near global coverage repeated at regular intervals. Therefore, the amount of hyperspectral imagery available should increase significantly in the near future as new satellite-based sensors are successfully launched.

Table 2.1: Current and Recent Hyperspectral Sensors and Data Providers

Sensors	Manufacturer	Number of Bands	Spectral Range
FTHSI on MightySat II	Air Force Research Lab www.vs.afrl.af.mil /Tech-Progs/MightySatII	256	0.35 to 1.05 μm
Hyperion on EO- 1	NASA Goddard Space Flight Center eo1.gsfc.nasa.gov	220	0.4 to 2.5 μm
AVIRIS (Airborne Visible Infrared Imaging Spectrometer)	NASA Jet Propulsion Lab makalu.jpl.nasa.gov/	224	0.4 to 2.5 μm
HYDICE (Hyperspectral Digital Imagery Collection Experiment)	Naval Research Lab	210	0.4 to 2.5 μm
PROBE- 1 Earth Search Sciences Inc.	www.earthsearch .com	128	0.4 to 2.5 μm
casi (Compact Airborne Spectrographic Imager)	ITRES Research Limited www.itres.com up to	228	0.4 to 1.0μm

2.5 Application of Hyperspectral Image Analysis

Multi spectral images like Landsat thematic Mapper and SPOT XS produce few broad wavelength bands but the hyperspectral image sensor will produce narrower wavelength bands. These bands can range from 100 to 200 or more. These measurement make it possible to derive continuous spectrum for each image cell, see below image. After sensor is adjusted, terrain and atmospheric condition are applied, then the results is verified against the collected spectrum value to categorize the type of vegetation or minerals or other features. Hyperspectral images contain ton of information, surface information and its spectrum behavior should be understand deeply and how it related to the hyperspectral images. This type of image are finding their importance in different fields as before it was just used for remote sensing application. Here are few applications of hyperspectral images.

1. **Remote Sensing:** In remote sensing technology it is very important to distinguish earth surface features, each features have different spectrum band. Multi spectral satellite can capture image up few bands for example Landsat 7 have 8 bands. But multi spectral imaging satellite can capture earth surface in more than 200 bands which helps scientist to differentiate objects that were not possible in multi spectral imaging because of spectral resolution.

2. **Seed Viability Study:** By using the hyperspectral image and plotting the reflectance spectrum one can conclude that whether those seed are viable or not viable. Seed might be looking same through naked eyes but its viability will be trace down by the hyperspectral image.

3. **Biotechnology:** Hyperspectral technology has become popular in the biological and medical applications. It is easy and quick to acquire the data that can be used in the laboratory. Mostly they are used in the study of the wound analysis, fluorescence microscopy, and cell biology.

4. **Environmental Monitoring:** Hyperspectral imaging is becoming widely popular for tracking changes in the environment. It is commonly used to understand surface CO2 emissions, map hydrological formations, tracking pollution levels, and more.

5. **Food:** Hyperspectral imaging is widely used in the food sector. It is used in different discipline of food industry, bruise detection in apples, freshness of the fish, citrus fruit inspection, distribution of sugar in melons, and sorting of potatoes. For example apple bruise is not visible on the early stage and it takes few days to show dark color mark. In this type of scenario hyperspectral imaging techniques can be used to track the early stage of bruise for the quality control.

6. **Pharmaceuticals:** Hyperspectral imaging technique is widely used to enhance the quality control. It is used widely to control the counterfeit or illegal drugs, managing the packaging of medicine and mixing of the powder.

7. **Medical Diagnose:** Early disease detection and disease prevention are very important for the healthy body. Hyperspectral imaging technology can be used to detect the early of various types of cancer or retinal disease.

2.6 Problems of Hyperspectral Data

Optical sensing has come a long way from grayscale to multispectral and now to hyperspectral images. The advances in imaging hardware over recent decades have enabled availability of high spatial, spectral, and temporal resolution imagery for a variety of applications. Hyperspectral imagery, also called imaging spectroscopy, entails acquiring images using a large number (typically a few hundreds) of narrow and often contiguous spectral bands, covering a wide range of the electromagnetic spectrum from the visible to the infrared regions. Compared to conventional color imagery (with 3 spectral bands covering the red, green and blue wavelengths, respectively), or compared to conventional multispectral imagery (typically a few spectral bands), hyperspectral data provide a very fine spectral characterization of the sensed materials, which facilitates their detection and characterization. Advances in hardware to acquire hyperspectral imagery have made such data easily accessible to a wide variety of application domains, but have also created unique challenges for researchers working on algorithms for the representation, exploitation, and analysis of such data. Unfortunately, this is often a double-edged sword. A direct consequence of the dense spectral sampling implies that each measurement corresponds to a vector with several hundreds of values. Consequently, the data are evolving in a vector space with several hundreds of dimensions. Traditional information-processing techniques cannot be used to process such data effectively. While it is a curse from an analytical, theoretical, and statistical point of view, the very high dimensionality of the data is also a blessing. Challenges of hyperspectral images are given below:

- Curse of dimensionality.
- High correlation.

- Irrelevant information for classification.

Chapter 3

Literature Review

3.1 Introduction

Relevant feature identification has become an essential task to apply data mining algorithms effectively in real-world scenarios. Therefore, many feature selection methods have been proposed to obtain the relevant feature or feature subsets in the literature to achieve their objectives of classification and clustering. The amount of high-dimensional data that exists and is publically available on the internet has greatly increased in the past few years. Therefore, machine learning methods have difficulty in dealing with the large number of input features, which is posing an interesting challenge for researchers. In order to use machine learning methods effectively,pre-processing of the data is essential. Feature selection is one of the most frequent and important techniques in data pre-processing, and has become an indispensable component of the machine learning process. It is also known as variable selection, attribute selection, or variable subset selection in machine learning and statistics. It is the process of detecting relevant features and removing irrelevant, redundant, or noisy data. This process speeds up data mining algorithms, improves predictive accuracy, and increases comprehensibility. Irrelevant features are those that provide no useful information, and redundant features provide no more information than the currently selected features.

3.2 Feature Mining

Hyperspectral sensors record the reflectance from the Earth's surface over the full range of solar wavelengths with high spectral resolution. The resulting high-dimensional data contain rich information for a wide range of applications. However, for a specific application, not all the measurements are important and useful. The original feature space may not be the most effective space for representing the data. Feature mining, which includes feature generation, feature selection (FS), and feature extraction (FE), is a critical task for hyperspectral data classification. Significant research effort has focused on this issue since hyperspectral data became available in the late 1980s. The feature mining techniques which have been developed include supervised and unsupervised, parametric and nonparametric, linear and nonlinear methods, which all seek to identify the informative subspace.In the process of feature mining, irrelevant and redundant features or noise in the data may be hinder in many situations, because they are not relevant and important with respect to the class concept such as microarray data analysis . When the number of samples is much less than the features, then machine learning gets particularly difficult, because the search space will be sparsely populated. Therefore, the model will not able to differentiate accurately between noise and relevant data . There are two major approaches to feature selection. The first is Individual Evaluation, and the second is Subset Evaluation. Ranking of the features is known as Individual Evaluation . In Individual Evaluation, the weight of an individual feature is assigned according to its degree of relevance. In Subset Evaluation, candidate feature subsets are constructed using search strategy. The general procedure for feature selection has four key steps as shown in Figure 3.1

- Subset Generation
- Evaluation of Subset
- Stopping Criteria
- Result Validation

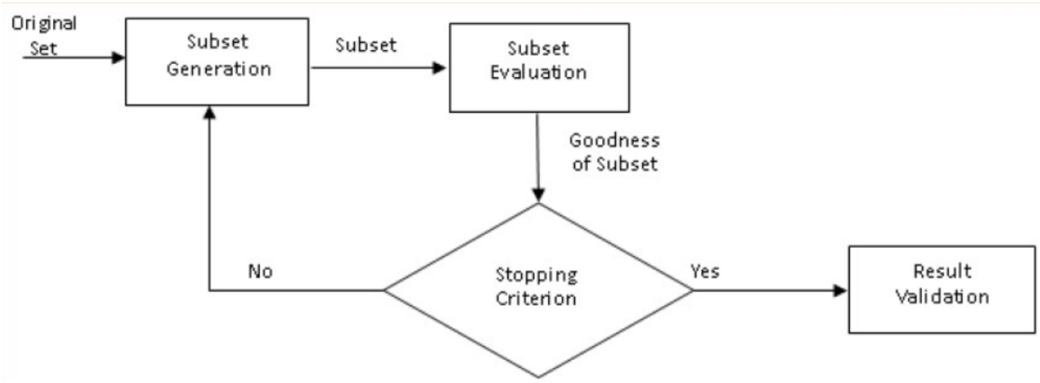

Figure 3.1: Four key steps for the feature selection process

Subset generation is a heuristic search in which each state specifies a candidate subset for evaluation in the search space. Two basic issues determine the nature of the subset generation process. First, successor generation decides the search starting point, which influences the search direction. To decide the search starting points at each state, forward, backward, compound, weighting, and random methods may be considered . Second, search organization is responsible for the feature selection process with a specific strategy, such as sequential search, exponential search or random search . A newly generated subset must be evaluated by a certain evaluation criteria. Therefore, many evaluation criteria have been proposed in the literature to determine the goodness of the candidate subset of the features. Base on their dependency on mining algorithms, evaluation criteria can be categorized into groups: independent and dependent criteria . Independent criteria exploit the essential characteristics of the training data without involving any mining algorithms to evaluate the goodness of a feature set or feature. And dependent criteria involve predetermined mining algorithms for feature selection to select features based on the performance of the mining algorithm applied to the selected subset of features. Finally, to stop the selection process, stop criteria must be determined. Feature selection process stops at validation procedure. It is not the part of feature selection process, but feature selection method must be validate by carrying out different tests and comparisons with previously established results or comparison with the results of competing methods using artificial datasets, real world datasets, or both. The relationship between the inductive learning method and feature selection algorithm infers a model. There are three general approaches for feature selection. First, the Filter Approach exploits the general characteristics of training data with independent of the mining algorithm [6].

Second, the Wrapper Approach explores the relationship between relevance and optimal feature subset selection. It searches for an optimal feature subset adapted to the specific mining algorithm [12]. And third, the Embedded Approach is done with a specific learning algorithm that performs feature selection in the process of training.

3.3 Related Works

Many feature selection methods have been proposed in the literature, and their comparative study is a very difficult task. Without knowing the relevant features in advance of the real data set, it is very difficult to find out the effectiveness of the feature selection methods, because data sets may include many challenges such as the huge number of irrelevant and redundant features, noisy data, and high dimensionality in term of features or samples. Therefore, the performance of the feature selection method relies on the performance of the learning method. There are many performance measures mentioned in the literature such as accuracy, computer resources, ratio of feature selection, etc. Most researchers agree that there is no so-called "best method". Therefore, the new feature selection methods are constantly increasing to tackle the specific problem (as mentioned above) with different strategy.

- To ensure a better behavior of feature selection using an ensemble method
- Combining with other techniques such as tree ensemble [86], and feature extraction
- Reinterpreting existing algorithms
- Creating a new method to deal with still-unresolved problems
- To combine several feature selection methods

Dimensionality reduction prior to classification is advantageous in hyperspectral data analysis because the dimensionality of the input space greatly affects the performance of many supervised classification methods [7]. Further, there is a high likelihood of redundancy in the features and it is possible that some features contain less discriminatory information than others. Moreover, the high-dimensionality imposes requirements for storage space and computational load. The analysis in [1] supports this line of reasoning and suggests that feature selection may be a valuable procedure in preprocessing hyperspectral data for classification by the widely used SVM classifier. In hyperspectral image analysis, feature selection is preferred over feature extraction for dimensionality reduction [1], [10]. Feature extraction methods involve transforming the data and hence, crucial and critical information may be compromised and distorted. In contrast, feature selection methods strive to discover a subset of features which capture the fundamental characteristics of the data, while possessing sufficient capacity to discriminate between classes. Hence, they have the advantage of preserving the relevant original information of the data. There are various studies which establish the usefulness of feature selection in hyperspectral data classification. [1] lists various feature selection methods for hyperspectral data such as the SVM Recursive Feature Elimination (SVM-RFE) [11], Correlation based Feature Selection(CFS) [12], Minimum Redundancy Maximum Relevance(MRMR) [13] feature selection and Random Forests [14]. In [6], a band prioritization scheme based on Principal Component Analysis (PCA) and classification criterion is presented. Mutual information is a widely used quantity in various feature selection methods. In a general setting, features are ranked based on the mutual information between the spectral bands and the reference map(also known as the ground truth). In [7], mutual information is computed using the estimated reference map obtained by using available a priori knowledge about the spectral signature of frequently-encountered materials. Recently, Brown et al [15] have presented a framework for unifying many information based feature selection selection methods. Based on their results and suggestions we have chosen the set of feature selection methods that they recommend outperform others, in various situations, which are elaborated in the next section for the purposes of our analysis. In [8] a feature selection method based on minimization of a tight bound on the VC dimension is presented. This paper presents the first application of this novel method to hyperspectral data analysis.

3.4 Feature Relevancy and Redundancy

3.5 General Approach for Feature Selection

The feature selection methods are typically presented in three classes based on how they combine the selection algorithm and the model building.

1. **Filter method:** Filter type methods select variables regardless of the model. They are based only on general features like the correlation with the variable to predict. Filter methods suppress the least interesting variables. The other variables will be part of a classification or a regression model used to classify or to predict data. These methods are particularly effective in computation time and robust to overfitting.

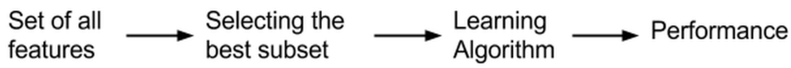

Figure 3.2: Filter Method for feature selection

However, filter methods tend to select redundant variables because they do not consider the relationships between variables. Therefore, they are mainly used as a pre-process method.

2. **Wrapper method:** Wrapper methods evaluate subsets of variables which allows, unlike filter approaches, to detect the possible interactions between variables. The two main disadvantages of these methods are :

   - The increasing overfitting risk when the number of observations is insufficient.
   - The significant computation time when the number of variables is large.

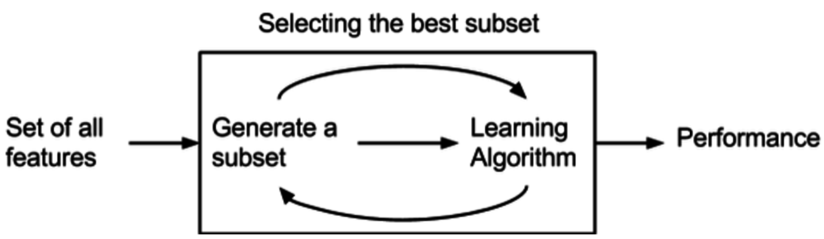

Figure 3.3: Wrapper Method for Feature selection

3. **Embedded method:** Embedded methods have been recently proposed that try to combine the advantages of both previous methods. A learning algorithm takes advantage of its own variable selection process and performs feature selection and classification simultaneously.

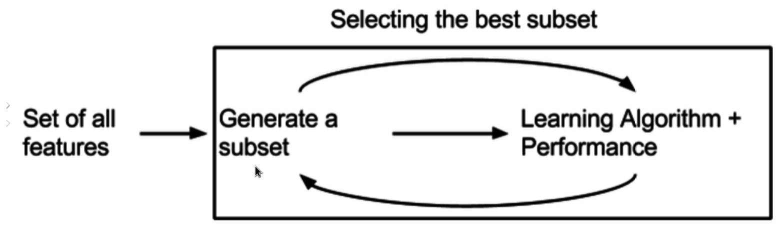

Figure 3.4: Embedde Method for Feature selection

3.6 Feature Mining for Hyperspectral Image

Hyperspectral sensors record the reflectance from the Earth's surface over the full range of solar wavelengths with high spectral resolution. The resulting high dimensional data contain rich information for a wide range of applications. However, for a specific application, not all the measurements are important and useful. The original feature space may not be the most effective space for representing the data. Feature mining, which includes feature generation, feature selection (FS), and feature extraction (FE), is a critical task

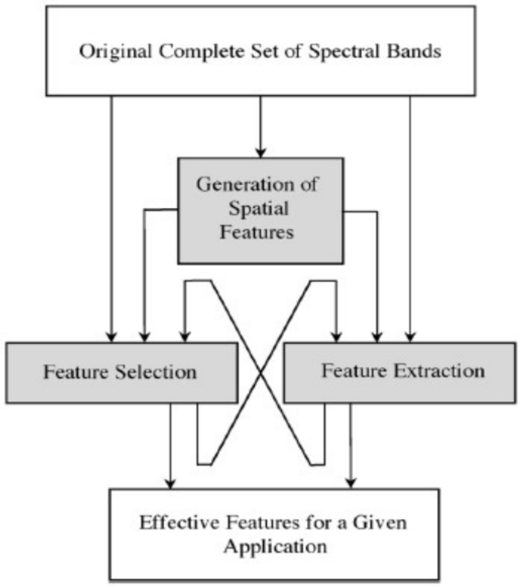

Figure 3.5: Possible operations for feature mining (only one path is used for any specific approach).

for hyperspectral data classification. Significant research effort has focused on this issue since hyperspectral data became available in the late 1980s. The feature mining techniques which have been developed include supervised and unsupervised, parametric and nonparametric, linear and nonlinear methods, which all seek to identify the informative subspace. This paper provides an overview of both conventional and advanced feature reduction methods, with details on a few techniques that are commonly used for analysis of hyperspectral data. A general form that represents several linear and nonlinear FE methods is also presented. Experiments using two widely available hyperspectral data sets are included to illustrate selected FS and FE methods.

3.7 Classification Techniques

Digital image classification techniques group pixels to represent land cover features. Land cover could be forest, urban, agricultural and other types of features. There are two main image classification techniques.

- Unsupervised Classification
- Supervised Classification

3.7.1 Unsupervised Classification

Pixels are grouped based on the reflectance properties of pixels. These groupings are called "clusters". The user identifies the number of clusters to generate and which bands to use. With this information, the image classification software generates clusters. There are different image clustering algorithms such as K-means and ISODATA. Unsupervised classification is a method which examines a large number of unknown pixels and divides into a number of classed based on natural groupings present in the image values. unlike supervised classification, unsupervised classification does not require analyst-specified training data. The basic premise is that values within a given cover type should be close together in the measurement space (i.e. have similar gray levels), whereas data in different classes should be comparatively well separated (i.e. have very different gray levels) (PCI, 1997; Lillesand and Kiefer, 1994; Eastman, 1995) The classes that result from unsupervised classification are spectral classed which based on natural groupings of the image values, the identity of the spectral class will not be initially known, must compare classified data to some from of reference data (such as larger scale imagery, maps, or site visits) to determine the identity and informational values of the spectral classes. Thus, in the supervised approach, to define useful information categories and then examine their spectral separability; in the unsupervised approach the computer determines spectrally separable class, and then define their information value. (PCI, 1997; Lillesand and Kiefer, 1994)

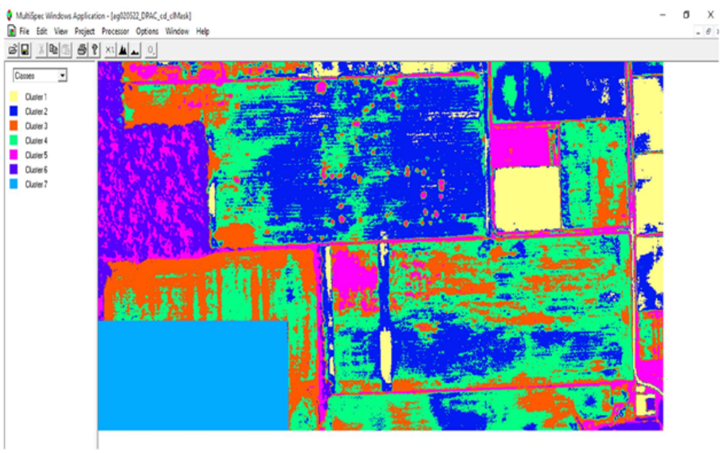

Figure 3.6: Unsupervised Classification

Unsupervised classification is becoming increasingly popular in agencies involved in long term GIS database maintenance. The reason is that there are now systems that use clustering procedures that are extremely fast and require little in the nature of operational parameters. Thus it is becoming possible to train GIS analysis with only a general familiarity with remote sensing to undertake classifications that meet typical map accuracy standards. With suitable ground truth accuracy assessment procedures, this tool can provide a remarkably rapid means of producing quality land cover data on a continuing basis.

3.7.2 Supervised Classification

The user selects representative samples for each land cover class in the digital image. These sample land cover classes are called "training sites". The image classification software uses the training sites to identify the land cover classes in the entire image. The classification of land cover is based on the spectral signature defined in the training set. The digital image classification software determines each class on what it resembles

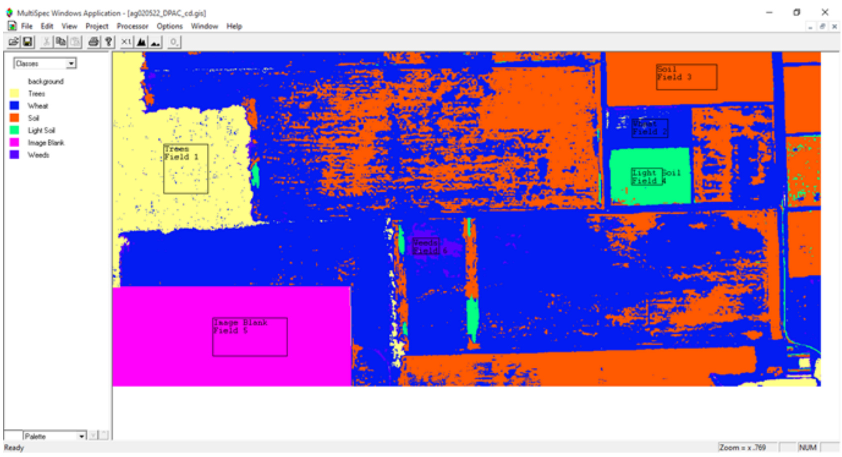

Figure 3.7: Supervised Classificationn

most in the training set. The common supervised classification algorithms are maximum likelihood and minimum-distance classification.

Chapter 4

Feature Selection and Classification

Feature mining generally refers to transform a high dimensional correlated dataset into a low dimensional uncorrelated data space. It can be done by following two steps:

- Feature extraction
- Feature selection

4.1 Feature Extraction Based on Principal Component Analysis

Principal component analysis is one of the most popular unsupervised feature extraction technique.The principal component analysis is based on the fact that neighboring bands of hyperspectral images are highly correlated and often convey almost the same information about the object. The PCA employs the statistic properties of hyperspectral bands to examine band dependency or correlation. This transformation is based on the mathematical principle known as eigenvalue decomposition of the covariance matrix of the hyperspectral image bands to be analyzed. The new transformed uncorrelated variables called principal components (PCs). First few variable contains most of the variation of the original data. If a hyperspectral image form M = i * j * k dataset where i is number of row, j is number of columns and k is the number of bands of the dataset. Then after applying PCA to our dataset, we can use a small portion of k and still find almost all variations of our data. if X1,X2....Xn is the dataset. Then for calculating PCA we first subtract mean form the dataset. The mean is calculated as

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{4.1}$$

The covariance matrix between any two dimension can be calculated as:

$$COV(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1} \tag{4.2}$$

Covariance matrix is a square matrix. Eigenvalue and eigenvector is calculated from covariance matrix. Eigenvector that has the highest eigenvalue is the first principal component. Arrangement of eigenvectors according to descending order of eigenvalues creates a matrix of new set of data whose first few column can be chosen for further use.

4.2 Feature Selection Based on Normalized Mutual Information

Mutual Information (MI) measures the amount of information one random variable holds about the other random variable by quantifying dependencies of those variables. Mutual information of two random varibles X and Y can be defined as:

$$I(X, Y) = \sum_{y=1}^{L1} \sum_{x=1}^{L2} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{4.3}$$

where x, y are the pixel values and L1, L2 are the maximum pixel values of the input images. P(x,y) is joint probability distribution function and P(x) and P(y) is marginal probability distribution function of variable X and Y respectively. If two variable x and y are independent, x does not know any information about y and vice versa. So their mutual information will be zero. According to Shannon's information theory, if the probability of a random variable x is P(x) then uncertainty of x measured by entropy is:

$$H(x) = -p(x)log(p(x)) \tag{4.4}$$

Here mutual information can also be described as:

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \tag{4.5}$$

Where H(A) and H(B) are the entropies of A and B and H(A,B) is their joint entropy. Now by calculating mutual information between class label $c = c_1, c_2, ...c_N$ and bands $f =$

$f_1, f_2, ...f_N$ of the hyperspectral image we can select relevant features for classification. Each feature will be selected in descending order $I_1 \geq I_2 \geq I_3 \geq .... \geq I_N$. Feature with highest mutual information contains most information. It is difficult to use the value given by (4.3) directly in an absolute sense, as it is affected by the entropy of the two variables and not bounded to [0, 1]. A few methods to normalize mutual information are available to give a new value between 0 and 1, where 0 is the value for the independent case and 1 for the identical or one-to-one relationship case. Normalized MI is defined as follows:

$$\hat{I}(X,Y) = \frac{I(X,Y)}{\sqrt{I(X,X)}\sqrt{I(Y,Y)}} \tag{4.6}$$

4.2.1 Original Dataset Plus nMI Approach

In feature selection, the relevant features have important information required to produce the classification output. The aim of feature selection is to find only informative features $k < n$ for a dataset. If the reference image is f and original dataset is x then the normalized mutual information between f and x is given below:

$$\hat{I}(x_i, f_j) = \frac{\sum_i \sum_j p(x_i, f_j) \log \frac{p(x_i, f_j)}{p(x_i)p(f_i)}}{\sqrt{I(x_i, x_i)}\sqrt{I(f_i, f_i)}} \tag{4.7}$$

4.2.2 PCA Dataset Plus nMI Approach

In order to find informative features from the transformed features obtained by unsupervised PCA, normalized mutual information (nMI) is implemented as below. If the reference image is f and PCA transformed dataset is y then the normalized mutual information between f and y is given below:

$$\hat{I}(y_i, f_j) = \frac{\sum_i \sum_j p(y_i, f_j) \log \frac{p(y_i, f_j)}{p(y_i)p(f_i)}}{\sqrt{I(y_i, y_i)}\sqrt{I(f_i, f_i)}} \tag{4.8}$$

4.3 Proposed Method

The proposed method is a target class oriented feature selection technique which select relevant features for each class of the dataset using principal component analysis (PCA) as feature extraction and normalized mutual information (nMI) as features selection. At first, the unsupervised feature extraction technique PCA is applied to find uncorrelated dataset in smaller data space. Then normalized mutual information (nMI) with two constraints to maximize general relevance and minimize redundancy on the components obtained via PCA is applied for feature selection. Here we consider one class of the dataset at a time and make all others as the background class. For each class of the dataset feature selection method is applied to obtain a relevant feature space to classify that class. Finally kernel support vector machine (SVM) is applied for classification purpose of the data set. Here first SVM use those selected features of each class to train a model for that specific class and test data to find test result for that specific class. RBF kernel is used in SVM as the relationship between class label C and datase X_i is nonlinear. K-fold cross validation is applied to find correct parameter C and γ of KSVM for each target class of the dataset. Finally average of those accuracy for each class is the final classification accuracy.

4.4 Implementation

4.4.1 Implementation of PCA Based Feature Extraction

4.4.2 Implementation of nMI Based Feature Selection

4.4.3 Implementation of Target Class Oriented Feature Selection

4.5 Experimental Results

4.5.1 Indian pines (92AV3C) Dataset

4.5.2 Feature Extraction Result

4.5.3 nMI Based Feature Selection Result

4.5.4 Target Class Oriented Feature Selection Result

4.6 Classification Results

Chapter 5

Conclusion and Future Works

5.1 Conclusion

5.2 Future Works

References

[1] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, p. 55–63, Jan. 1968.

[2] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, 4th ed. Germany: Springer-Verlag, 2006.