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 17, 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	**External Examiner**
__________________	__________________
Dr. Md. Ali Hossain	**Dr. X khan**
Assistant Professor	Assistant Professor
Department of Computer Science & Engineering	Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology	Rajshahi University of Engineering & Technology
Rajshahi-6204	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

List of Figures

List of Tables

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.

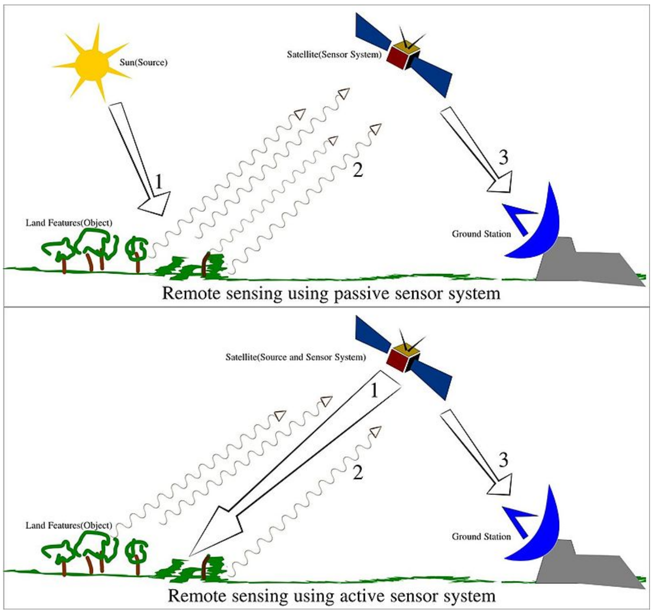

Figure 1.1: Remote Sensing

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.

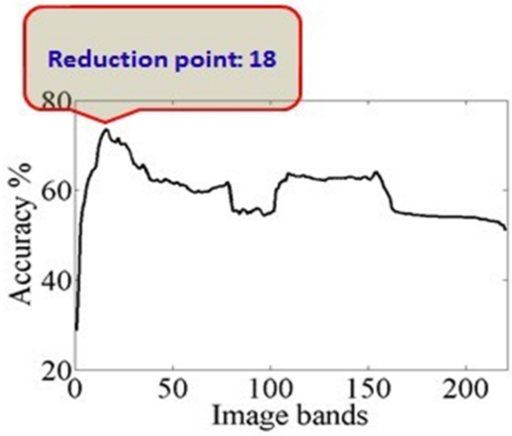

Figure 1.2: Curse of dimensionality

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. 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

2.2 Spectral Image Basics

2.3 Hyperspectral Data

2.4 Sensors

2.5 Application of Hyperspectral Image Analysis

2.6 Problems of Hyperspectral Data

Chapter 3

Literature Review

3.1 Introduction

3.2 Feature Mining

3.3 Related Works

3.4 Feature Relevancy and Redundancy

3.5 General Approach for Feature Selection

3.6 Feature Mining for Hyperspectral Image

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)

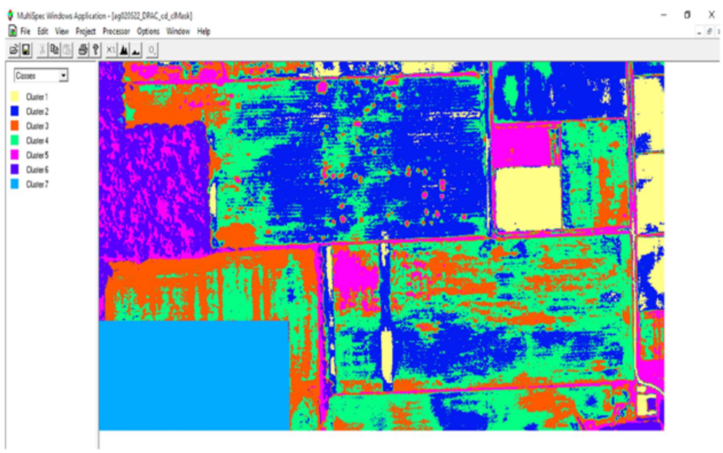

Figure 3.1: Unsupervised Classification

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

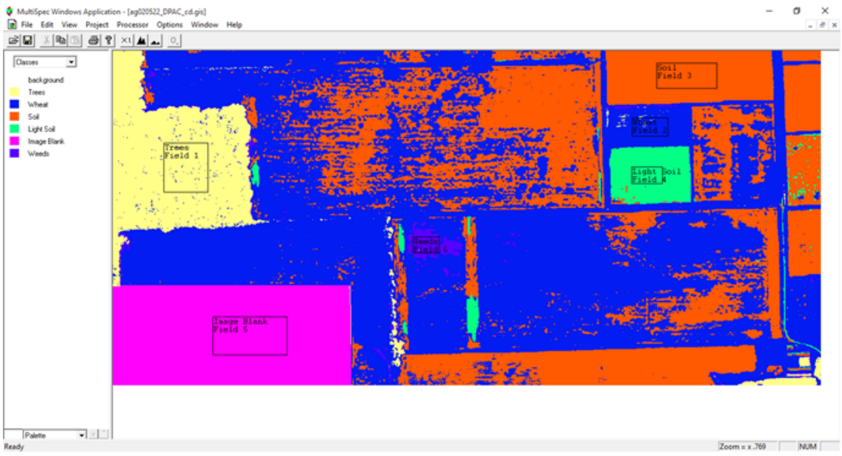

Figure 3.2: Supervised Classificationn

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