People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University M'Hamed BOUGARA – Boumerdes

Institute of Electrical and Electronic Engineering

Department of Electronics

Final Year Project Report Presented in Partial Fulfilment of the Requirements for the Degree of

MASTER

In **Telecommunication**

Option: **Telecommunications**

Title:

Spectral-spatial features for hyperspectral image classification

Presented by:

- **MOUNIR Zakaria**
- **MEROUANI Mawloud**

Supervisor:

Mr DAAMOUCHE

Registration Number:……/2018

Dedication

I dedicate this work

To my mother, my father and

To all my family

To all people who have helped me to obtain my

modest knowledge

To all good people that I met in Boumerdess and

shared the last five years together

To all my friends

ZAKARIA

Dedication

I dedicate this modest work

To my dear parents,

To the people who made me who I am today.

To all the good people that shared five years of their lives with me.

To all my friends.

Mouloud

Acknowledgement

First and foremost, we are thankful to God, the most merciful, the most beneficent at times when it seemed impossible to go on.

We would like to express our deep gratitude to our supervisor Mr. Abdelhamid Daamouche for his assistance, his wisdom orientations, ideas, effort and suggestions during the period of this work. it has been a privilege to work under his supervision.

Table of content

Chapter 3 : Data Classification.

Chapter 4: Data classification & Results

List of tables

List of Figures

List of abbreviations

SVM: Support Victor Machine.

AVIRIS: Airborne Visible InfraRed Imaging Spectrometer.

HSI: HyperSpectral Imaging

PCA: Principle Component Analysis.

MM: Mathematical Morphology.

MO: Morphological Operator.

MF: Morphological Filter.

SE: Structure Element.

C: Penalty Parameter.
g: Kernel Parameter.

Abstract

Image classification is one among important branches of artificial intelligence field. Generally, it translates the information contained in images into thematic categories which are suitable for use in many applications using low-level visual features.

Nowadays, there exists a large number of machine learning algorithms used for image classification.

The main objective of this work is to perform a classification of hyperspectral data by means of spectral-spatial features. The principle component analysis was exploited as a tool to decorrelate and reduce the dimension of the original hyperspectral data. The mathematical morphology is used to extract the spatial features; its parameters were generated empirically. The combination of the morphological features and the spectral features were fed to the state-of-the-art classifier which is the Support Vector Machines (SVM).

The obtained results over two benchmark datasets show that the achieved performance using the developed method is promising.

Introduction

Recently, immense research efforts have been devoted to developing advanced classification approaches and techniques to improve the hyperspectral image classification accuracy. However, classifying hyperspectral data into thematic information remains a challenge because of many factors, such as the complexity of the land shapes in a study area, and the difficulty to assign a label to a pixel vector such that it can be identified as belonging to a given class due to overlapping between the classes.

Classification techniques can be divided into unsupervised and supervised approaches, of which supervised classification methods are more widely used. However, the supervised classification of high-dimensional data sets, especially hyperspectral images, remains a challenging endeavor. The Hughes phenomenon caused by the large number of spectral bands poses a major problem during this process. Additionally, the presence of noise and that of mixed pixels, represent further hurdles hindering accurate hyperspectral image classification.

The focus of this project is on the classification of Hyperspectral data using principle component analysis as a technique to reduce the dimensions of the data to reduce computational time, with the tendency to preserve most of the spectral information then using morphological filtering in order to add spatial features to enhance the classification accuracy.

This report is organized in four chapters, chapter 1, which is an introductory chapter, deals with the fundamentals of the remote sensing data and some generalities about hyperspectral data and its applications. Chapter 2 explains the tools used in this work which are the Principle Component Analysis (PCA) and the Morphological operators. Chapter 3 highlights generalities about classification and describes the proposed classification approach. Chapter 4 presents the experimental results of the work and discusses the obtained results.

Chapter 1

Hyperspectral data

1. Hyperspectral data

1.1. Introduction to hyperspectral data

Since the beginning of remote sensing observation, scientists have created a "toolbox" with which to observe the varying dimensions of the Earth's dynamic surface.
Hyperspectral imaging represents one of the later additions to this toolbox, emerging from the fields of aerial photography, ground spectroscopy and multi-spectral imaging. This new tool provides capacity to characterize and quantify, in considerable detail, the Earth's diverse environments.

The concept of hyperspectral imaging originated from geological applications in the early 1980s, mainly for the purpose of mineral exploration [1]. Imaging spectrometers acquire data in many contiguous narrow bands leading to a complete reflectance or emittance spectrum for each pixel of the wavelength region covered.

1.2. Why hyperspectral?

Hyperspectral images are spectrally overdetermined; they provide ample spectral information to identify and distinguish between spectrally similar but unique materials. Consequently, hyperspectral imagery provides the potential for more accurate and detailed information extraction than it is possible with other types of remotely sensed data.

1.3. Basic concept

Hyperspectral imaging, or imaging spectroscopy, combines the power of digital imaging and spectroscopy. For each pixel in an image, a hyperspectral camera acquires the light intensity (radiance) for a large number (typically a few tens to several hundred) of contiguous spectral bands. Every pixel in the image thus contains a continuous spectrum (in radiance or reflectance) and can be used to characterize the objects in the scene with great precision and detail [2].

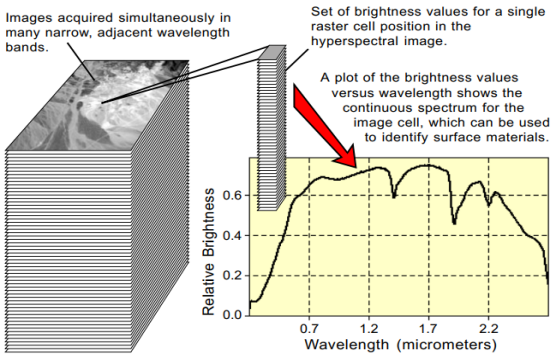


Figure 1-1 The concept of hyperspectral imagery.

Hyperspectral images obviously provide much more detailed information about the scene than a normal color camera, which only acquires three different spectral channels corresponding to the visual primary colors red, green and blue. Hence, hyperspectral imaging leads to a vastly improved ability to classify the objects in the scene based on their spectral properties.

Recent advances in sensor design and processing speed has cleared the path for a wide range of applications employing hyperspectral imaging, ranging from satellite based/airborne remote sensing and military target detection to industrial quality control and lab applications in medicine and biophysics. Due to the rich information content in hyperspectral images, they are uniquely well suited for automated image processing, whether it is for online industrial monitoring or for remote sensing.

1.4. Hyperspectral data acquisition

There are four basic techniques for acquiring the three-dimensional (x,y,λ) dataset of a hyperspectral cube. The choice of technique depends on the specific application, seeing that each technique has context-dependent advantages and disadvantages.

1.4.1. Spatial scanning

In spatial scanning, each two-dimensional (2-D) sensor output represents a full slit spectrum (x,λ). Hyperspectral imaging (HSI) devices for spatial scanning obtain slit spectra by projecting a strip of the scene onto a slit and dispersing the slit image with a prism or a grating.

1.4.2. Spectral scanning

In spectral scanning, each 2-D sensor output represents a monochromatic ('single-colored'), spatial map of the scene. HSI devices for spectral scanning are typically based on optical band-pass filters. The scene is spectrally scanned by exchanging one filter after another while the platform must be stationary

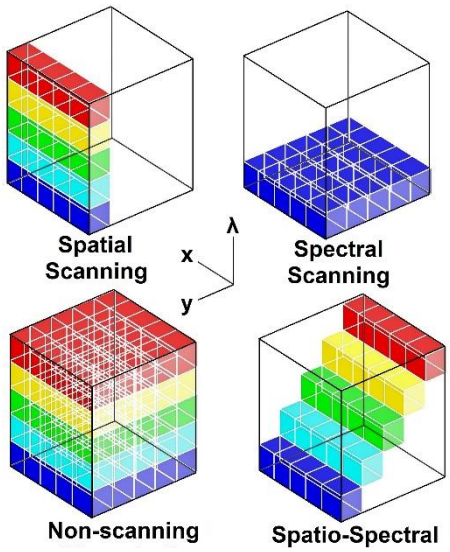


Figure 1-2 HSI data cube with its spatial dimensions (x,y) and spectral dimension

1.4.3. Non-scanning

In non-scanning, a single 2-D sensor output contains all spatial (x,y) and spectral (λ) data. HSI devices for non-scanning yield the full data cube at once, without any scanning. Figuratively speaking, a single snapshot represents a perspective projection of the data cube, from which its three-dimensional structure can be reconstructed.

1.4.4. Spatiospectral scanning

In spatiospectral scanning, each 2-D sensor output represents a wavelength-coded ('rainbow-colored', λ = λ(y)), spatial (x,y) map of the scene.

1.5. Hyperspectral V.s Multispectral Data

Multispectral and hyperspectral imagery gives the power to see as humans (red, green and blue), goldfish (infrared) and bumble bees (ultraviolet). Actually, we can see even more than this as reflected Electromagnetic radiation to the sensor.

The main difference between multispectral and hyperspectral is the number of bands and how narrow the bands are:

Multispectral imagery generally refers to 3 to 10 bands covering the spectrum from the visible to the longwave infrared. To be clear, each band is obtained using a remote sensing radiometer. Multispectral imaging with several images is what distinguishes multispectral imaging in the visible wavelength from color photography. A multispectral sensor may have many bands Multispectral images do not produce the "spectrum" of an object.

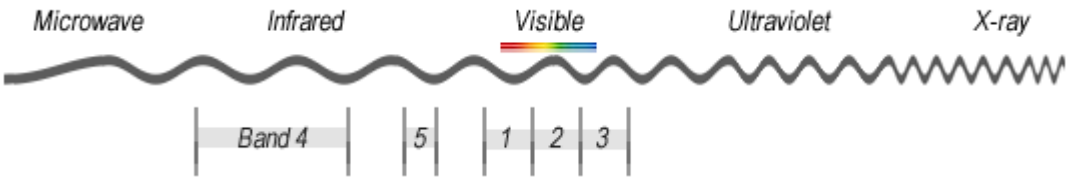


Figure 1-3 Multispectral Example: 5 wide bands.

Hyperspectral imagery consists of much narrower bands (10-20 nm). A hyperspectral image could have hundreds or thousands of bands. In general, it comes from an imaging spectrometer.

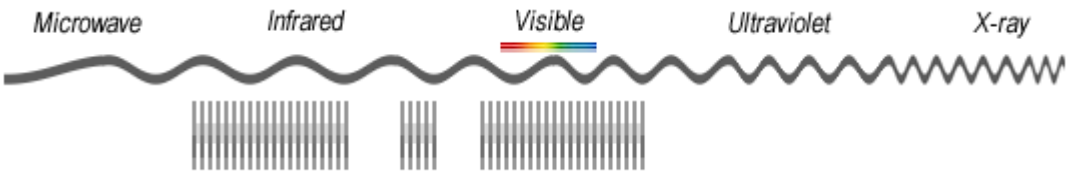


Figure 1-4 Hyperspectral Example: hundreds of narrow bands.

1.6. Hyperspectral data analysis

The traditional approach is centered upon the concept of delineating scene space areas (objects, or part object) that are as homogeneous and/or spatially coherent as possible, in order to define representative training classes. But it was needed to discriminate between severe inconsistencies which lead to suboptimal class representation [2][3].

Sensor	Organization	Country	Number Of band	Wavelength range(µm)
Satellite Sensors				
FTHSI on Mighty Sat II	Air Force Research Lab	United States	256	0.35 – 1.05
Hyperion on EO-1	NASA Goddard Space Flight Center	United States	220	0.4 – 2.5
Airborne Sensors				
AVIRIS (Airborne Visible Infrared Imaging Spectrometer)	NASA	United States	224	0.4 - 2.5
HYDICE (Hyperspectral Digital Imagery Collection Experiment)	Naval Research Lab	United States	210	0.4 - 2.5
AISA	Spectral Imaging Ltd.	Finland	286	0.45 - 0.9
CASI	Itres Research	Canada	288	0.43 - 0.87
DAIS 2115	GER Corp.	United States	211	0.4 - 12.0
HYMAP	Integrated Spectronics Pty Ltd	Australia	100 to 200	Visible to thermal infrared
PROBE-1	Earth Search Sciences Inc.	United States	128	0.4 - 2.45
EPS-H (Environmental Protection System)	GER Corp.	United States	VIS/NIR (76), SWIRI (32), SWIRI-2 (32), TIR (12)	VIS/NIR (0.43 to 1.05) SWIR-I (1.5 to 1.8), SWIR-II (2.0 to 2.5) and TIR

Table 1-1 Current and Recent Hyperspectral Sensors and Data Providers.

Table 1-1 lists some of the imaging spectrometers currently being operated for research or commercial purposes. The hyperspectral images produced by these sensors present a challenge for the analyst. They provide the fine spectral resolution needed to characterize the spectral properties of surface materials but the volume of data in a single scene can seem overwhelming. The difference in spectral information between two adjacent wavelength bands is typically very small and their grayscale images therefore appear nearly identical.

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. Hence, 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.

1.7. Advantages and disadvantages of HSI

The primary advantage to hyperspectral imaging is that, because an entire spectrum is acquired at each point, the operator holds all available information from the dataset to be mined. HSI can also take advantage of the spatial relationships among the different spectra in a neighborhood, allowing more elaborate spectral-spatial models for a more accurate segmentation and classification of the image.

The primary disadvantages are cost and complexity. Fast computers, sensitive detectors, and large data storage capacities are needed for analyzing hyperspectral data. Moreover, in high dimensionality, it comes difficult to perform accurate parameters estimation, for example in the Bayesian context, and the distance measures lose their efficiency to distinguish between different vectors .

1.8. Applications of Hyperspectral Image Analysis:

Having a higher level of spectral detail in hyperspectral images gives better capability to see the unseen. Hyperspectral imagery has been used to detect and map a wide variety of materials having characteristic reflectance spectra. For example:

- Hyperspectral images have been used by geologists for mineral mapping [3].
- To detect soil properties including moisture, organic content, and salinity [4].
- Vegetation scientists have successfully used hyperspectral imagery to identify vegetation species [3].
- Study plant canopy chemistry [4].
- To detect vegetation stress [5].
- Military personnel have used hyperspectral imagery to detect military vehicles under partial vegetation canopy, and many other military target detection objectives [6].

There are hundreds more applications where multispectral and hyperspectral enable us to understand the world. For example, we use it in the fields of agriculture, ecology, oil and gas, oceanography and atmospheric studies.

But one of the downfalls is that it adds a level of complexity. If you have 200 narrow bands to work with, how can you reduce redundancy between channels?

Chapter 2

Principle Component Analysis & Morphological Operators

2.1. Principle Component Analysis

2.1.1. Introduction principal component analysis

The use of hyperspectral images brings in new capabilities along with some difficulties in their processing and analysis. Unlike the widely used multispectral images, hyperspectral images can be used not only to distinguish different categories of land cover, but also the defining components of each land cover category, such as minerals, and soil and vegetation type. On the other hand, there are also difficulties in processing so many bands. The large amount of data involved with hyperspectral imagery will, however, dramatically increase processing complexity and time and reduce the classification accuracy this called the Hughes phenomena. Effectively reducing the amount of data involved or selecting the relevant bands associated with a particular application from the entire data set becomes a unique, yet primary task for hyperspectral image analysis.

In this work we use the principal component analysis (PCA) to reduce the dimension of the hyperspectral data and eliminate the Hughes phenomena to improve the accuracy of the classification.

2.1.2. Hughes phenomena

Hyperspectral data, consisting of hundreds of spectral bands with a high spectral resolution, enables acquisition of continuous spectral characteristic curves, and therefore have served as a powerful tool for classification. The difficulty of using hyperspectral data is that they are usually redundant, strongly correlated and subject to Hughes phenomenon.

Hughes phenomenon is a phenomenon that the classification precision increases gradually in the beginning as the number of spectral bands or dimensions increases, but when the band numbers reached at some point, the estimation accuracy begin to decrease dramatically. Due to Hughes phenomenon, the practical use of hyperspectral data is under restrictions. So, it is important to identify the reasons for Hughes phenomenon and find out the solution methods. In order to avoid “the curse of dimensionality” in classification, a better approach is to reduce the data dimensionality while trying to maintain the most vital and useful information in the dataset.

2.1.3. Dimension Reduction

Although HSI images provide abundant information about bands, in addition to the curse of dimensionality due to their high dimensionality they also substantially increase the computational burden. An important task in HSI data processing is to reduce the redundancy of the spectral and spatial information without losing any valuable details. Therefore, these conventional methods may require a pre-processing step, namely dimension reduction. Dimension reduction can be seen as a transformation from a high order dimension to a low order which eliminates data redundancy.

The proposed technique is Principal Component Analysis, PCA is one such data reduction technique, which is often used when analyzing remotely sensed data. The collected HSI image data are in the form of three dimensional image cube, with two spatial dimensions (horizontal and vertical) and one spectral dimension. In order to reduce the dimensionality and make it convenient for the subsequent processing steps, the easiest way is to reduce the dimensions by PCA.

2.1.4. Principal Component Analysis

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 analysis is used to transform the original data so to remove the correlation among the bands. In the process, the optimum linear combination of the original bands accounting for the variation of pixel values in an image is identified. The PCA employs the statistic properties of hyperspectral bands to examine band dependency or correlation. Though, one may find many synonyms for PCA, such as the Hotellling transformation or Karhunen-Loeve transformation [7], all these transformations are based on the same mathematical principle known as eigenvalue decomposition of the covariance matrix of the hyperspectral image bands to be analyzed.

Below is a brief formulation of the principle [8]:

$$x_i = [x_1, x_2, \dots \dots . x_N]_i^T \qquad (1.1)$$

with all pixel values x_1, x_2, ... x_N at one corresponding pixel location of the hyperspectral image data.

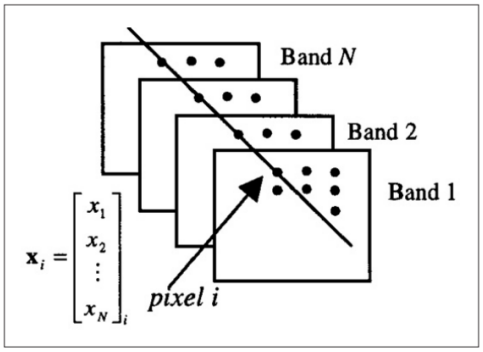


Figure 2.1-1 Pixel vector in principal component analysis.

The dimension of that image vector is equal to the number of hyperspectral bands N. For a hyperspectral image with m rows and n columns there will be M=m*n such vectors, namely i=1,…M. The mean vector of all image vectors is denoted and calculated as:

$$mean = \frac{1}{M}\sum_{i=1}^{M}[x_1, x_2, \dots \dots . x_N]_i^T \tag{1.2}$$

The covariance matrix of x is defined as:

$$Cov(x) = E\{\big(x - E(x)\big)\big(x - E(x)\big)^T\} \tag{1.3}$$

where:

E = expectation operator, **Cov** = notation for covariance matrix.

T superscript = transpose operation.

The covariance matrix is approximated via the following calculation:

$$C_x = \frac{1}{M}\sum_{i=1}^{M}(x_i - mean)(x_i - mean)^T \tag{1.4}$$

The PCA is based on the eigenvalue decomposition of the covariance matrix, which takes the form of:

$$C_x = DAD^T$$

where:

$$D=\begin{bmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_N\end{bmatrix}$$

D is the diagonal matrix composed of the eigenvalues $\lambda_1, \lambda_2 ... \lambda_N$ of the matrix C_X, and A is the orthonormal matrix composed of the corresponding N dimension eigenvectors a_k (k=1,2,…, N) of C_X as follows:

$$A = (a_1, a_2, \dots \dots, a_N) \tag{1.5}$$

The linear transformation defined by:

$$Y_i = A^T x_i \qquad (i = 1,2, \dots \dots, N) \tag{1.6}$$

Y_i is the PCA pixel vector, and all these pixel vectors form the PCA (transformed) bands of the original images.

Let the eigenvalues and eigenvectors be arranged in descending order so that $\lambda_1 > \lambda_2 > \dots > \lambda_N$, thus the first K (K < N, usually K<<N) rows of the matrix A^T, namely the first K eigenvectors $a_j^T \ (j = 1,2, \dots . K)$, can be used to calculate an approximation of the original images in the following way:

$$Z_i = \begin{bmatrix} z_1 \\ z_2 \\ . \\ . \\ z_K \end{bmatrix}_i = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1K} & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2K} & a_{2N} \\ . & .. & .. & .. & . \\ a_{K1} & a_{K2} & \dots & a_{KK} & a_{KN} \end{bmatrix} . \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \\ \vdots \\ x_N \end{bmatrix}_i \tag{1.7}$$

where pixel vector Z_i will form the first K bands of the PCA images. Such formed PCA bands have the highest contrast or variance in the first band and the lowest contrast or variance in the last band. Therefore, the first K PCA bands often contain the majority of information residing in the original hyperspectral images and can be used for more effective and accurate analyses because the number of image bands and the amount of image noise involved are reduced [9].

2.2. Morphological operatores

2.2.1. Introduction Morphological operators

Morphological operators or Mathematical Morphology (MM) has been in principle introduced as an image processing method, based on set theory [10]. The basic concept of morphological signal processing is to modify the shape of a signal, equivalently considered as a set, by transforming it through its interaction with another object, called the structuring element. In practice, the structuring element is compact and of a simpler shape than the original object.

2.2.2. Morphological Processing of sets

Mathematical filtering aims to quantitatively describe operations effective for the shape of objects in an image. Operations are described by combinations of a basic set of numerical manipulations between an image A and a small object B, called a structuring element, which can be seen as a probe that scans the image and modifies it according to some specified rule. The shape and size of B, typically much smaller than the image A, together with the specific rule, define the characteristics of the performed process.

Binary mathematical morphology is based on two basic operators: Dilation, and erosion. Both are defined in terms of the interaction of the original image A to be processed, and the structuring element B. Next the 4 basic operators are defined [11].

a. Morphological dilation

It is defined as the set union of the objects A obtained after the translation of the original image for each coordinate pixel b in the structuring element B:

$$A \oplus B = \bigcup_{b \in B} T_b(A) \tag{2.1}$$

Binary dilation can be interpreted as the combination of two sets by using the vector additions of set elements, called the Minkowski Addition. This operation is expressed as:

$$A \oplus B = \{ r | r = a + b \; \forall \in A \; and \; b \in B \} \tag{2.2}$$

b. Morphological erosion

Erosion is the morphological dual of the dilation. It is defined in terms of the Minkowski subtraction as:

$$A \ominus B = \{ r(|r + b)) \in A \forall b \in B \} \quad (2.3)$$

This definition can be expressed in terms of set intersections as:

$$A \ominus B = \cap_{b \in B} T_{-b}(A) \quad (2.4)$$

c. Opening filter

An important operator, which is the backbone of the pattern spectrum, is the *opening* morphological filter, defined as an erosion operation followed by a dilation using the same structuring element. The opening operator is defined as:

$$A \circ B = (A \ominus B) \oplus B \quad (2.5)$$

Dilation tries to undo erosion operation. However, some details closely related with the shape and size of the structuring element will vanish. Furthermore, an object disappearing as consequence of erosion cannot be recovered.

d. Closing filter

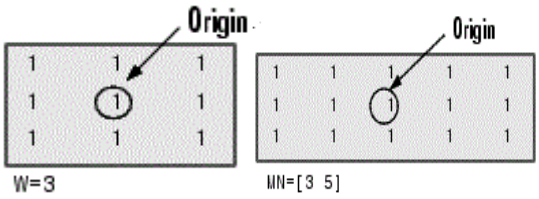
Closing morphological operator, which is the dual of the opening, is defined as dilation followed by an erosion operation using the same structuring element:

$$A \circ B = (A \oplus B) \ominus B \quad (2.6)$$

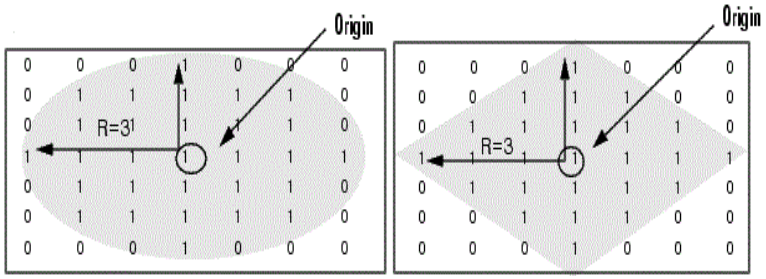
Opening and closing filters have been used as discriminators for filtering, segmentation, edge detection, differential counting, or numerical analysis of shapes.

2.2.3. Morphological Structuring Element

The structuring element is sometimes called the kernel, but we reserve that term for the similar objects used in convolutions. The structuring element consists of a pattern specified as the coordinates of a number of discrete points relative to some origin. Normally Cartesian coordinates are used and so a convenient way of representing the element is as a small image on a rectangular grid. **Figure 2.2-1** below shows a number of different structuring elements. In each case the origin is marked by a ring around that point.



Square 3*3 element Rectangular 3*5 element



Disk element with R=3 Diamond element with R=3

Figure 2.2-1 Some example structuring elements.

Note that each point in the structuring element may have a value. In the simplest structuring elements used with binary images for operations such as erosion, the elements only have one value, conveniently represented as a one. More complicated elements, such as those used with thinning or grayscale morphological operations, may have other pixel values.

Determining the size and shape of a structuring element is largely an empirical process. However, the overall selection of a structuring element depends upon the geometric shapes you are attempting to extract from the image data. For example, if you are dealing with biological or medical images, which contain few straight lines or sharp angles, a circular structuring element is an appropriate choice. When extracting shapes from geographic aerial images of a city, a square or rectangular element will allow you to extract angular features from the image.

The size of the structuring element depends upon what features you wish to extract from the image. Larger structuring elements preserve larger features while smaller elements preserve the finer details of image features.

When a morphological operation is carried out, the origin of the structuring element is typically translated to each pixel position in the image in turn, and then the points within the translated structuring element are compared with the underlying image pixel values. The details of this comparison, and the effect of the outcome depend on which morphological operator is being used.

Chapter 3
Classification

3. Classification

3.1. Introduction

Hyperspectral data are used in many applications. Typically, an image classification process is converting this data into meaningful information. Unfortunately, image classification is not a trivial task. Classification of HSI data is particularly daunting because most of the supervised learning schemes require sufficiently large amount of training samples, yet definition and acquisition of reference data is often a critical problem. Various classification techniques, both parametric and non-parametric, have been developed and used in different contexts.

3.2. Image classification

To classify an image is to assign each pixel in the image to a class. The spectral band values for each pixel in an image forms a cloud of points when plotted in multidimensional space. A classifier is a computer algorithm that takes the data cloud of points and groups the data into clusters or classes.

The assumption used by classifiers is that pixels of like materials will plot close together in the data cloud, and that the closest cluster to an individual data point will consist of other pixels of the same category. This is demonstrated schematically in **Figure 3-1**, which shows a data cloud in 2-D image space [12].

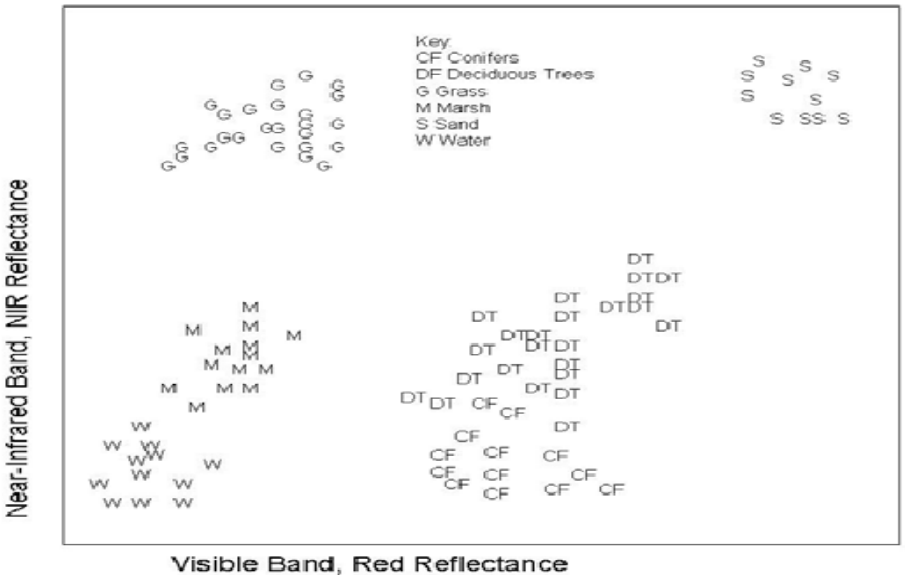


Figure 3-1 Image pixels grouped into classes in a 2D scatter plot.

3.3. Types of classifications

Image classification is perhaps the most important part of digital image analysis. It is very helpful to have a "pretty picture" or an image, showing a magnitude of colors illustrating various features of the underlying terrain, but it is quite useless unless to know what the colors mean. Two main classification methods are Supervised Classification and Unsupervised Classification.

3.3.1. Supervised classification

With supervised classification, we identify examples of the Information classes (i.e., land cover type) of interest in the image. These are called "training sites". The image processing software system is then used to develop a statistical characterization of the reflectance for each information class. This stage is often called "signature analysis" and may involve developing a characterization as simple as the mean or the rage of reflectance on each bands, or as complex as detailed analyses of the mean, variances and covariance over all bands.

Once a statistical characterization has been achieved for each information class, the image is then classified by examining the reflectance for each pixel and making a decision about which of the signatures it resembles most.

3.3.2. Unsupervised classification

Unsupervised classification is a method which examines a large number of unknown pixels and divides into a number of classes 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).

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 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.

3.4. Support Vector Machine (SVM)

SVMs (Support Vector Machines) are a useful technique for data classification. Although SVM is considered easier to use than Neural Networks, users not familiar with it often get unsatisfactory results at first. Although users do not need to understand the underlying theory behind SVM, we briefly introduce the basics necessary for explaining our procedure.

In general, there are two approaches to develop classifiers:

- A parametric approach, in which a priori knowledge of data distributions is assumed
- A nonparametric approach, in which no a priori knowledge is assumed.

Support vector machines (SVMs) is a supervised non-parametric statistical learning technique, therefore there is no assumption made on the underlying data distribution. In its original formulation [Vapnik, 1979] the method is presented with a set of labeled data instances and the SVM training algorithm aims to find a hyperplane that separates the dataset into a discrete predefined number of classes in a fashion consistent with the training examples. The term optimal separation hyperplane is used to refer to the decision boundary that minimizes misclassifications, obtained in the training step. Learning refers to the iterative process of finding a classifier with optimal decision boundary to separate the training patterns (in potentially high-dimensional space) and then to separate simulation data under the same configurations (dimensions) [Zhu and Blumberg, 2002].

In its simplest form, SVMs are linear binary classifiers that assign a given test sample a class from one of the two possible labels. An instance of a data sample to be labeled in the case of remote sensing classification is normally the individual pixel derived from the multi-spectral or hyperspectral image. Such a pixel is represented as a pattern vector, and for each image band, it consists of a set of numerical measurements. Elements of the feature vector may also include other discriminative variable measurements based on pixel spatial relationships such as texture.

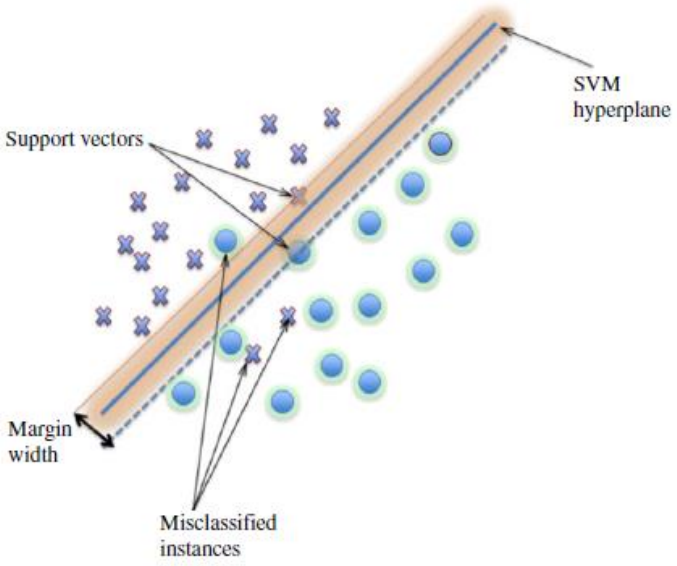


Figure 3-2 Linear support vector machine example.

Figure 3-2 illustrates a simple scenario of a two-class separable classification problem in a two-dimensional input space. An important generalization aspect of SVMs is that frequently not all the available training examples are used in the description and specification of the separating hyperplane. The subsets of points that lie on the margin (called support vectors) are the only ones that define the hyperplane of maximum margin [13][14].

3.4.1. How to use SVM in classification

The SVM package used in our project is LIBSVM, developed by **Chang and Lin** (2007). LIBSVM is an easy-to-use freeware module for the MATLAB environment. In preparing the inputs for LIBSVM, one can use the *read-sparse* command to turn raw data into a LIBSVM recognized format. For practical usage, LIBSVM mainly provides two simple commands: *Svm-train* is used to train an SVM, and *Svm-predict* are used for class predictions. Both commands require two files as input. One contains pixel attributes (an $m \times n$ matrix with m samples and n attributes), and the other is the pixel class (an $m \times 1$ matrix), in such a way that both training error and testing error can be automatically computed and displayed by LIBSVM. The user may also use the relevant parameter selection facilities provided by LIBSVM to test various SVM kernels and corresponding parameters. For further description see **Chang and Lin** (2007) [15].

3.4.2. SVM Format

In general, the hyperspectral data (train and test) is fed to the SVM classifier just after setting it in the SVM format as exerted by the LibSVM developers.

For the hyperspectral data to be recognized and processed by SVM classifier, it must be in the following format:

<label><index1>:<value1><index2>:<value2> .....

.

.

.

<label>

<label> is an integer indicating the class label.

The pair <index>:<value> gives a feature (attribute) value.<index> is an integer starting from 1, and <value> is a real number. Indices must be in ASCENDING order. Labels in the testing file are only used to calculate accuracy or errors.

3.4.3. Regularization parameter

We can define two main terms of regularization parameter **C** and **gamma.** These are tuning parameters in SVM classifier. Varying those we can achieve considerable nonlinear classification line with more accuracy in reasonable amount of time.

a) C parameter

C tells the SVM optimization how much you want to avoid misclassifying each training example.

For large values of C, the optimization will choose a smaller-margin hyperplane if that hyperplane does a better job of getting all the training points classified correctly. Conversely, a very small value of C will cause the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points.

b) Gamma parameter

The gamma parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’. In other words, with low gamma, points far away from plausible separation line are considered in calculation for the separation line. Whereas high gamma means the points close to plausible line are considered in calculation.

Chapter 4

Data classification

&

Results

4. Data classification & Results

4.1. Introduction data classification and result

In this chapter, we use the theoretical background obtained from the previous chapters to perform the classification of the hyperspectral data using spectral-spatial features with SVM classifier all this after doing dimension reduction using PCA. Then, we discuss the obtained results.

4.2. Data set description

The hyperspectral dataset used in our experiments is a section of scene gathered by AVIRIS sensor over the Indian Pines test site in North-western Indiana and consists of 145*145 pixels and 224 spectral reflectance bands in the wavelength range 0.4–2.5 10^{-6} meters. This scene is a subset of a larger one. The Indian Pines scene contains two-thirds agriculture, and one-third forest or other natural perennial vegetation. Since the scene is taken in June some of the crops present, corn, soybeans, are in early stages of growth with less than 5% coverage. The ground truth shown in **Figure 4-2** is designated into sixteen classes and is not all mutually exclusive.

There is ground truth for over 75% of this scene and it is comprised of the three row crops, Corn-notill, Soybean-notill, Soybean-mintill, and Grass-Trees. The full 145 by 145 scene for which there is ground truth covering 49% of the scene and it is divided among 16 classes ranging in size from 20 pixels to 2468 pixels. We have also reduced the number of bands to 200 by removing bands covering the region of water absorption: [104-108], [150-163], 220. Indian Pines data are available through Pursue's univeristy MultiSpec site [16][17].

Figure 4-1 Sample band of Indian Pines dataset.

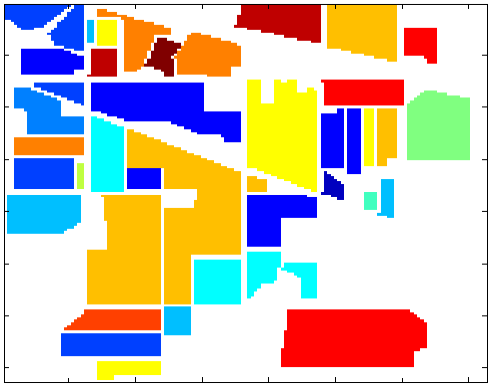


Figure 4-2 Ground truth of Indian Pines dataset.

Table 4-1 Ground truth classes for the Indian Pines scene and their respective samples number

#	Class	Samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

Table 4-1 Classes and number of samples used in the experiments.

4.3. Experiment setup

The work we propose is described by the following chart:

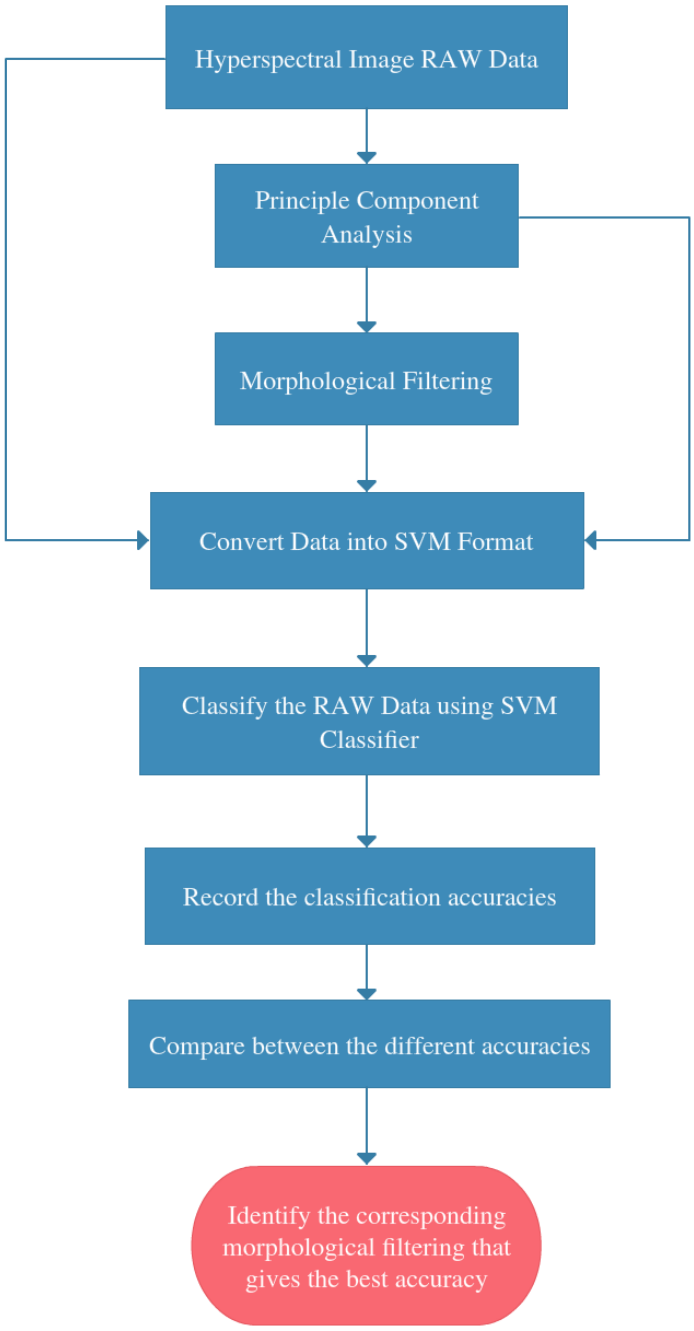


Figure 4-3 overall proposed method.

The block diagram shown in **Figure 4-3** illustrates how the procedures is done. This methodology adapted in the work is focused on the steps to achieve the objective i.e. performance classification of the hyperspectral data using PCA and Morphological filtering as feature extraction.

Working with high dimensional data like hyperspectral data cause the classifier to over fit the training data and consume a lot of computational power due to the high dimensional space in HSI [18].Dimension reduction is suggested and PCA is introduced as a feature extraction solution to this problem.

PCA transform hyperspectral image data into a new, uncorrelated co-ordinate system or vector space. It produces a space in which the data have maximum variance along its first axis, the next largest variance along a second mutually orthogonal axis and so on, thereby reducing the essential dimensionality of the classification space and thus improving classification speed. Stated differently, the purpose of this process is to compress all the information contained in an original n band data set into fewer than n components. The components are then used in lieu of the original data. These transformations may be applied as a preprocessing procedure prior to automated classification process of the data.

The main goal of using morphological filtering is to generate spatial features and remove the imperfection that mainly affects the shape and texture of our HIS. It is obvious that morphological operations can be very useful in image segmentation as the process directly deals with 'shape extraction' in an image. Morphology in context of image processing means description of shape and structure of the object in an image. It works on the basis of set theory and rely more on relative ordering of the pixel instead on their numerical value. This characteristic makes them more useful by providing more information for image processing.

In our work, we consider reducing the dimension of the data to minimum using the Principle Component Analysis. This will provide us with the optimum number of features vector without any redundancy in it and that would minimize the process time. In the Morphological Filtering part, we proposed to fuse information from two filtered data along with original data, this technique adds spatial features from morphologically filtered data to the existing spectral features in the original data to provide Spectral-spatial feature to improve the classification process, so the final data resulted before the classification will be a concatenation of two morphologically filtered data with the original one, the order of concatenation is not important.

The choice of the Morphological operator we use to filter the data affects the accuracy of the classification. The Morphological operator usage depends on the type, shape and size of the operator.

So, to experimentally to find the best Morphological operator for our application, our program was designed to apply different combinations using: opening, closing, erosion and dilation and different shapes such as disk, diamond, rectangle and square.

To be able to use apply PCA and MF, we wrote a MATLAB code that do a preprocessing on HSI data before feeding it to the classifier. It should be noted that different combinations of MO with different shapes are tested to give the best result since the hole process is empirical.

In general, the Matlab code is represented by the following chart:

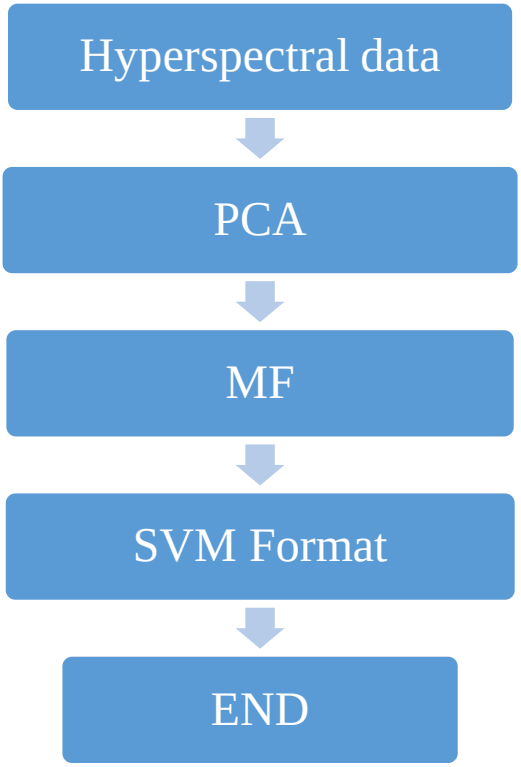


Figure 4-4 Flowchart of hyperspectral data classification based on feature extraction using Principal component analysis and Morphological filter

First, we took 100 training samples and 200 test samples from each one of our 16 classes in original HSI which contains 200 bands, in regards the classes that contains less than 100 samples we took half of the classes samples, this process was made to know the accuracy of classification before doing our preprocesses.

It should be noted that the classes to which the training and test data belong is already known and we processed the training and test data in the same manner.

We fed the training and the testing data to the classifier in the SVM format.

The training and test data are then stored in two separate .pat files. here is a sample of the training data shown in **Figure 4-5**.

```
1 1:2773 2:4119 3:4213 4:4188 5:4467 6:4847 7:4897 8:4831 9:4761 10:4569
1 1:2604 2:4241 3:4301 4:4183 5:4527 6:4805 7:4801 8:4707 9:4663 10:4525
1 1:3199 2:4111 3:4030 4:4022 5:4286 6:4633 7:4571 8:4552 9:4469 10:4274
1 1:3193 2:3994 3:4292 4:4183 5:4470 6:4752 7:4854 8:4668 9:4711 10:4476
1 1:2602 2:4116 3:4393 4:4183 5:4406 6:4703 7:4844 8:4662 9:4670 10:4464
1 1:2604 2:4241 3:4301 4:4183 5:4527 6:4805 7:4801 8:4707 9:4663 10:4525
1 1:2607 2:3985 3:4030 4:3957 5:4399 6:4649 7:4711 8:4653 9:4473 10:4311
1 1:2597 2:3871 3:4119 4:4117 5:4345 6:4704 7:4797 8:4623 9:4598 10:4425
1 1:2607 2:3985 3:4030 4:3957 5:4399 6:4649 7:4711 8:4653 9:4473 10:4311
1 1:2591 2:4114 3:4217 4:4179 5:4462 6:4810 7:4842 8:4753 9:4723 10:4566
1 1:3707 2:4239 3:4309 4:4107 5:4535 6:4753 7:4946 8:4781 9:4759 10:4565
1 1:2604 2:4120 3:4211 4:4111 5:4525 6:4708 7:4801 8:4673 9:4615 10:4531
1 1:3033 2:3986 3:4114 4:4103 5:4279 6:4649 7:4754 8:4568 9:4515 10:4338
1 1:2607 2:3995 3:4302 4:4112 5:4519 6:4808 7:4859 8:4763 9:4761 10:4581
1 1:3199 2:3994 3:4122 4:4032 5:4341 6:4542 7:4716 8:4579 9:4515 10:4237
1 1:3204 2:4232 3:4390 4:4037 5:4451 6:4752 7:4759 8:4658 9:4581 10:4378
1 1:2597 2:3869 3:4034 4:4031 5:4511 6:4747 7:4849 8:4589 9:4695 10:4531
```

Figure 4-5 Example of SVM Format

The resulting training data is used to find the optimum **c** and **g** and create a model that trains the SVM classifier to classify the test data, after adjusting the optimum parameters we get the flowing results in **Figure 4-6**.

```
Overall accuracy=77.5802% (2007/2587) (classification)

Accuracy for each class:
class 1:  OA=100.000000
class 2:  OA=44.000000
class 3:  OA=53.000000
class 4:  OA=94.000000
class 5:  OA=90.000000
class 6:  OA=86.500000
class 7:  OA=89.285714
class 8:  OA=99.500000
class 9:  OA=35.000000
class 10:  OA=65.500000
class 11:  OA=67.500000
class 12:  OA=76.500000
class 13:  OA=91.000000
class 14:  OA=78.500000
class 15:  OA=74.000000
class 16:  OA=95.698925


Average Accuracy=77.499040
```

Figure 4-6 Average and overall accuracies of the classification of the original data.

Then, the data is reduced into a smaller dimension. The reduced dimension depends on our selection we chose to take the 3 most weighted bands. So the Principle Component Analysis will reduce 200 dimension into 3, the resulted band shown in **Figure 4-7**.

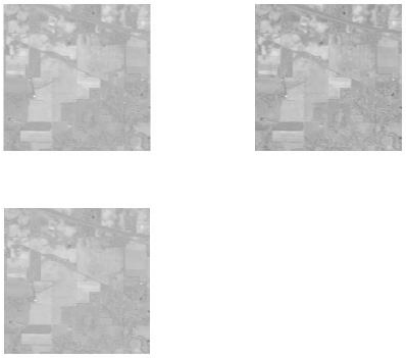


Figure 4-7 Resulted data from the PCA.

We redo the classification on the PCA resulted data as the same steps previously taking in consideration the selection of the best regularization parameter, we find the best c corresponding to the highest cross validation accuracy. In this case ***C***=200 and **gamma**=1.75, we assign them to the classifier as shown in the **Figure 4-8**.

```
svm-train -c 200 -g 1.75 traindata.pat model_cv

svm-predict testdata.pat  model_cv est_label.pat
pause
```

Figure 4-8 Content of the bat file of the classifier.

The classification result is shown in the **Figure 4-9.**

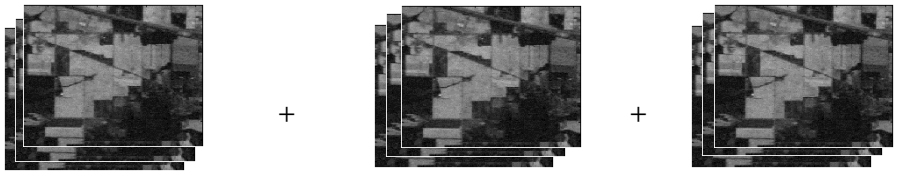
```
Overall accuracy=66.6796% (1725/2587) (classification)

Accuracy for each class:
class 1:  OA=71.739130
class 2:  OA=36.500000
class 3:  OA=52.500000
class 4:  OA=85.000000
class 5:  OA=81.000000
class 6:  OA=92.000000
class 7:  OA=67.857143
class 8:  OA=89.000000
class 9:  OA=0.000000
class 10:  OA=63.500000
class 11:  OA=53.500000
class 12:  OA=48.000000
class 13:  OA=91.500000
class 14:  OA=71.500000
class 15:  OA=28.500000
class 16:  OA=94.623656


Average Accuracy=64.169996
```

Figure 4-9 Average and overall accuracies of the classification PCA resulted data.

After finishing with dimension reduction we move now to morphological filtering since we are using concatenation we tried to find the best combination that gives us the optimum results by changing the operators, structure element, shape and size, as mention in previous chapter, even though selecting the parameters is empirical, however the overall selection depends on the geometric shapes and the size of the features that we are working on, in our work we proposed 4 shapes: disk, square, diamond and rectangular and we vary the size of the SE from 1 till 100.



PCA resulted data Morphological filtered -1-

PCA resulted data

PCA resulted data Morphological filtered -2-

Figure 4-10 Illustration of the concatenation of PCA resulted data with 2 PCA resulted data Morphological filtered

By varying the parameters, operator combinations and size different results are achieved and it will be shown in the following graphs depends on the SE shape:

a. Square SE

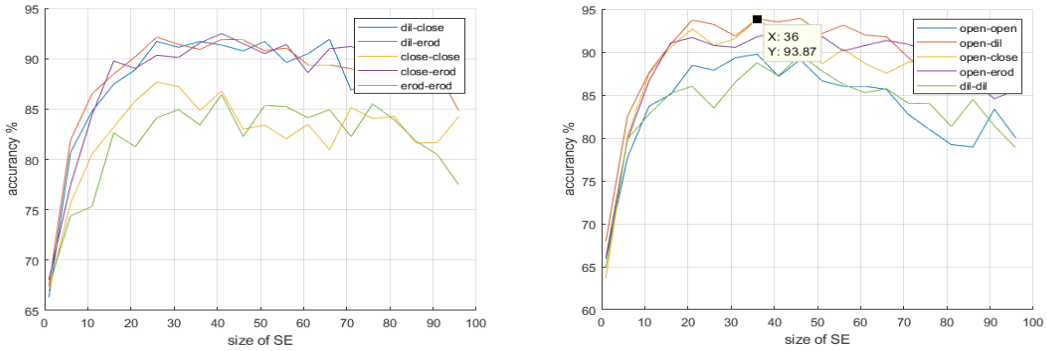


Figure 4-11 Graphical representation of the accuracy Vs size of the square SE for different combination of MO.

b. Rectangular SE

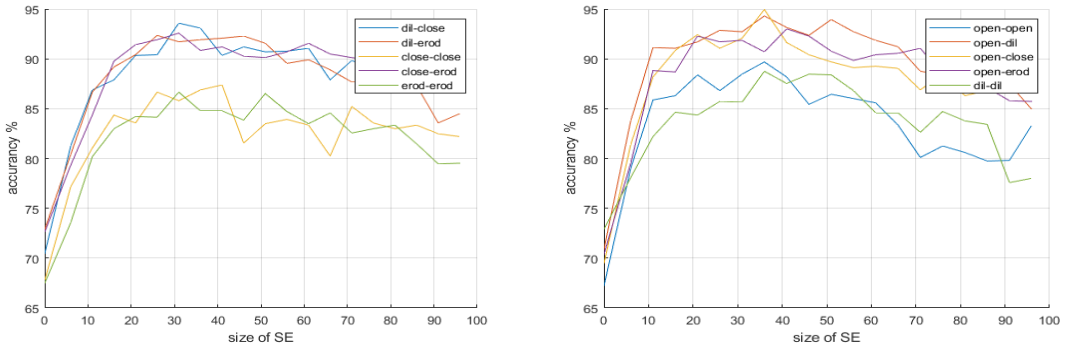


Figure 4-12 Graphical representation of the accuracy Vs size of the rectangle SE for different combination of MO.

c. Disk SE

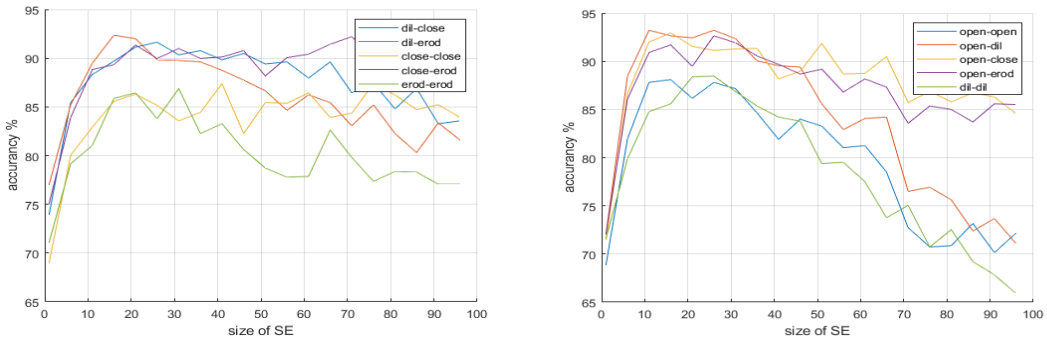


Figure 4-13 Graphical representation of the accuracy Vs size of the disk SE for different combination of MO.

d. Diamond SE

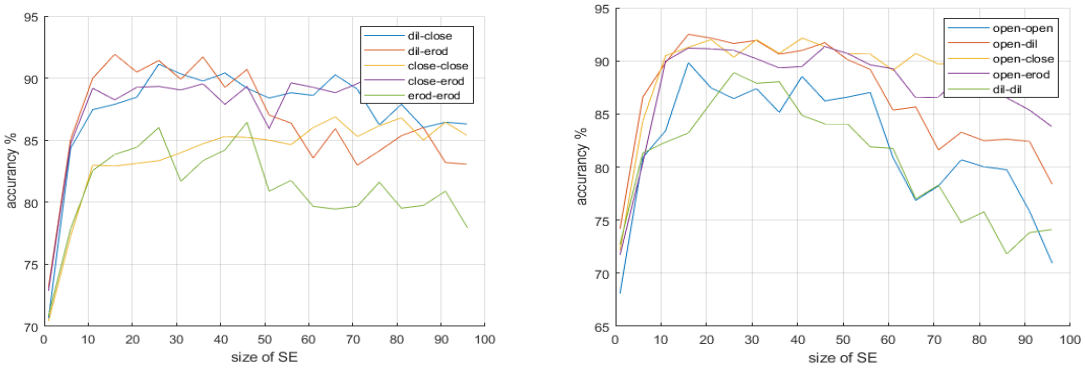


Figure 4-14 Graphical representation of the accuracy Vs size of the diamond SE for different combination of MO.

Regularization parameter used in the above graphs where varying between g=0.25 and 1.25 and select the best results, Where C was constant at 200 since it gives the optimum result.

4.4. Results

- Principle Component Analysis reduce the process time and eliminates the Hughes phenomena but it decreases the accuracy of the classification, the results decreased from 77.58% to 66.67%.
- Using Morphological filtering concatenation with the original data optimize the classification by adding spatial features to the existing spectral features, this technique called Spector-spatial features, it enhances the classification in some cases by more than 15% in our experiment compared to the original data classification.
- From the above graphs, we notice that the classification accuracy is proportional to many parameters: Morphological operator's combinations, structuring element shape and size. The best resulted accuracy is 94.74% with the 36*38 rectangle SE opening-closing operators with SVM parameters c=200 and g=0.75.
- We remark also that increasing the size of the SE further more decreases the classification accuracy.

In following the method used for the final classification by using the best resulted parameters: the rectangular 36*38 SE with the closing-opening filtering.

Best c =200, obtained from the following cross validation accuracy shown in **Figure 4-15**.

```
optimization finished, #iter = 129
nu = 0.006520
obj = -62.581222, rho = 0.218874
nSV = 10, nBSV = 0
Total nSV = 400
Cross Validation Accuracy = 89.9784%
```

Figure 4-15 Cross validation accuracy of the best **c.**

Best g =0.75, obtained from the cross validation shown in **Figure 4-16**.

```
optimization finished, #iter = 95
nu = 0.003614
obj = -34.696837, rho = 0.083661
nSV = 13, nBSV = 0
Total nSV = 405
Cross Validation Accuracy = 90.9877%
```

Figure 4-16 Cross validation accuracy of the best **g.**

The classification accuracies shown in **Figure 4-17**.

```
Overall accuracy=94.7429% (2451/2587) (classification)

Accuracy for each class:
class 1:  OA=100.000000
class 2:  OA=88.000000
class 3:  OA=90.000000
class 4:  OA=99.500000
class 5:  OA=97.500000
class 6:  OA=99.500000
class 7:  OA=96.428571
class 8:  OA=100.000000
class 9:  OA=90.000000
class 10:  OA=95.000000
class 11:  OA=80.000000
class 12:  OA=90.000000
class 13:  OA=100.000000
class 14:  OA=98.500000
class 15:  OA=97.000000
class 16:  OA=96.774194


Average Accuracy=94.887673
```

Figure 4-17 overall and average accuracies.

Now to classify our HSI data we transferred the data into SVM format after prepressed and feed it to the classifier, the classifier will use the model created during training to classify the vector pixels, reshaping the output of the SVM classifier results gives **Figure 4-18**.

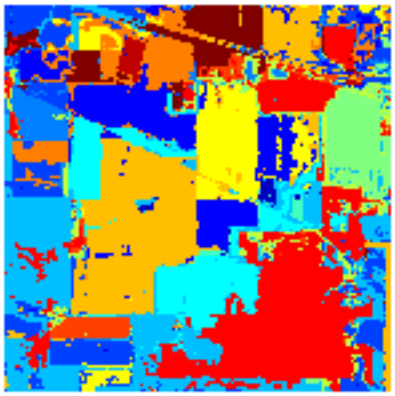


Figure 4-18 final classification result using the optimum parameters.

if we eliminate the non-classified areas

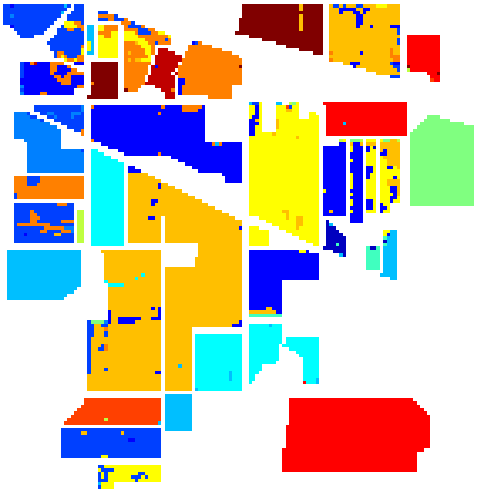


Figure 4-19 final classification result without non-classified areas.

Moreover, we implemented our model on the Pavia University data sets in order to check the robustness of the method.

The classification results shown in **Figure 4-20.**

```
Accuracy = 94.4444% (1700/1800) (classification)


Accuracy for each class:
class 1:  OA=83.500000
class 2:  OA=94.000000
class 3:  OA=93.500000
class 4:  OA=99.500000
class 5:  OA=99.500000
class 6:  OA=99.000000
class 7:  OA=96.500000
class 8:  OA=84.500000
class 9:  OA=100.000000
Average Accuracy=94.444444
Mean squared error = 1.22222 (regression)
Squared correlation coefficient = 0.822779 (regression)
```

Figure 4-20 classifications accuracies of Pavia university.

Reshaping the output of the classifier we get **Figure 4-21**.

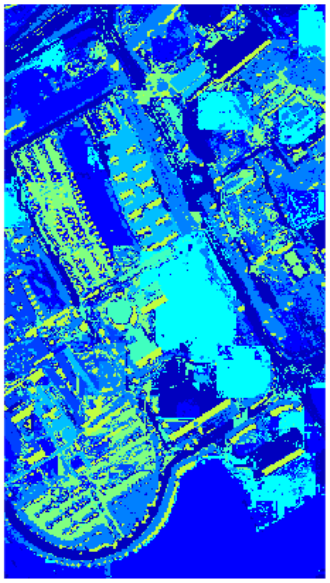
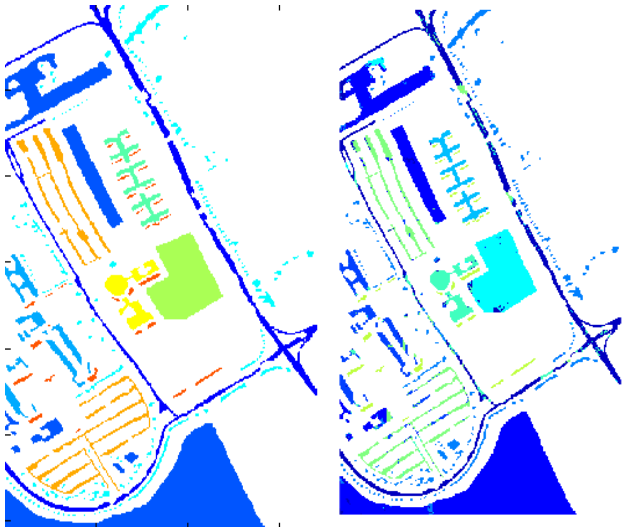


Figure 4-21 Classification of Pavia university using our proposed model



(a) **(b)**

Figure 4-22 a) Pavia university ground truth. **-b)** Classification of Pavia university without non-classified areas.

4.5. Discussion

In this work, we have performed feature extraction tending to preserve most of the information contained in the original data. Our goal was to reduce the processing time, we based our work on the assumption that the hyperspectral data suffers from Hughes phenomenon. However, the dimensionality reduction gives low accuracy results. The decrease in the classification accuracy is mainly due to the loss of information during the process. Add more feature such as spatial feature will minimize the loss of information and gives a higher accuracy, so the proposed technique is adding 2 spatial features victors to the PCA data victors to compensate the lost information, the morphological filtering was suggested as spatial features extraction technique, a different combination of two operators in the filtering was used with changing in the parameters of the operator to detect the best filtering parameter in our case.

From the obtained results, we can see that the change of the used morphological filtering parameter has a direct impact on the spatial features, which affects the classification accuracy due to the changing in the delivered information. We can notice that there exist good accuracies for the same data set with different spatial feature combinations. In other words, there may not be found a unique feature combination which gives the best results even for the same data set. However, we notice also that using combination of the same morphological operators in the spatial feature extraction results lower classification accuracies compared to the ones using different operators, its mainly due to the fact that using the same operators will generate duplicate information in the Spectral-spatial vector and this fact will cause a lower information content compared to other combinations. In other word a lower classification accuracy.

Moreover, we implemented our model on the Pavia University which has different specifications: high spatial resolution 610*430 pixel per band (Pavia University) versus low spatial resolution 145*145 pixel per band (Indian Pines), low spectral resolution 103 band (Pavia University) versus high spectral resolution 220 band (Indian Pines), and agricultural (Indian Pines) versus urban (Pavia University) which could help us to form a better judgement. The results were very promising where the accuracy reached 94.44%, this result indicates the robustness of the method and its reliability to preform high accuracy classification.

Conclusion

In this project, a method for classification of Hyperspectral Image is presented. This method presents dimension reduction by Principal Component Analysis. The reduced data classified then through the support vector machine (SVM) classifier using Spectral-spatial features by combining information from original and spatially filtered hyperspectral images. In particular, we exploited different combinations of parameters for the morphological filter as a spatial feature extraction technique to deliver a Spectral-spatial feature vectors as input to the support vector machine classifier, and then searched for the parameters that provide the best data representation in terms of classification accuracies.

Using the above technique overall accuracy of 94.74% was achieved with specific parameters which is very promising in comparison to original data Support Vector Machine classification which had an overall accuracy of 75.58% in our case, with the same dataset. Another Dataset which is Pavia university also tested and we got promising results.

At this point it may be clear to conclude that the classification of hyperspectral data based on the Spectral-spatial information gives better results with almost 15% comparing to classification based original data spectral information. However, considering a dimension reduction technique is very important when dealing with this type of data due to its large dimension that can overfit the classifier and take a lot of computation power.

Finally, through this modest work, we can say that we have succeeded in implementing Spectral-spatial features extraction technique as a model to classify the hyperspectral data.

References

[1]. Goetz A.F.H., Vane G., Solomon J., Rock B.N. “Imaging spectrometry for earth remote sensing”. Science. 1985;228:1147–1153.

[2]. Randall B. Smith, Ph.D. “Introduction to hyperspectral” with TNTmips. MicroImages, Inc. 5 January 2012.

[3]. Peg Shippert. Ph.D. “Introduction to Hyperspectral Image Analysis. Earth Science Applications Specialist. Research Systems, Inc”.

[4]. Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., and Yakhini, Z. “ Tissue Classification with Gene Expression Profiles”. J. Comput. Biol. 2000.

[5]. Merton, Robert C. "Commentary: Finance Theory and Future Trends: The Shift to Integration.”. July 1999.

[6]. Clark, *C.*, Dyson, A. and Millward, A. (Eds.). “Towards Inclusive Schooling**”.** London. (1995).

[7]. Rafael C. Gonzalez, Richard E. Woods “Digital Image Processing”.1993

[8]. Craig Rodarmel and Jie Shan, Principal Component Analysis for Hyperspectral Image Classification ,Surveying and Land Information Systems, Vol. 62, No. 2, 2002, pp.115-000.

[9]. Richard Eugene Woods, Rafael C. Gonzalez, Digital Image Processing 1993.

[10]. J. Serra 1982 Im*a*g*e* Analysis and Mathematical Morpholo*g*y New York: Academic Press.

[11]. Pilar Gómez-Gil, Manuel Ramírez-Cortés1, Jesús González-Bernal, Ángel García Pedrero1, César I. Prieto-Castro, Daniel Valencia, Rubén Lobato, José E. Alonso. A Feature Extraction Method Based on Morphological Operators for Automatic Classification of Leukocytes.2008

[12]. Dr. Susan L. Ustin; Mary Andrews; Margaret E. Andrew; Yen-Ben Cheng; Deanne DiPietro; John Kefauver; Shawn Kefauver; Nina Noujdina; Karen S.Olmstead; Carlos M. Ramirez; Edward T. Tom; Emma Underwood, ” Project SI-1143: Application of Hyperspectral Techniques to Monitoring and Management of Invasive Plant Species Infestation “,January 2008.

[13]. Giorgos Mountrakis, Jungho Im, Caesar Ogole, “Support vector machines in remote sensing”, copyright©

2010 International Society for Photogrammetry and Remote Sensing, Inc.

[14]. Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2 (2), 121–167.

[15]. Brandt Tso, Paul Mather, “Classification methods for remotely sensed data”, © 2009 by Taylor & Francis Group, LLC

[16]. AVIRIS NW Indiana’s Indian Pines 1992 Data Set. [Online]. Available:

ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.lan (original files); ftp.ecn. purdue.edu/biehl/PC_MultiSpec/

ThyFiles.zip (ground truth).

[17]. Grupo de Inteligencia Computacional de la Universidad del País Vasco (UPV/EHU) http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Pavia_University_scene.

[18]. Donoho DL. (2000). High-dimensional data analysis: The curses and blessings of dimensionality.