SPECTRAL-SPATIAL APPROACHES FOR HYPERSPECTRAL DATA CLASSIFICATION

SATHI RANI ROY
Bachelor of Science, Shahjalal University of Science and Technology, 2009

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Mathematics and Computer Science
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

SPECTRAL-SPATIAL APPROACHES FOR HYPERSPECTRAL DATA CLASSIFICATION

SATHI RANI ROY

Date of Defense: December 18, 2014

Dr. Howard Cheng Co-Supervisor	Associate Professor	Ph.D.
Dr. Karl Staenz Co-Supervisor	Professor	Ph.D.
Dr. Jackie Rice Committee Member	Associate Professor	Ph.D.
Dr. Nadia Rochdi Committee Member	Adjunct Professor	Ph.D.
Dr. Hadi Kharaghani Chair, Thesis Examination Committee	Professor	Ph.D.

Dedication

I dedicate this thesis to my brother, Sujit Kumar Roy.

Abstract

Classification of hyperspectral data is very challenging and mapping of land cover is one of its applications. Improving the classification accuracy and computation time of hyperspectral data were achieved incorporating contextual information in combination with spectral information for correcting classification errors along class boundaries and within class. In the proposed method, the original hyperspectral image was first classified using the Support Vector Machine (SVM) classifier, followed by the Markov Random Field (MRF) approach applied to the boundary areas and Unsupervised Extraction and Classification of Homogeneous Objects (UnECHO) classifier used for the interior parts of regions to produce the final classification map. In this study two agricultural (Hyperion and AVIRIS) and one urban (ROSIS) datasets were used. Investigations of the spectral and various contextual approaches including feature reduction show that the SVM-MRF method with grid search works best for all of the datasets. The highest overall accuracy of 97.35% was achieved for the urban dataset.

Acknowledgments

This study was carried out at the Department of Mathematics and Computer Science, the Department of Geography, and the Alberta Terrestrial Imaging Centre (ATIC) at the University of Lethbridge and was funded primarily by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University of Lethbridge. At first, I wish to express my profound gratitude to my co-supervisor, Dr. Howard Cheng for his continuous support, mentorship, insightful advice and guidance throughout my research work. I would also like to express my deep acknowledgement to my co-supervisor, Dr. Karl Staenz, for his help, expertise and encouragement. I would like to thank my committee members, Dr. Jackie Rice and Dr. Nadia Rochdi for their support and valuable suggestions. I would also like to thank to Dr. Jinkai Zhang, graduate students Ilya Parshakov and Hoimonti Rozario, my labmates Tom Arjannikov and Zahra Ghasemaghai for their moral support. I am very much grateful to my parents, Satya Ranjan Roy and Chanchala Rani Chowdhury, my parents-in-law, Subodh Chowdhury and Rupasree Chowdhury, my family members, and finally, my husband, Subir Chowdhury, for their endless support and encouragement.

Contents

3 Method 39

List of Tables

List of Figures

Chapter 1

Introduction

Hyperspectral remote sensing datasets are one of the major sources of land cover information, which can be analysed using digital image processing techniques [28, 43]. Remote sensing data are acquired by sensors on board aircraft, spacecraft or Unmanned Aerial Vehicle (UAV) platforms [30]. The spectral resolution of remote sensing data is defined as the width of wavelength interval and the number of bands. The number of different intervals, or bands, which can be detected has increased from panchromatic to multispectral, hyperspectral and ultraspectral due to the advancing sensor technology in remote sensing. For example, hyperspectral sensors can simultaneously acquire data with more than a hundred bands for a single pixel and can have spectral resolution as high as 0.67 nm [11]. Therefore, hyperspectral images are a rich source of information for characterizing objects and materials of the Earth surface. Classification and identification of objects is one of the important applications of hyperspectral data.

For classifying hyperspectral data, two main approaches are used: unsupervised and supervised [13]. Unsupervised techniques do not require any information about the data [28]. In supervised techniques, a training collection of pixels with known correct labels is used to compute parameters that are used for classification. Each pixel is labelled based on its spectral information independently of other pixels. The Support Vector Machine (SVM) method is a supervised classification technique based on statistical learning theory and has been shown to be a very effective method for the classification of hyperspectral data [19]. Most supervised classification methods including SVM use only spectral information for classifying images. However, spatial information has also been used for classification [28]. Some classification methods use contextual (i.e., both spectral and spatial) information [33]. In these cases, a spectral information based classification algorithm is applied at first. After that, spatial information from neighbouring pixels is combined with the spectral information to improve classification results.

Although the amount of information in hyperspectral datasets is large compared to other remote sensing datasets, sometimes it is difficult to classify these datasets because of the large number of bands. This is especially true when the number of training samples is small, and this phenomena is known as Hughes phenomenon [43]. Moreover, the computational cost is high for processing datasets with a large number of bands. Hence, band reduction of hyperspectral datasets is very important for improving both accuracy and computation time. There are two types of band reduction techniques: band extraction and band selection. In band extraction, the original hyperspectral dataset is transformed into a new smaller dataset that contains most of the data variance [28]. Band selection algorithms select a best subset of features by removing noisy and redundant bands [22].

1.1 Problem Formulation

Many of the spectral-spatial classification methods consider the initial classification result of a neighbourhood to classify a particular pixel. There are also spectral-spatial classification methods that work only on the boundary pixels of the spectral classification. In this case, misclassification is caused by pixels containing multiple classes near the borders of regions. In both of these cases, misclassified or incorrectly classified pixels may appear on the boundary as well as the interior parts of the regions. Misclassification may occur mainly because of noise or mixing of multiple types of objects. These two reasons may decrease classification accuracies. If we are able to correct both misclassified interior and boundary pixels, the classification accuracy should be higher.

Band extraction and selection are used in previous studies for improving the accuracy of mostly spectral information based classifier [43, 50]. The accuracy of contextual classifiers may improve after band extraction and selection.

1.2 Objective

The primary objective of this study is hyperspectral image classification by incorporating contextual (i.e., spectral and spatial) information in order to improve classification accuracy as well as processing time. At first a standard spectral classification algorithm is applied. Spatial methods are then investigated to correct errors of both interior pixels of the regions and boundary pixels that surround the regions in the result of spectral classification. Different band extraction and selection techniques are investigated to determine their effects on accuracy of both spectral and contextual classification.

1.3 Summary of the Proposed Approach

In this thesis, the SVM method as a spectral classifier is first applied on the original hyperspectral data. After this step, two methods are examined: use of the Unsupervised Extraction and Classification of Homogeneous Objects (UnECHO) classifier for reducing misclassified pixels inside the constructed class boundaries and the Markov Random Field (MRF) regularization process to improve classification of mixed spatial boundary pixels of the first classified image [29, 33]. The spectral and spatial information are combined by both of these methods in this study. If we use both of these methods for reducing errors in a region and boundary, the classification accuracy may be higher than using each method individually. For investigating the effect of band extraction and selection techniques on classification accuracies of these contextual methods, the Principal Component Analysis (PCA) and the Independent Component Analysis (ICA) approaches are evaluated for feature extraction, and the SVM Recursive Feature Elimination (SVM-RFE), minimum-Redundancy–Maximum-Relevance (mRMR) and Correlation based Feature Selection (CFS) methods are evaluated for feature selection in this thesis. These methods have been used in other studies on classification problems [26, 31, 43].

In the literature, many of the classification approaches based on contextual information have been tested for urban areas or man-made objects only [29]. In this study, the proposed methods are tested for the classification of both agricultural and urban hyperspectral data.

1.4 Research Findings

Two agricultural and one urban datasets are used in the thesis. One of the agricultural datasets is used to evaluate the proposed method including SVM, MRF, and UnECHO methods. The MRF method improved the accuracy of the SVM result. On the other hand, the UnECHO method failed to improve the accuracy of the SVM classification result and because of this the accuracy of the final result is not improved.

We also found that a variation of the MRF method applied only on the boundary pixels of the SVM classification showed good results both numerically and visually for all datasets. This variation also reduces the computation time for classification. However, the classification accuracy is not improved for any of the datasets when the UnECHO method is applied to the SVM classification result. Feature reduction techniques showed different results for different datasets. For example, PCA and ICA techniques improved the accuracy of the SVM classification for one agricultural dataset to almost 100% but for other datasets, these techniques could not provide better results. Finally, using the SVM method with appropriate training parameters can improve the accuracy significantly compared to using the default parameter values.

1.5 Overview of this Thesis

This thesis is organized as follows:

Chapter 2 gives some background information on remote sensing as well as an overview of some spectral classification methods, contextual classification methods, and feature reduction techniques.

Chapter 3 briefly describes the problem, proposed methodology and implementation of all the methods used in this thesis.

In Chapter 4, a description of one agricultural dataset, and all experiments and results are presented.

Chapter 5 presents all experiments and results of two other datasets.

Finally, in Chapter 6, we summarize our findings and discuss future work.

Chapter 2

Background

Remote sensing can be defined as gathering information about an object without any physical contact with it. Terrestrial remote sensing is commonly used to refer to the identification of features of the Earth's surface by detecting the characteristics of electromagnetic radiation that is received from the Earth system. The property of electromagnetic radiation, which is reflected or emitted from the objects, is used to sense the Earth's surface from space in order to improve natural resource management, land use, and the protection of the environment among others [30].

The most important source of electromagnetic radiation is the Sun. Solar radiation is either reflected from, absorbed by, or transmitted through objects when it interacts with a surface. The basic property that permits identification of the type of an object in remote sensing imagery is called spectral signature [30]. An object's spectral signature is its reflectance values in the various wavelengths that are covered by the sensor [6]. The idea is that each type of object has a unique signature by which it can be identified [30]. For example, spectral signatures of water, green vegetation and soil are illustrated in Figure 2.1.

Spectral resolution is the sensor's ability to define wavelength intervals where the finer spectral resolution represents narrower wavelength range for a band. Each band has its Relative Spectral Response Function (RSRF) that is characterized by the center wavelength and bandwidth at the Full-Width Half-Maximum (FWHM) [28]. In the pattern recognition literature, features can be defined as spectra of remotely sensed data [28]. If there are d bands, the feature space can be represented as a scatter plot of d-dimensional vectors whose components are the reflectance values in each band such that each vector represents a pixel in the scene.

In multispectral sensors, only a few bands are available and for this reason, the spec-

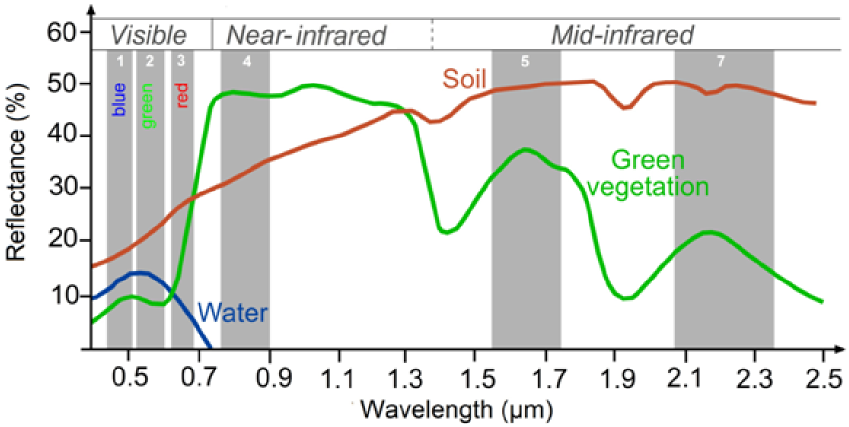


Figure 2.1: Spectral Signatures of water, green vegetation and soil [6].

tral resolution is low; for example, images obtained from Landsat-8 have 11 bands and QuickBird-2 have 4 bands [1, 42]. Some multispectral sensors have high spatial-resolution (area covered by a pixel). For example, each pixel of Quickbird-2 and SPOT-6 imagery represents an area of 2.4 m × 2.4 m and 6 m × 6 m on the ground, respectively [9, 42]. On the other hand, hyperspectral images are obtained using many narrow contiguous spectral bands. They contain higher resolution information than those of multispectral images. For example, images obtained from Hyperion have 220 spectral bands with 30 m spatial resolution [57]. Hyperspectral images can be used for separating land cover classes for mapping of the Earth's surface more accurately than multispectral images. The classification of hyperspectral datasets has several important applications, such as mapping of pipeline leakage through vegetation stress, finding sources of water pollution, and monitoring precision agriculture, vegetation health, and mine tailings site re-vegetation [53]. However, classification of this large amount of information from hyperspectral imagery is very challenging. Before performing classification of hyperspectral data, several pre-processing steps are mandatory. For example, removal of sensor artifacts and atmospheric effects that are described in the next Section [28]. The pre-processing steps are important, because the classification accuracy depends on the quality of data correction.

2.1 Image Preprocessing

In order to achieve the best classification accuracy, it is necessary to ensure that preprocessing steps are done properly. Major data pre-processing steps include removal of atmospheric and geometric effects [28]. In addition, assessment and removal of sensor artifacts and calibration effects are required. Data can be calibrated by converting Digital Numbers (DN) to radiance values using calibration coefficients, such as gain and offset in the metadata file provided by the data provider [46].

Reducing Sensor Artifacts There are a number of common types of sensor-specific artifacts that can be reduced.

1. *Instrument alignment* : Sometimes Visible Near-InfraRed (VNIR) and Short-Wave InfraRed (SWIR) images are not spatially and/or spectrally aligned properly, because of the misalignment of the VNIR and SWIR detectors. Shifting or rotating one of the images to the other one and spectrally shifting the VNIR or SWIR portion of the spectrum remove these artifacts [28].

2. *Random Noise* : The random influence of various sources, such as sensor electronics, by which the images are affected is called noise. Noise can be removed using smoothing kernels, convolution filters or by first transforming the data to another state, such as frequency space [28].

3. *Striping* : Striping is a systematic noise, which has a pattern to its distribution and variation [51]. Stripes occur along or across imagery due to improper radiometric

detector-to-detector calibration, temperature change, etc. The algorithms for destriping images can be grouped into two categories: digital filtering and statistical approaches (histogram matching and moment matching).

4. *Smile/Frown* : The position of wavelength changes across-track within an image in a specific band. This occurs because of spatial distortions caused by the diffusion element, such as prism or grating and the imaging optics [28]. Atmospheric absorption features of a radiative transfer model is used to determine the size of the smile/frown. The smile/frown is corrected by resampling image data based on the wavelength calibration data.

5. *Keystone* : Inter-band spatial misregistration in spectrographs means that the exact same area on the ground is not measured in all the bands due to distortion in camera lenses [28]. Spatial resampling is used to correct keystone [28].

Atmospheric Correction: Sensor radiance of a target can vary depending on the time of data acquisition and atmospheric properties. Scattering, transmission and absorption occur when radiation interacts with the atmosphere [28]. Water vapour and other gases and particles change the radiation reflected from the objects. Absolute atmospheric correction of converting the measured at-sensor radiance to surface reflectance is needed for obtaining accurate spectral characteristics [55]. Generally, two different approaches are used to carry out this correction. In the first approach, information collected on the ground is used for removing the effects of illumination, atmospheric scattering, and gaseous absorption [55]. The second approach makes use of atmospheric radiative transfer models to remove the effects of atmospheric scattering and gaseous absorption under varying illumination and viewing conditions [55]. Hyperspectral sensors have absorption feature bands, which can be used to improve the atmospheric correction [55].

Geometric Correction: Geometric correction is used to reduce or eliminate the distortion in acquired imagery so that individual elements or pixels are in their proper planimetric map location [28]. This is usually done by co-registering the image with another image or by adjusting image coordinates to a map reference system.

2.2 Image Classification

Image classification is an information extraction procedure in remote sensing that is used for mapping land-use and land-cover types. The two main types of classification are supervised and unsupervised classification.

Supervised classification requires selecting training areas of remotely sensed data with assigned labels [24, 28]. The training areas, which represent homogeneous examples of known land-cover types are identified by an analyst through a combination of ground data and personal experience. The remainder of the image is classified using the classification algorithm, which is trained based on these training areas.

In unsupervised classification, classes are produced automatically by grouping similar pixels into clusters based on their spectral characteristics [28]. In the next step, the user has to manually label the classes as one land-cover type or another. As a general rule, the larger the number of classes, the more difficult it is to assign meaningful class labels.

There are various kinds of statistical pattern recognition techniques that can extract land-cover information for both supervised and unsupervised classification [28]. Generally, we can divide them as follows:

Per-pixel classification: the whole image is processed pixel by pixel. The labelling of each pixel depends mainly on the spectral information of that pixel.

Spatial-information based classification: the spatial information of groups of pixels is used to classify the pixels. For example, the shape of objects may be used.

Contextual classification: per-pixel classification results of nearby pixels (called a neighbourhood) are examined to determine the classification of each pixel.

In this thesis, only per-pixel and contextual classification methods are examined.

2.3 Classification Accuracy

Image classification algorithms are often evaluated based on the accuracy of the classification results. In order to perform the evaluation, “ground-reference” data must be available for the areas corresponding to the pixels classified.

A number of different measures are used to evaluate the accuracy of a classification algorithm [28]. The overall classification accuracy is the percentage of correctly classified pixels:

$$\text{Overall Accuracy (\%)} = \frac{\text{number of correctly classified pixels}}{\text{total number of pixels}} \times 100. \tag{2.1}$$

Sometimes the classification results are more accurate for some class and not as accurate for others. In these cases, it is useful to also study the classification accuracy for individual classes. For each individual class y, we define user accuracy and producer accuracy as follows [28]:

$$\text{User Accuracy (\%)} = \frac{\text{number of pixels correctly assigned to } y}{\text{total number of pixels classified as } y} \times 100 \tag{2.2}$$

and

$$\text{Producer Accuracy (\%)} = \frac{\text{number of pixels assigned to } y}{\text{total number of pixels in } y} \times 100. \tag{2.3}$$

The user accuracy for a class *y* is the percentage of pixels classified as class *y* that are correct, while the producer accuracy for *y* is the percentage of pixels in class *y* that are classified correctly.

Two closely related types of errors for each class can be defined in terms of user accuracy and producer accuracy:

$$\text{Commission Error (\%)} = 100\% - \text{User Accuracy} \tag{2.4}$$

and

$$\text{Omission Error (\%)} = 100\% - \text{Producer Accuracy.} \tag{2.5}$$

Commission error for a class *y* represents the pixels that are incorrectly assigned to *y*, while omission error represents the pixels in class *y* that are assigned to some other class.

2.4 Per-pixel classification

Per-pixel classification is a process in which each pixel of the entire image is classified. Most image classification techniques are based on per-pixel classification. These techniques include ISODATA and K-means (unsupervised classification), parallelepiped, minimum distance to mean, Maximum Likelihood (ML, parametric classifier), Mahalanobis distance (parametric classifier), Support Vector Machine (SVM), decision tree, and neural network [28]. We will only describe the Support Vector Machine (SVM) and Maximum likelihood methods in details in this study. Previous literature [20] has shown that SVM can provide high classification accuracies compared to other techniques in many cases. In addition, many remote sensing data users prefer the ML classifier, because it is also competitive with other classifiers as shown in the recent literature [3].

2.4.1 Support Vector Machine

SVM was first developed to deal with only two classes. Later multiple-class problems were handled by the expanded SVM [28]. The goal of SVM is to determine a hyperplane (decision boundary) between classes based on the training data. In the case of linearly separable problems, the classes do not overlap in the feature space. However, classes can often overlap in linearly non-separable problems [28].

In the case of a two-class classification problem, the linear SVM algorithm considers a training set consisting of N vectors (X_i, $i = 1, \ldots, N$) from the d-dimensional feature space R^d [36]. Each vector X_i is associated with a class label or target $y_i \in \{-1, +1\}$. In the case of linearly separable problems, there exists a hyperplane separating the vectors of the two classes (-1 and $+1$). The hyperplane can be described by a normal vector w and a bias $b \in R$. That is, the hyperplane consists of all vectors $X \in R^d$ satisfying the equation

$$w \cdot X + b = 0. \tag{2.6}$$

For any vector $X \in R^d$, we can decide, which class X belongs to by computing a discriminating function $f(X)$ defined as

$$f(X) = w \cdot X + b. \tag{2.7}$$

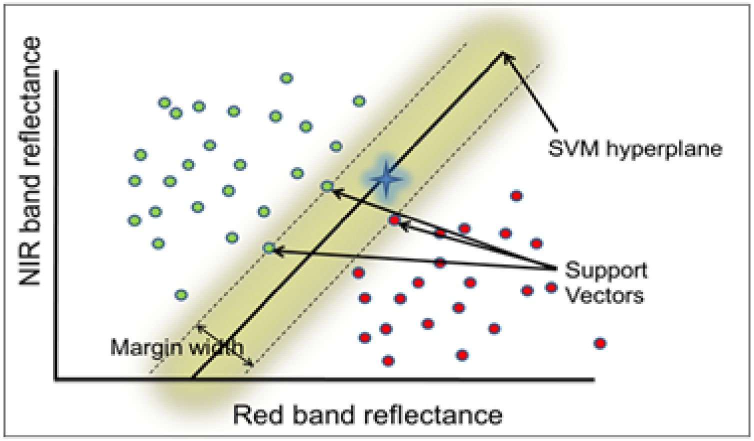


Figure 2.2: Concepts of Support Vector Machine (SVM) classification [28].

If $f(X) > 0$, then X is labeled as class $+1$. If $f(X) < 0$, then X is labeled as class -1. If $|f(X)|$ is large, then the labelling is likely correct. The value of $|f(X)|$ can be interpreted as a probability that the labelling is correct [49].

The optimal hyperplane acquired by the SVM approach maximizes the distance between the separating hyperplane and the closest training samples. The training samples that are closest to the hyperplane are called *support vectors.* The distance between these *support vectors* and the optimal separating hyperplane is the margin, and is equal to $1/\|w\|$ [36]. The concept of margin is essential in the SVM approach, because it is an indication of its generalization capability and measures how accurately the classifier labels new input different from the training samples. The larger the margin, the higher the expected generalization capability. These concepts are illustrated in Figure 2.2.

In the case of linearly non-separable data, the SVM approach allows some training samples to be incorrectly classified. It chooses a hyperplane to balance between maximizing the margin and minimizing incorrectly classified samples. This trade-off is controlled by parameters that are specified by the user or chosen by exhaustive search [36]. Separability between classes can also be improved by mapping feature vectors into a higher dimensional space using nonlinear kernel functions [28]. Kernel functions are used to embed data in a higher dimensional space and by adding extra dimensions, inseparable data may become separable. A kernel function, K, is defined as follows:

$$K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j), \tag{2.8}$$

where the optimal margin in the feature space can be written by replacing $X_i \cdot X_j$ with $\phi(X_i) \cdot \phi(X_j)$. Using the kernel function, (2.7) becomes

$$f(X) = \sum_i \lambda_i y_i K(X_i, X_j) + b, \tag{2.9}$$

where λ_i is a Lagrange multiplier. A detailed description of the computational aspects of SVM can be found in [10, 60].

Among different types of kernel functions, the Radial Basis Function (RBF) has been found to provide better classification results. Mathematically, the RBF kernel is defined as follows:

$$K(X_i, X_j) = \exp\left(-\gamma ||X_i - X_j||^2\right), \gamma > 0, \tag{2.10}$$

where γ is a kernel parameter. The default value of $\gamma = \frac{1}{d}$ is commonly used [10].

For multiclass problems, the SVM approach uses a number of two-class SVM classifiers to perform the classification [36]. There are two common strategies of using SVM in a multiclass problem:

1. *One-Against-All Strategy*: Each SVM solves a two-class problem defined by one class against all the others. If one of these SVMs labels an input vector as a particular class, this class label is given to the vector. If multiple labels are possible, the one

with the highest probability is chosen.

2. *One-Against-One Strategy*: There is a two-class SVM for each pair of classes. Each SVM assigns a label to an input vector, and the results of all SVMs can be combined based on the probabilities to obtain the final label.

The training data samples for SVM are considered as independent and identically distributed. However, in practice, the data is non-gaussian and erroneous, because of many reasons, such as motion (i.e., externally induced signals in the measurement) and sensor artifacts, shade and illumination effects, spectral variability, and other environmental effects [28, 52]. A posterior probability is established to express the uncertainty of labels assigned using the Bayesian decision theory—a basic method of statistical pattern recognition [61]. For a two-class problem, the distance of a feature vector from the separating hyperplane can be used to calculate the posterior probability [49], and this can be extended to the multi-class problem [62]. The formulas to compute these probabilities are similar to the ones used in the Maximum Likelihood Supervised Classifier described in the next section.

According to [19], the classification performance of SVM is very satisfactory using NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) hyperspectral data. In addition, in [44], SVM achieved better accuracy compared to other advanced approaches, such as neural network and maximum likelihood as shown in Figure 2.3.

2.4.2 Maximum Likelihood Supervised Classifier

One of the most popular supervised classifier in remote sensing is the maximum likelihood method [28]. It was described by Swain and Davis in 1978 [45] and has been used since then. The labeling decision is taken based on the likelihood that a pixel belongs to a par-

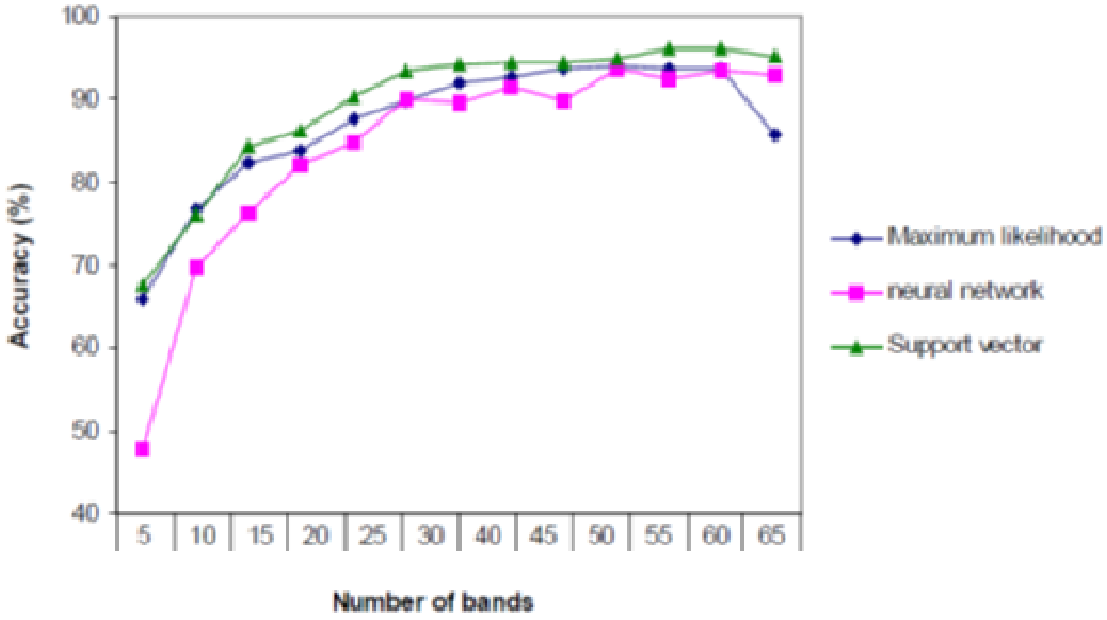


Figure 2.3: Classification accuracies of hyperspectral data (training data size = 200 pixels/class) using SVM, neural network, and Maximum Likelihood [44].

ticular class [28]. That is, a pixel is assigned to a certain class for which the likelihood or probability is maximum. In Figure 2.4, the pixel is assigned to Class 3 as this class has the highest likelihood over all classes for that pixel.

The likelihood function is used in the maximum likelihood method. This function is derived from Bayes' theorem [14]. According to Bayes' theorem, the posterior probability of a pixel belonging to class y given the feature vector X of that pixel is defined as

$$p(y|X) = \frac{p(y).p(X|y)}{p(X)}, \tag{2.11}$$

where $p(X|y)$ is the likelihood function or probability density function to observe the attribute values X given that it belongs to class y [14], $p(y)$ is the prior probability of class y, and $p(X)$ is the probability of the feature vector X being observed. Here, $p(X)$ is the same for all classes. In the maximum likelihood method, it is usually assumed that each

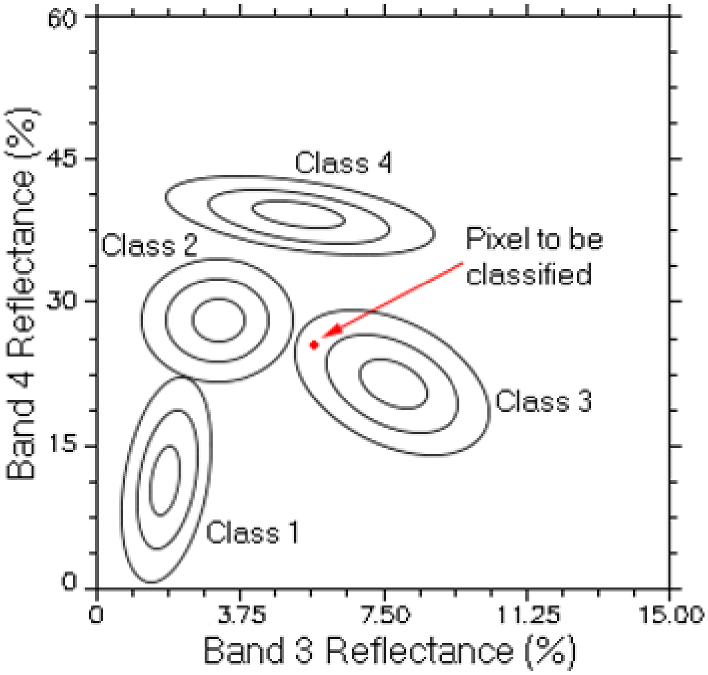


Figure 2.4: Maximum likelihood classification with equal probability contours [28].

class has an equal probability of occuring, which makes it possible to eliminate the prior probability term $p(y)$ [28]. Therefore, likelihood depends only on the probability density function [28, 41]. Figure 2.5 illustrates an example of probability density functions of two training classes (e.g., forest and agriculture) overlapping in the feature space, where the pixel X is assigned to the forest class because the probability density is greater in the forest class at that point than in the agriculture class.

In this method, the normal distribution or Gaussian distribution of the training data is assumed for each class in each band [28]. The probability density function in band k for class y_i can be computed using the following formula:

$$p(x_k|y_i) = \frac{1}{(2\pi)^{\frac{1}{2}}\hat{\sigma}_i} e^{\left(-\frac{1}{2}\frac{(x-\hat{\mu}_i)^2}{\hat{\sigma}_i^2}\right)} \tag{2.12}$$

where x is the pixel value (e.g., reflectance) in band k, $\hat{\mu}_i$ is the estimated mean of all pixel

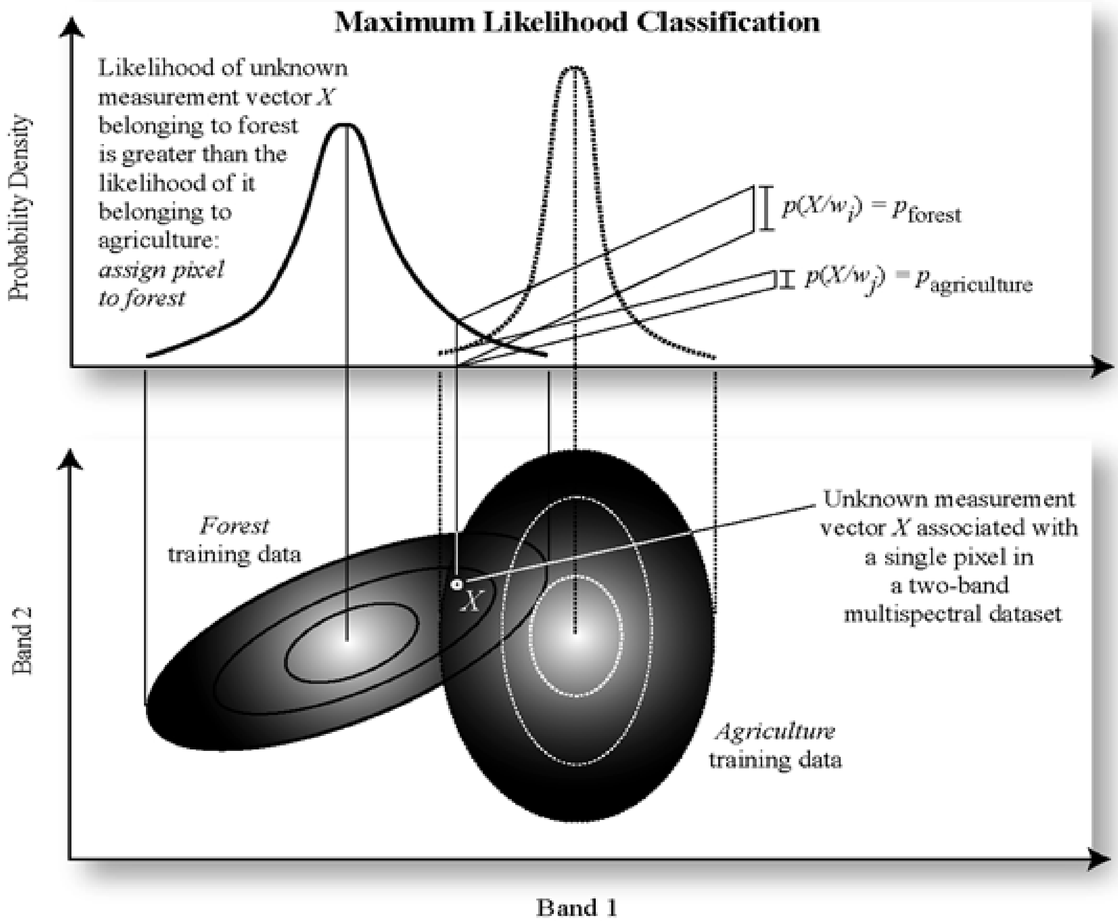


Figure 2.5: An example of overlapping probability density functions of two training classes [28].

values in band k of class y_i, and $\hat{\sigma}_i$ is the estimated standard deviation of all the reflectance values in band k of class y_i. Therefore, the probability function of individual reflectance values is computed based on the mean and standard deviation (variance) of each training class. In addition, an d-dimensional multivariate normal density function is computed using the covariance matrix for multiple bands in the training data of each class. For example, the bivariate probability densitiy functions of normally distributed six hypothetical classes in red and near infrared feature space are illustrated in Figure 2.6.

2.5 Contextual classification

Contextual classification methods make use of both spectral and spatial information. In these cases, a spectral class (per-pixel) classification algorithm is initially applied. The spatial information is then combined with the spectral information. Two methods studied in this thesis are described in this section.

2.5.1 Classification based on the SVM-MRF method

SVM classification is initially used and the probability of each pixel belonging to each class is obtained from the SVM. A number of techniques can be obtained to compute the posterior probabilities. In the next step, certain and uncertain pixels are extracted from the classified map by applying the erosion technique (an image processing technique). The pixels that are situated near the borders of regions or spatial boundaries are referred as uncertain pixels. The assigned labels for these pixels may not be reliable, because the pixels may contain a mixture of multiple classes. Contextual information is used only for the uncertain pixels by the application of the Markov Random Field (MRF) regularization process [33]. The steps of this classification process are shown in Figure 2.7. Applying the

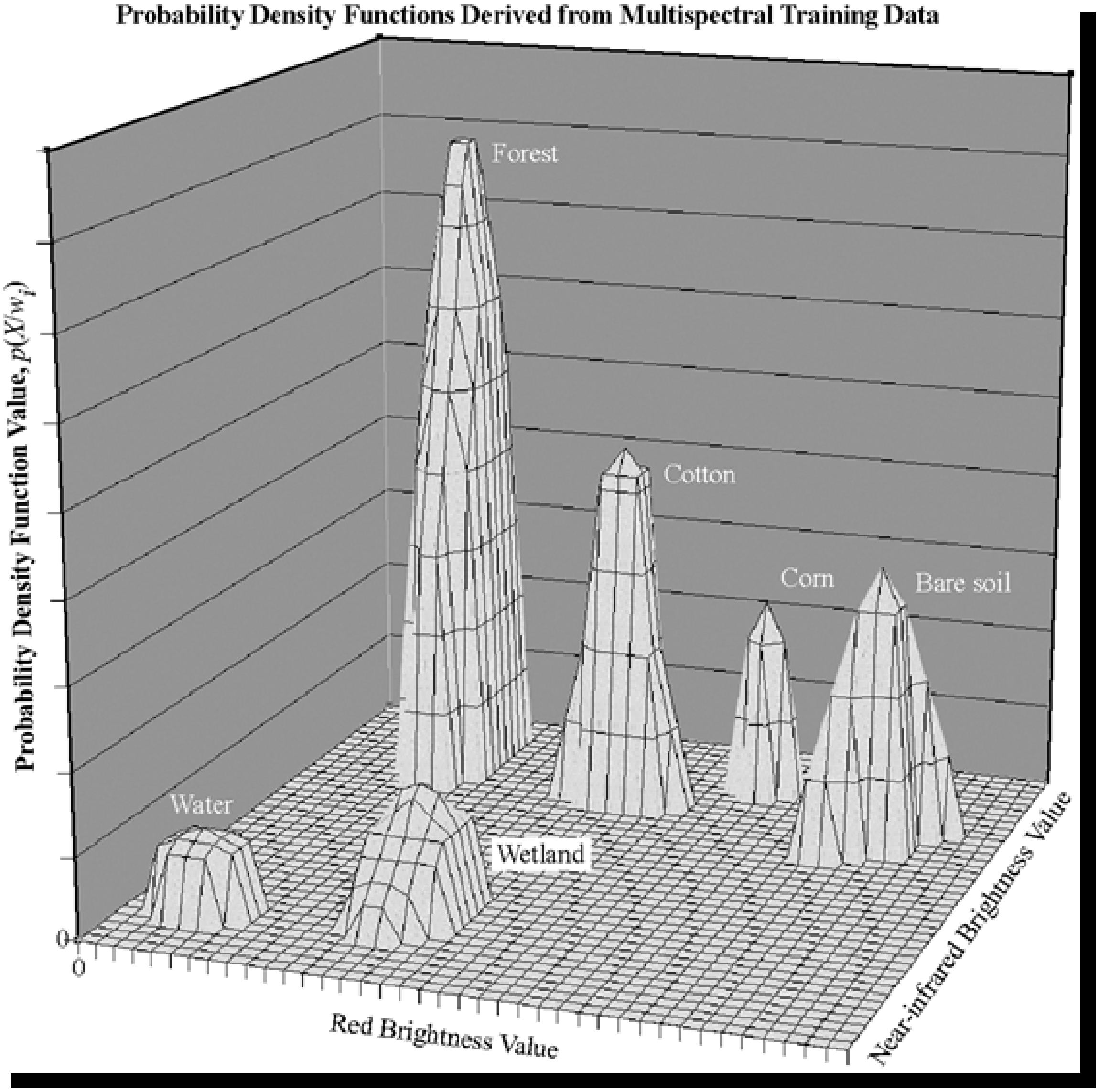


Figure 2.6: Probability density function derived from the training data [28].

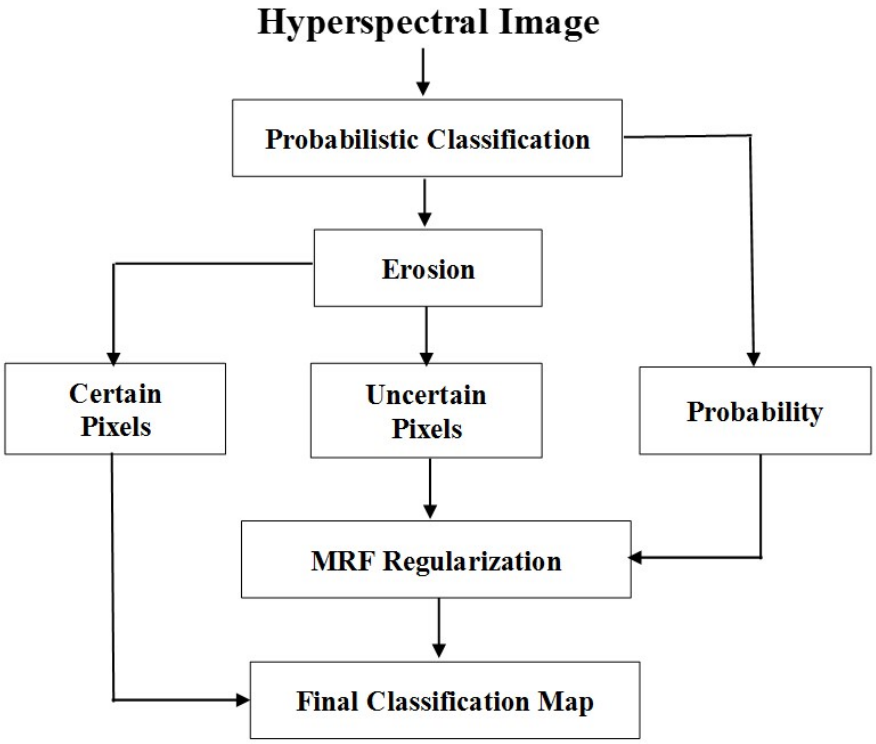


Figure 2.7: SVM-MRF algorithm layout [33].

MRF method only for uncertain pixels reduces processing time compared to applying it to the whole image. It also improves the classification accuracy of boundary pixels.

Probabilistic SVM Classification: The SVM assigns a class label to a pixel based on the posterior probabilities computed as described in Section 2.4.1. Given the posterior probability $p(y_i|X)$ (i = 1,2,...,M, where M is the total number of classes) and the feature vector X, X is assigned to class y_i if

$$p(y_i|X) > p(y_j|X) \quad \forall j \neq i. \tag{2.13}$$

Certain and Uncertain Pixels Extraction: In this step, certain and uncertain pixels are separated from the classified image using the erosion technique.

Erosion is one of the fundamental operations of morphological image processing, which is related to the shape or morphology of features in an image. Morphological techniques extract image components, which describe and represent region shape. These components include skeleton, boundaries and the convex hull. Morphological techniques on shapes only work on binary images. The pixels are either foreground pixels (containing the objects or shapes of interest) or background pixels. A binary image may also be represented as the set of pixel coordinates at which foreground pixels are located. An image is processed by the morphological techniques with a small shape or template, and this shape is called a structuring element. It is positioned at all possible locations in the image, and it is compared with the corresponding neighbourhood of pixels. Some operations shown in Figure 2.8 test whether the element "fits" within the neighbourhood, while others test whether it "hits" or intersects the neighbourhood.

The erosion technique produces a new binary image such that the foreground objects are reduced in size [18]. Let us assume A and B are binary images represented as sets of pixel coordinates. The erosion of A by B, where A is a binary image and B is a structuring element, can be written as $A \ominus B$ and defined as $A \ominus B = \{z | (B)_z \subseteq A\}$. That means erosion of A by B is the set of all points z such that B, translated by z, is contained in A. In other words, the result of erosion is the set of points such that the structuring element B fits in A when placed at those points.

After erosion, the results are the interior parts of a region, which are considered to be certain pixels [33]. Figure 2.9 shows an example of erosion with a 3×3 square structuring element. Erosion with a square structuring element shrinks the foreground objects in an image and the gaps between different regions are enlarged by eliminating small details. The SVM classified image is considered to be a collection of binary images in which each binary image has as foreground the pixels labelled as one particular class. The erosion technique is applied to each binary image to obtain the interior of the regions, and the

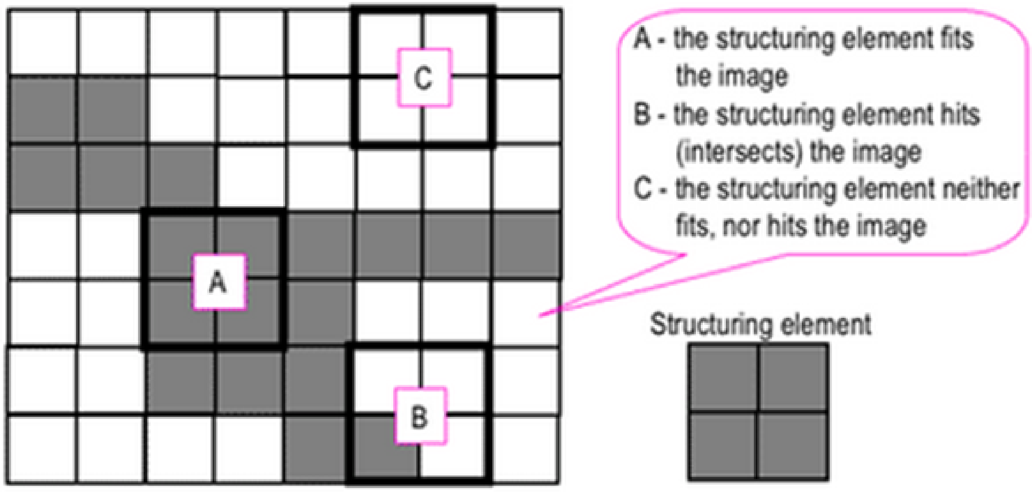


Figure 2.8: Probing of an image with a structuring element [12].

interior is subtracted from the original binary image to obtain the uncertain or boundary pixels.

MRF-Based Regularization of Uncertain Pixels: This is the final step where uncertain pixels of the SVM classified image are regularized by incorporating spatial information. The "energy" at each pixel is defined to measure the uncertainty in the assigned label based on the posterior probability at that pixel and the labelling of adjacent pixels. The MRF process minimizes the total energy over all uncertain pixels until the labelling has stabilized. This process attempts to integrate spatial information into the per-pixel classification result. The local energy is defined as

$$U(a_i) = -\ln\{P(a_i|y_i)\} + \sum_{a_j \in n_i} \beta(1 - \delta(y_i - y_j)) \tag{2.14}$$

where $P(a_i|y_i)$ in the first term is the likelihood of a pixel a_i being observed from class y_i, which can be computed from the posterior probability and Baye's rule. A large likelihood results in low energy in the first term while a small likelihood results in high energy. The

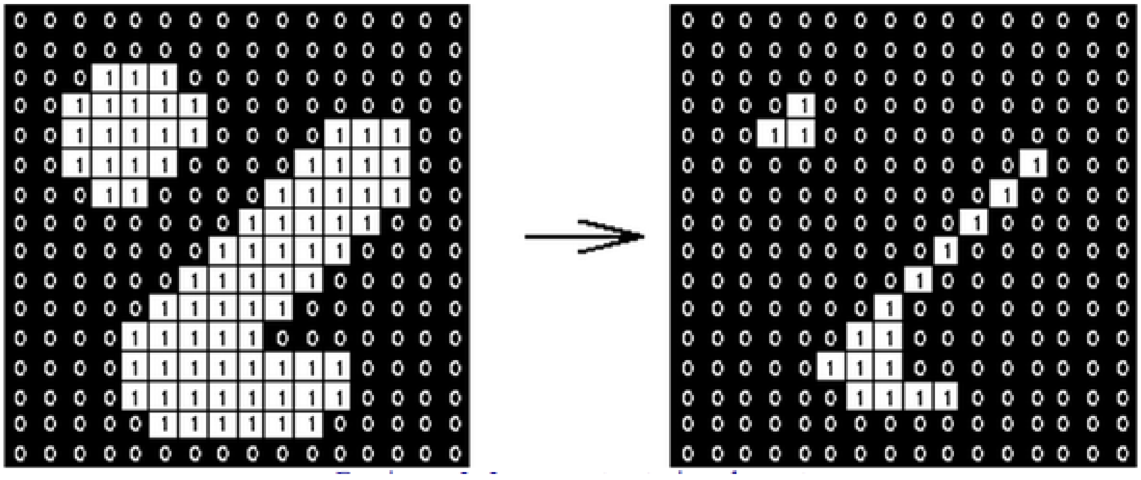


Figure 2.9: Erosion is shown using a 3 × 3 square structuring element [17].

first term is computed only from the spectral information of a pixel. The second term of (2.14) without β is $\sum_{a_j \in n_i} (1 - \delta(y_i - y_j))$, where $\delta(Y)$ is defined as

$$\delta(Y) = \begin{cases} 1 & \text{if } Y = 0 \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{2.15}$$

If the class label y_i for a pixel a_i is the same as the class label of a surrounding pixel, the energy from that term will be 0, otherwise it will be 1. In the second term, n is the total number of pixels in a neighbourhood. The more the labels of neighbourhood pixels match with the given pixel, the smaller the local energy will be from the second term. n_i is the total number of neighbourhood pixels. In the MRF approach, eight neighbourhood pixels were used [33]. We can consider the second term in (2.14) as a way to incorporate spatial information. The defined local energy is a combination of spectral and spatial term, and β controls the balance between the two. For example, a large value of β favours labelling pixels in a neighbourhood the same way. This minimization of local energies is performed over uncertain pixels and continues until labelling has stabilized [33].

It has been shown that the classification accuracy improves by incorporating spatial information and processing time can be reduced by applying this process only on uncertain pixels. The authors in [33] used the AVIRIS hyperspectral data of Indiana Pines in Northwestern Indiana and sixteen classes of different crop types for experiment. The training set was selected randomly from ground-reference data and the remaining samples were selected as the test set. The SVM classification with the RBF kernel was applied, and the probabilistic estimates were then calculated. Certain and uncertain pixels were extracted after applying the erosion technique on the classified image using a disc structuring element with radius 1. Combination of spectral and spatial information improved the SVM classification accuracy by achieving overall accuracy from 82.52 % to 92.07 % as well as reduced the processing time by 57 % compared to the conventional SVM-MRF method in which the MRF method is applied to the entire image instead of only the uncertain pixels.

2.5.2 Integration of Spatial Information Using the UnECHO Method

The Unsupervised Extraction and Classification for Homogenous Objects (UnECHO) classification method attempts to enhance pixel homogeneity of local neighbourhoods [29]. This is actually a refinement of the supervised Extraction and Classification for the Homogenous Objects (ECHO) classifier, where spectrally homogeneous pixels in a local neighbourhood are enforced to the same class [29]. Pixels in a local neighbourhood are classified to the same class if their spectral characteristics are similar. The threshold values required to measure the conditions of homogeneity of all the neighbourhoods are estimated by the algorithm.

The UnECHO method classifies a set of pixels in a neighbourhood based on the degree of heterogeneity or homogeneity. High homogeneity and low heterogeneity of a neighbourhood to a particular class means that all pixels of a neighbourhood belong to that class [29].

UnECHO has two steps. In the first step, the original image is classified using a per-pixel based classifier such as C-means. In the second step, the classification results of the first stage and spectral information in a neighbourhood are used for contextual classification. The image is divided into a number of non-overlapping neighbourhoods. All pixels of neighbourhood t, $\hat{X}^{(t)}$, belong to the i-th class if

$$Q_i\left(\hat{X}^{(t)}\right) < D_i \quad \text{and} \quad Q_i\left(\hat{X}^{(t)}\right) < Q_j\left(\hat{X}^{(t)}\right) \quad \forall j \neq i, \tag{2.16}$$

where $Q_i\left(\hat{X}^{(t)}\right)$ is the distance between the t-th neighbourhood $\hat{X}^{(t)}$ to class i and D_i is the threshold for class i. The first part means that the spectral homogeneity of t-th neighbourhood is large enough and can be classified as class i. The second part means that the degree of homogeneity of all pixels of neighbourhood $\hat{X}^{(t)}$ to class i is the largest compared to any other class j. $Q_j\left(\hat{X}^{(t)}\right)$ is defined as follows:

$$Q_j\left(\hat{X}^{(t)}\right) = \frac{1}{L}\sum_{m=1}^{L} q_j\left(X_m\right), \tag{2.17}$$

where L is the total number of pixels in the t-th neighbourhood, X_m is the m-th pixel of that neighbourhood and $q_j\left(X_m\right)$ is the distance of pixel X_m to the j-th class. Various metrics such as Euclidean distance, Mahalanobis distance, and Maximum likelihood can be used for the measurement of these distances.

1) Euclidian distance [29] :

$$q_j\left(X_m\right) = \left(X_m - \mu_j\right)^T \left(X_m - \mu_j\right), \tag{2.18}$$

2) Mahalanobis distance [29] :

$$q_j(X_m) = \left(X_m - \mu_j\right)^T \sum_j^{-1} \left(X_m - \mu_j\right), \tag{2.19}$$

3) Maximum likelihood [29] :

$$q_j(X_m) = |\sum_j| \left(X_m - \mu_j\right)^T \sum_j^{-1} \left(X_m - \mu_j\right), \tag{2.20}$$

where μ_j is the expected value and $\sum_j$ is the covariance matrix of class j. We can use any one of these based on various situations. For example, the Euclidean distance can be used when the eigenvalues of covariance matrices are nearly zero. Therefore, $Q_j\left(\hat{X}^{(t)}\right)$ is the average of the distances of every pixel to the centroid of the j-th class. If the average is too high, the pixels in the neighbourhood t are not classified to class j.

In the UnECHO method, square neighbourhoods of 2×2, 3×3, or 4×4 are chosen because they are simple and the computational performance is better. Since using larger neighbourhoods require more time to compute the threshold, we should select a small neighbourhood size. However, too small is not good either, because less spatial information is considered. Also if we use a overlapping neighbourhoods, it will affect the execution time, and classification results of overlapping neighbourhoods would need to be merged.

The threshold D_i depends on the composition of θ of the neighbourhood. The composition of a neighbourhood for an M-class problem is defined as the vector $\theta = [n_1, n_2, \ldots, n_m]$, where n_i is the number of pixels in the neighbourhood belonging to the i-th class. The vector θ represents classification distribution of pixels of one neighbourhood. If $\hat{X}^{(t)}$ has a composition θ, the corresponding threshold for the ith class is defined as follows:

$$D(I,\theta) = \frac{\Lambda(I,\theta)}{N_\theta}, \tag{2.21}$$

where $\Lambda(I,\theta)$ is the sum of the degrees of heterogeneity of all neighbourhoods that have the same composition θ and are closer the Ith class than any other class, and N_θ is the number of neighbourhoods that have the same θ composition.

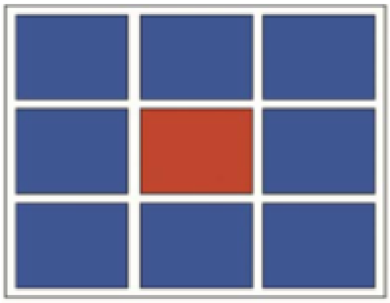
Let us explain the estimation of threshold D_i using an example. Assume we have three 3×3 neighbourhoods with two classes (red and blue) in the initial classified image as shown in Figure 2.10. $Q_j\left(\hat{X}^{(t)}\right)$ is the average distance of the pixels in the t-th neighbourhood $\hat{X}^{(t)}$ to the j-th class. In our example, we define θ = [Number of pixels that belong to the red class, Number of pixels that belong to the blue class] = [1,8]. $\hat{X}^{(t)} \in \theta$ means that neighbourhood t has a θ composition. In our case θ = [1,8] for $t = \{1,2,3\}$ and N_θ= 3.

Let $I^{(t)} = \arg\min_i Q_i\left(\hat{X}^{(t)}\right)$ such that $I^{(t)}$ represent the class having the minimum average spectral distance to neighbourhood t. Let us suppose for our example that $I^{(1)}$= blue, $I^{(2)}$ = blue, and $I^{(3)}$ = red, $Q^{(1)}_{blue}$ = 3, $Q^{(2)}_{blue}$ = 2, $Q^{(3)}_{blue}$ = 4. So Λ(blue, [1,8]) = $Q^{(1)}_{blue}$ + $Q^{(2)}_{blue}$ = 3 +2 = 5. Therefore, D (blue, [1,8]) = 5 / 3.

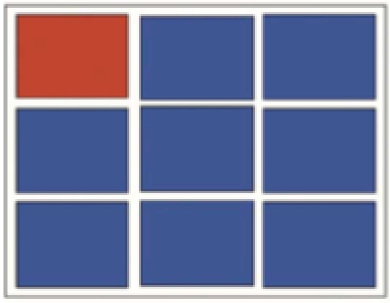
A number of experiments were conducted using hyperspectral datasets to test the UnECHO classifier [29]. The first experiment was conducted using a HYDICE hyperspectral dataset of an urban area with 210 bands and 3.5 m spatial resolution. The UnECHO improved the classification accuracy of C-means clustering by 16 % and detected detailed spatial structures and shapes of large to small scale objects, such as buildings, roads, vehicles, and narrow lines. The second experiment was conducted using AVIRIS data of the Kennedy Space Flight Center in Florida that has 224 bands and 20-m spatial resolution. As in the first experiment, the UnECHO classifier was able to reveal more detailed spatial structure of civil infrastructure, such as road compared to C-means.

2.6 Feature Reduction

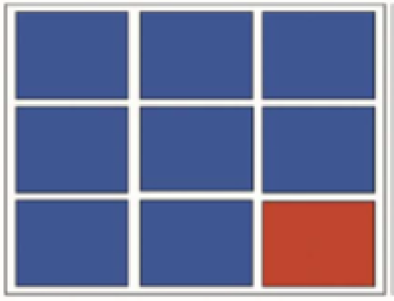
Hyperspectral images are used to classify different land-cover types more accurately than multispectral images because of the high number of contiguous spectral bands provided in



(a) Neighborhood $X^{<1>}$



(b) Neighborhood $X^{<2>}$



(c) Neighborhood $X^{<3>}$

Figure 2.10: Example of classification of a 3 × 3 neighbourhood in the initial classified image [29].

hyperspectral images. However, the performance of many supervised classification methods is strongly affected by the increased number of bands or dimensions in the input data [4, 5]. Every pixel in a hyperspecral image that has hundreds of spectral samples is considered as a vector. The number of dimensions of the vector is the same as the number of bands. The contiguity of bands makes spectral samples within a vector highly correlated. In addition, some bands may contain less discriminative information than others and data redundancy may occur between bands. Data redundancy may obscure information that is important for classification. Moreover, the number of training samples available for classification is limited, but in high dimensional data, a larger number of training samples is needed to avoid error. This problem is called the Hughes phenomenon [43]. Finally, large storage space and computation time are required for high dimensional data. Therefore, dimension reduction without significant loss of essential information is a very important issue. Dimension reduction techniques are generally categorized into two types: feature extraction and feature selection.

2.6.1 Feature Extraction

Feature extraction is a process of feature reduction by projecting or transforming a higher dimensional correlated space into a lower dimensional uncorrelated space [4]. The newly transformed features are actually a linear combination of the original features. Feature extraction produces a set of small and rich attributes that contain the maximum information. There are many feature extraction methods existing in the literature. Among these, commonly used feature extraction methods are Principal Component Analysis (PCA) [28], Independent Component Analysis (ICA) [27], Minimum Noise Fraction (MNF), discriminant analysis feature extraction [16], decision-boundary feature extraction (DBFE) [16], Euclidian distance measurement (EDM) [59], the discrete measurement criteria function

(DMCF) [59], the minimum differentiated entropy method (MDE) [59], and the probability distance criterion (PDC) [59]. In this thesis, only Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are investigated.

Principal Component Analysis (PCA): Principal component analysis is one of the classical dimensionality reduction techniques that can be applied to multispectral and hyperspectral datasets. In PCA, a set of correlated spectral bands is transformed into an equivalent set of uncorrelated components in a feature vector, so that the first fewer components represent most of the information in original data. This technique condenses the information of intercorrelated bands into a few bands or features and these features are called the principal components [28]. To analyze the correlation among the bands, a covariance matrix among the bands is formed. The eigenvectors of this matrix are applied to the band values as a set of weights to obtain the principal component and the associated eigenvalues indicate a measure of variance in that component.

Let us assume each pixel of hyperspectral image data is represented as the vector $\mathbf{X}_i = [x_1, x_2, \dots, x_d]_i^T$, where $x_1, x_2, \dots, x_d$ are the values of the d bands. Let $E = e \times r$, where e and r are the number of rows and columns of the image [50]. The mean vector of all image vectors is denoted as follows:

$$\mathbf{M_v} = \frac{1}{E} \sum_{i=1}^{E} [x_1, x_2, \dots, x_d]_i^T. \tag{2.22}$$

Now, the covariance matrix is calculated by the following formula:

$$\mathbf{C_x} = \frac{1}{E} \sum_{i=1}^{E} (\mathbf{X_i} - \mathbf{M_v})(\mathbf{X_i} - \mathbf{M_v})^T. \tag{2.23}$$

The PCA is dependent on the eigenvalue decomposition, $\mathbf{c_x} = \mathbf{AHA}^T$ of the covariance matrix where $\mathbf{H}$ is the diagonal matrix composed of the eigenvalues and $\mathbf{A}$ is the

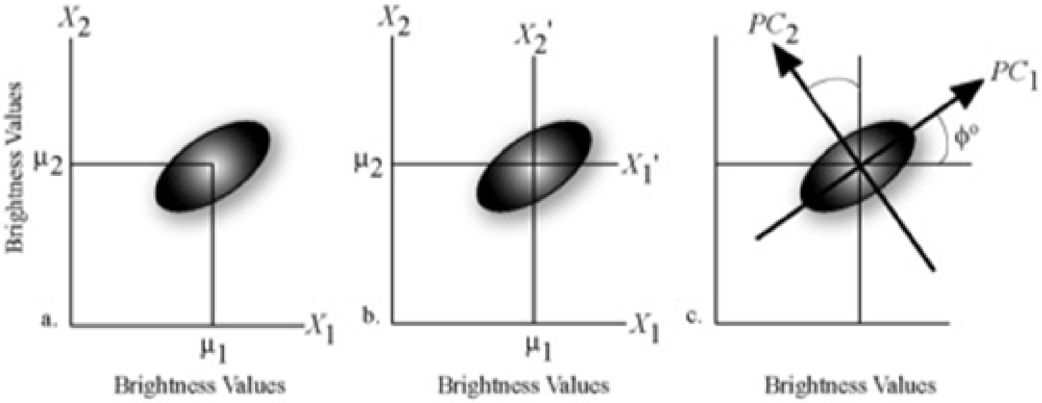


Figure 2.11: PCA transformation process [28].

orthonormal matrix composed of the corresponding eigenvectors. Each feature vector is transformed by $\mathbf{G}_i = \mathbf{A}^T\mathbf{X}_i$ (x_i, $i = 1, \ldots, d$). The principal components are the bands in the transformed vectors.

The resulting principal components are ordered in a way such that the first few principal components contain the maximum portion (e.g., 90 %) of the explanatory power or variance of the original dataset. The remaining components can be dropped in the subsequent analysis without significant loss of information [28]. Generally the first three principal components describe the vast majority of the variance within the dataset. If this is true, the classification process can be performed using just these three principal components. To compute the principal components, a transformation is applied to reproject the data onto each principal component. The transformation is performed by rotating original coordinate axes so that original pixel values are projected onto the principal components (Figure 2.11).

Independent Component Analysis (ICA): In Independent Component Analysis (ICA), a set of mixed, random signals is transformed into mutually independent components and is used in both multispectral and hyperspectral datasets. The transform uses higher order statistics and it is based on the assumption that independent sources are non-Gaussian [26, 27].

ICA has been developed to find a linear representation or transformation of the original data so that the transformed components are mutually independent. Let us assume an observed data vector $\mathbf{X} = [x_1, x_2, \ldots, x_d]^T$ which has a zero mean and d-dimensions or features. The o-dimensional transform of the data vector is $\mathbf{s} = [s_1, s_2, \ldots, s_o]^T$, where s_i is the coefficient of the i-th feature. The linear transformation of the observed vector is then as follows:

$$\mathbf{s} = \mathbf{W}\mathbf{X}, \tag{2.24}$$

where $\mathbf{W}$ is the weight matrix and it is constant. An optimal dimension or feature reduction is computed by estimating the transformation. Therefore, the main objective of ICA is to find a representation by estimating the transformation so that the transformed components become statistically independent as much as possible. In this manner ICA can be considered a special case of redundancy reduction. The matrix $\mathbf{W}$ can be computed by considering the following standard generative model:

$$\mathbf{X} = \mathbf{F}\mathbf{s}, \tag{2.25}$$

where $\mathbf{s}$ is an u-dimensional random vector. The components of $\mathbf{s}$ can not be directly observed and these components are mutually independent. $\mathbf{F}$ is a constant matrix that needs to be estimated. The weight matrix $\mathbf{W}$ in equation (2.24) can be obtained as the inverse of $\mathbf{F}$. Non-Gaussianity of the independent components is necessary, because the matrix $\mathbf{F}$ is not identifiable for Gaussian independent components. In fact the ICA model can be estimated if at most one of the independent components is Gaussian.

The definition of ICA based on the concept of mutual information is provided in [26]. The natural information-theoretic measure of the independence between random variables is called mutual information and is used as a criterion to find the ICA transformation. According to this transformation, the matrix **W** is found in such a way that the mutual information among the transformed components s_i is minimized. Mutual information is non-negative and equals zero when the variables are statistically independent.

2.6.2 Feature Selection

In feature selection techniques, a subset of features is selected to represent the entire feature space. The original representation of the features is not altered. That is, the feature space is not transformed, but instead a subset of the original features is selected that maintains the useful information for separating the classes. Feature selection is a common technique in pattern recognition and machine learning to remove irrelevant, redundant, or noisy features to get a good classification result. A wide variety of feature selection methods have been applied to remotely sensed data. The methods that we used in this thesis, are described in this section.

SVM-Recursive Feature Elimination (SVM-RFE): The SVM-RFE method is a feature selection technique using SVM as a base classifier. SVM-RFE is an efficient, scalable method for many feature selection applications including feature selection of hyperspectral data. SVM-RFE was proposed in [21] for gene selection using linear SVM in a backward elimination procedure. The weight value calculated in the training stage of SVM is used by this method as a feature ranking criterion to produce a list of features ordered by apparent discriminatory ability. The ranking score of feature k is $(V_k)^2$, where V_k is the k-th component of the weight vector defining the hyperplane in equation (2.6). The magnitude of V_k indicates the importance of the feature. At each step, the feature with the smallest ranking score is eliminated. This process is repeated until the desired number of features is reached.

Correlation based Feature Selection (CFS): The CFS method is a simple and fast algorithm proposed by Hall [22] and can be applied to both discrete and continuous problems [23]. The CFS algorithm searches for the best subset of features based on the following hypothesis: *"Good feature subsets contain features highly correlated with the classification, yet uncorrelated to each other"* [22]. A subset of features is selected on the basis of a correlation-based heuristic evaluation function. The heuristic algorithm considers two concepts: the usefulness of individual features for predicting the class (called feature classification correlation) and the level of intercorrelation among the features (called feature-feature correlation). The following equation describes the heuristic function:

$$J_{uv} = \frac{d\bar{J_{uk}}}{\sqrt{d + d(d-1)\bar{J_{kk}}}}, \tag{2.26}$$

where J_{uv} is the correlation between summed features and class variable, d is the number of feature, $\bar{J_{uk}}$ is the average correlation between feature and class, $\bar{J_{kk}}$ is the average feature-feature intercorrelation. Both $\bar{J_{uk}}$ and $\bar{J_{kk}}$ are computed based on conditional entropy. The numerator of this equation indicates how predictive a subset of features is to the class and the denominator indicates the redundancy among the features. The heuristic evaluation function selects the best subset of features that are highly predictive with the class and contains less redundant information. Thus irrelevant features are discarded because of the poor prediction with the class and high correlation with one or more of the other features. The procedure starts with the complete set of features and repeatedly removes one feature at a time to maximize the quantity in equation (2.26).

Minimum-Redundancy–Maximum-Relevance (mRMR): The mRMR method is an improved version of the Max-Relevance feature selection method that implements the maxdependancy scheme [43, 47]. In the mRMR method, features are selected in a way so that they are maximally dissimilar to each other and have the largest dependency on the target class [39, 47]. Mutual information is used in this method to calculate the dependency between features.

Let us assume that f_k represents the feature k and y is the class label. The mutual information of f_k and y in terms of their probabilistic density functions $p(f_k)$, $p(y)$, $p(f_k, y)$ is defined as follows:

$$I(f_k;y) = \iint p(f_k,y)\log\frac{p(f_k,y)}{p(f_k)p(y)}df_k dy. \tag{2.27}$$

According to Max-Relevance, the selected feature f_k should have the highest mutual information value, $Z(f_k;y)$, with the target class y. For the best classification result, it is necessary to consider the dependency between features beside relevance of features on the target class because dependent features create redundancy. In order to minimize the redundancy, a series of calculations of relevance and redundancy is used to select features in mRMR.

In Max-Relevance, a set G of m features are selected in the descending order of $Z(f_k;y)$ or by computing the maximum of the average of all mutual information value of the individual features f_k and class y:

$$\max_G \frac{1}{|G|}\sum_{f_k\in G} Z(f_k;y). \tag{2.28}$$

Because these selected features using the Max-Relevance method may contain high correlation or redundant information, these are not the best selected features. In order to remove redundant information, the Min-Redundancy criterion is used to select the top features. The mutual information is considered in this case too for determining the dependency between each pair of features as follows:

$$\min_{G} \frac{1}{|G|^2} \sum_{f_k, f_j \in G} Z(f_k; f_j). \tag{2.29}$$

Both of the above optimization equations (2.28) and (2.29) are combined in mRMR to get the final set of features. A sequential incremental algorithm is used to perform the simultaneous optimizations, starting from an empty set of features. At each step, a feature is added to the selected subset. Let F_d be the set of all features, and G_m be the set of features selected after the m-th step. Then the m-th feature to be added to G_m is

$$\text{argmax}_{f_j \in F_d - G_{m-1}} \left[Z(f_j; y) - \frac{1}{m-1} \sum_{f_k \in G_{m-1}} Z(f_j; f_k) \right]. \tag{2.30}$$

Alternatively, the m-th feature can also be selected by maximizing the ratio of relevance and dependency as follows:

$$\text{argmax}_{f_j \in F_d - G_{m-1}} \left[\frac{Z(f_j; y)}{\frac{1}{m-1} \sum_{f_k \in G_{m-1}} Z(f_j; f_k)} \right]. \tag{2.31}$$

Chapter 3

Method

3.1 Problem and Solution Description

Various classification techniques based on the spectral properties alone have been proposed in the literature [28]. Among these, two efficient methods are described in Section 2.4. In particular, SVM classification has shown good performance in terms of classification accuracy even if there is a limited number of training samples available for hyperspectral images [44]. Therefore, SVM is used in the proposed method as a spectral classifier. However, this technique classifies pixels without giving attention to the spatial information or the labels of neighbouring pixels.

For any particular pixel, the imaging sensor acquires a significant portion of radiant energy from neighbouring pixels, and for this reason, pixels in the neighbourhood are likely to be assigned the same labels. Spatial information can be used to overcome the salt-and-pepper artifacts of the classification result. These misclassified pixels can be corrected by spatial information to improve classification accuracy. Furthermore, some important information can be extracted from the spatial domain. This includes the shape and the size of the ground-cover type to which a pixel belongs. It is also possible to discriminate between various structures made of the same materials. For example, the roof of a small house and a large building is detected as the same type using spectral information alone, but using spatial information it is possible to classify them into two separate classes in terms of the size of the roof. Therefore, integration of spectral and spatial information is very important.

Many of the existing contextual (spectral and spatial) based approaches consist of one type of spectral and one type of spatial classifiers and spatial information is used either only for the boundary pixels or the whole image of the spectral based classified image (Section 2.5). However, one spatial classifier may not be appropriate for correcting pixels of both the boundaries and the interior parts. The former are found along the class boundaries. These pixels usually contain mixed spectral or class information from adjacent regions due to rasterization of vector data [8]. Consequently, these boundary pixels may have different properties from the pixels found in the interior parts of regions. As a result, classification accuracy may be significantly affected.

The MRF approach is a probabilistic model that is commonly used to integrate spatial context into the image classification process. It is used for correcting misclassified pixels situated in the boundary parts. For correcting errors in the interior part, the UnECHO method is selected, because it integrates spatial information by enhancing pixel homogeneity of local neighborhoods. This process labels a set of pixels in a neighborhood to a class based on the degree of heterogeneity or homogeneity. High homogeneity and low heterogeneity of a neighborhood to a particular class comparing with a threshold means that all pixels of neighborhood belong to that class. Interior parts of regions of a spectral class classified image always have high homogeneity even if any entire region is misclassified. The MRF method may not be able to correct the entire misclassified region as this method labels a pixel examining the probability values and labels of its surrounding pixels. In this case, the UnECHO method may correct the entire region or a group of pixels of this region.

3.2 Proposed Method

The proposed method is divided into the following several steps.

Step 1: The original hyperspectral image is initially classified using the SVM classification algorithm as described in Section 2.4.1. The probability values for each pixel and every class are also obtained from this classification.

Step 2: The erosion technique is applied to the SVM classified image to extract the boundary pixels of each region for each class. The remaining non-boundary pixels are considered to be the interior pixels of those regions.

Step 3: The MRF regularization method is employed only to the boundary pixels of the SVM classified image obtained from step 2. The boundary pixels are regularized by this method in this step.

Step 4: In order to incorporate spatial information, the UnECHO approach is employed in the interior of every region of every class of the SVM classified image.

Step 5: The result of the boundary pixels from applying the MRF method and the result of the interior part of each region from the UnECHO method are combined into one final classification map. The entire process of creating the final classification map is illustrated in Figure 3.1.

The proposed approach (steps 1, 2, 3, 4 and 5), MRF method (steps 1 and 3), UnECHO method (steps 1 and 4) and SVM method (step 1) were investigated separately in order to compare the results.

3.3 Training and Test Data

For supervised classification, ground-reference data, which are used for both the training and test data, are essential. These classification algorithms require training data before the classification procedure can be carried out. The pixels of the unclassified input image are then classified using these training data, which represent the spectral signature of each

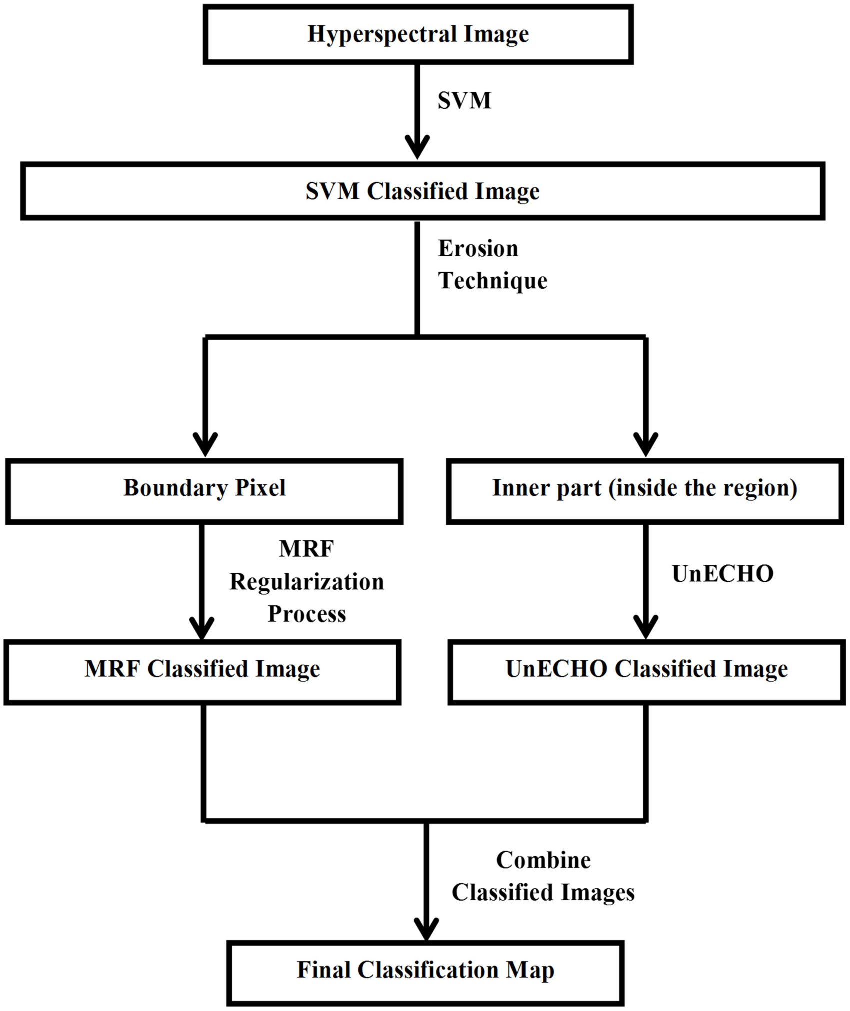


Figure 3.1: Flowchart of the proposed method.

class. Testing data is used for comparing the result of classification or for the accuracy assessment.

The ground-reference data, such as from an agricultural area, forest area or man-made area, are collected at the same time of the acquisition of airborne data. From this data, classes of homogenous regions are created using Regions of Interest (ROIs) by visual examination of the images and with the help of local knowledge of the area from field experts. ROIs can be any combination of polygons, points, or vectors. Therefore, these ROIs or ground-reference data contain all the spectral values with known correct class labels. In this thesis, ground-reference data are split into two parts. One part (10 % of ground-reference data) is used for training and the other part (the ground-reference data excluding the training part) is used for testing.

Supervised classification techniques require a large amount of training data so that the classes are accurately represented but this is very time consuming. However, SVM can perform well even with a small number of training data. Therefore, small ground-reference ROIs data are created, and training ROIs are used as input to the SVM for classification in this thesis. SVM, as a supervised classification method, is trained in order to find the optimal hyperplane for the classification as described in Section 2.4.1. In this thesis, only the SVM method needs to be trained and the reflectance values of training samples are used to compute the centroid of each class for the UnECHO classification. MRF does not require any training data because it is completely unsupervised.

3.4 Spectral Classification (SVM)

The original hyperspectral data is classified using a SVM classifier. The reason for choosing this classifier is that the SVM classifier has performed very well for the classification of remote sensing multispectral and hyperspectral datasets as described in Section 2.4.1.

The SVM classification is carried out using the ENVI remote sensing software [54]. Although SVM was first developed to deal with two-class classification problems, we know from Section 2.4.1 that it can function as a multi-class approach by combining several binary-class classifiers. In ENVI, the One-Against-One Strategy is used by the SVM classifier for the multi-class problem. In addition, ENVI produces the probability values of each pixel for each class, which are actually the decision values stored in ENVI as “rule images”. A pixel is classified to a class for which the probability value is highest. The performance of the SVM classification method relies on several parameters such as the kernel type, K, and penalty parameter, C.

The kernel embeds data into a higher dimensional feature space. It also requires kernel specific parameters. The ENVI SVM classifier has four types of kernels: linear, polynomial, Radial Basis Function (RBF), and sigmoid. Among these four types, the RBF kernel is used, because it works well in most cases [48]. It requires a small number of parameters compared to the other kernels and has shown better results for classifying various satellite imageries [48, 54]. The value of γ in the RBF kernel is a real value greater than zero and by default, it is set as the inverse of the number of spectral bands of the original image [48].

The penalty parameter C, which was mentioned in Section 2.4.1, is used in case of non-separable data to control the trade-off between allowing training errors and forcing rigid margins. The penalty parameter allows some misclassifications. The cost of misclassified pixels is increased according to the increasing value of the penalty parameter. The default value in ENVI is $C = 100$, which allows a few misclassification during the training process, and an accurate model is created [48, 54].

There is an optional threshold that allows pixels with probability values less than the threshold to remain unclassified. In this thesis, the value of the classification probability threshold is set to 0 so that no pixel becomes unclassified [48, 54].

3.5 Boundary Pixels Extraction

The boundary pixels are extracted from the SVM classified image by applying the erosion technique as described in Section 2.5.1. The rule images of each class obtained from the SVM classification are used for this purpose. As the erosion technique works only on binary images, every rule image is converted to a binary image of the pixels belonging to a class for first extracting the interior parts. The boundary pixels were then obtained by subtracting the interior pixels from the rule images. A circular structuring element with radius 1 is used so that one pixel thick boundaries are extracted. Finally, all resulting images containing boundaries of each class are combined into a single image.

3.6 Spectral-Spatial Classification (SVM + UnECHO)

The UnECHO method is applied to the interior pixels, which are obtained using the erosion technique to incorporate spatial information. Using this method, pixel spectral homogeneity is improved in a local neighborhood. The class label of all pixels in a neighborhood are the same if these pixels are spectrally homogeneous.

At the first stage of this method, the entire SVM classified image is divided into a square and non-overlapping neighborhoods. Non-overlapping and square neighborhoods are recommended for this method in order to achieve better results [29]. In this study, 2×2, 3×3, and 4×4 neighbourhoods are examined. The distance is then measured from each neighborhood to every class mean spectrum. The class mean is extracted from the training ROI classes.

Each of the distance measurement techniques in (2.18), (2.19) and (2.20) is examined. The average distance of each neighborhood to each class is computed. Among all of the classes, the class for which the average distance is minimum from the neighborhood is obtained. This class will be assigned to that neighborhood if the distance is less than the threshold computed as described in Section 2.5.2. The overall process of the UnECHO method is shown in Figure 3.2.

3.7 Spectral-Spatial Classification (SVM + MRF)

The MRF regularization method is applied to the SVM classified image consisting of only the boundaries of all classes. In the MRF method, the probability of a pixel belonging to a certain class obtained from the SVM classification, and the number of neighbouring pixels whose labels are the same as that pixel's label, are combined. This is achieved by the MAP-MRF framework where MAP (Maximum A Posteriori) estimate computation of the classification map is carried out by the adopted Metropolis algorithm [33]. As the MRF approach labels each pixel based on its both spectral properties and the labels of its surrounding pixels, it should be effective for correcting misclassified boundary pixels of the SVM classified image.

The main idea of MRF is the minimization of the local energy. Recalling from Section 2.5.1, the computation of the local energy can be rewritten as follows:

$$U(a_i) = U_{spectral}(a_i) + U_{spatial}(a_i), \tag{3.1}$$

where $U_{spectral}(a_i)$ is the spectral energy function from the observed data and $U_{spatial}(a_i)$ is the spatial energy term computed over the local neighbourhood. The spectral energy function is defined as

$$U_{spectral}(a_i) = -\ln\{P(a_i|y_i)\}, \tag{3.2}$$

where, $P(a_i|y_i)$, the posterior probability of pixel a_i for class y_i, is estimated from the SVM

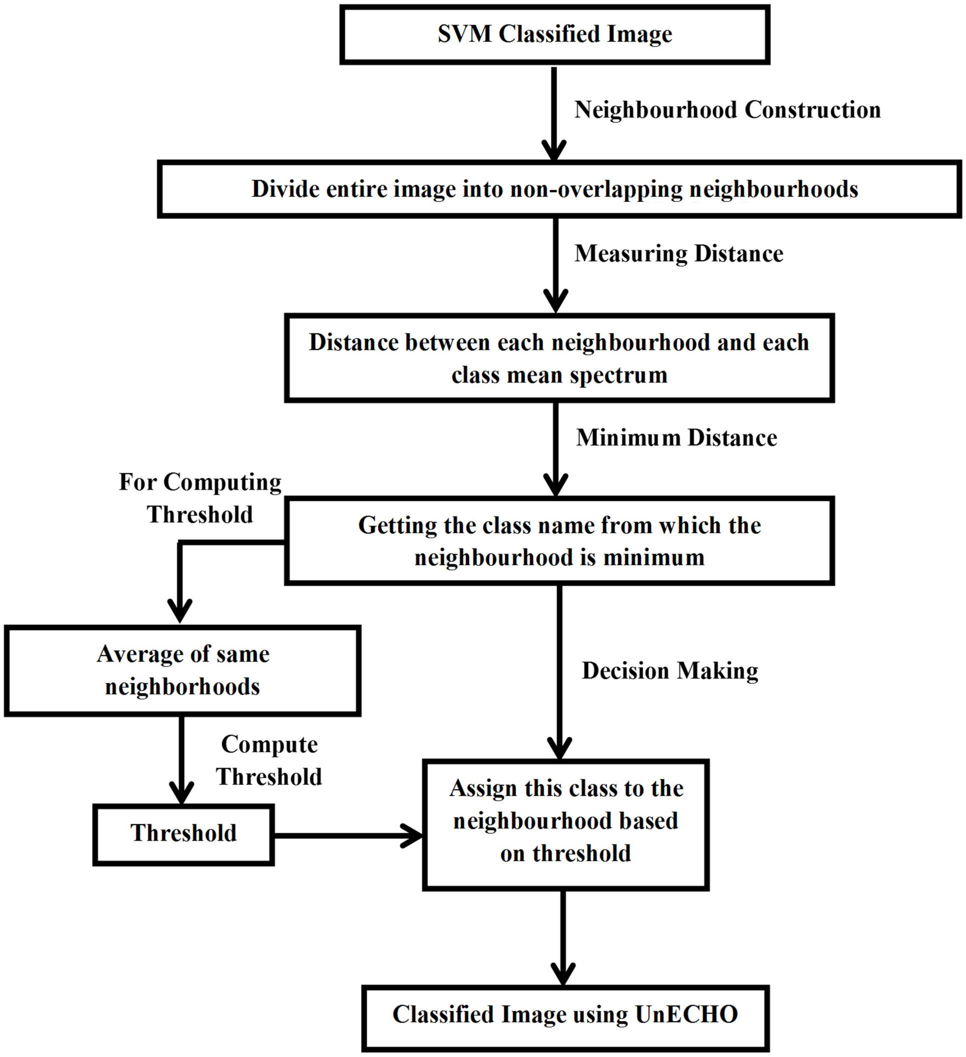


Figure 3.2: Flowchart of UnECHO algorithm.

classification. The spatial energy term is computed as follows:

$$U_{spatial}(a_i) = \sum_{a_j \in n_i} \beta(1 - \delta(y_i, y_j)), \tag{3.3}$$

where $\delta(y_i, y_j)$ is the Kronecker delta function ($\delta(g,h) = 1$ if $g = h$, and $\delta(g,h) = 0$ if $g \neq h$). The label of a given pixel is matched with the labels of each neighbouring pixels. If the labels are same, the value of this term is 1 and because of $(1 - \delta(y_i, y_j))$, the value becomes 0, otherwise, 1. All the values are then summed up. For minimizing the spatial energy, the number of neighbouring pixels whose labels are the same as the given pixel should be greater than the number of neighbouring pixels whose labels are different from the label of the given pixel. The lowest spatial energy is obtained when all the neighbours are labeled as the same as the given pixel. The term $U_{spatial}(a_i)$ is proportional to the number of neighbouring pixels of a_i assigned to one of the classes different from y_i. In the equation of minimizing the local energy, the importance of the spectral and spatial energy term is controlled by the parameter β. The value of β is set based on the priority of spectral and spatial part. In this thesis, the value of the β parameter is examined separately by starting at 1 and then gradually decreasing it by 0.1.

This method works better for images with large spatial structures because of its dependency on the labels of neighbouring pixels. On the other hand, if there is a one-pixel object, which is different from its neighbouring pixels' object, the spatial term then labels this pixel to the class of the surrounding pixels. In this situation, emphasizing the importance to the spectral term might be efficient for keeping the pixel's class label the same as before.

The process of labeling pixels using this method is not an easy problem, because changing the label of a specific pixel may affect the labels of neighbouring pixels. The chain effect makes it difficult to minimize energy for all pixels. Trying to minimize the energy for a pixel might increase the energy of other pixels. Therefore, two approaches are considered to minimize the energy in (3.1) for all pixels.

MRF applied to all pixels of boundaries The energy function (3.1) is applied to every boundary pixel. For each pixel, the local energy is computed for every class label. The class label for which the local energy is the minimum is assigned to the pixel. The entire process for minimizing the energy for all pixels is examined over a number of iterations until the labels stabilizes or if a maximum number of iteration is reached. The drawback of this approach is the large amount of computations required to minimize the energy for all pixels. However, it should provide a good result, because every pixel label is considered. A flowchart of the method is shown in Figure 3.3.

MRF applied to randomly selected pixels of boundaries The minimization of the energy function (3.1) is processed mainly based on the Metropolis algorithm [58]. In this algorithm, the computation of the local energy is done over randomly chosen pixels and the potential new class label for this pixel is also chosen randomly. It is argued that this is an efficient algorithm and takes less computation time. A brief summary of the optimization of the energy function is described in the following paragraph.

In the optimization process, a boundary pixel a_i is randomly chosen repeatedly. Using equation (3.1) the local energy of this pixel $U(a_i)$ is computed. A new class label y_i is randomly selected for this pixel and the new local energy $U^{new}(a_i)$ of this pixel is then computed. The difference between the old and the new local energy is as follows:

$$\Delta U = U^{\text{new}}(a_i) - U(a_i). \tag{3.4}$$

If ΔU, is less than zero, a new (randomly chosen) class label y_i^{new} is assigned to that pixel. Otherwise, the new class assignment is accepted with a probability:

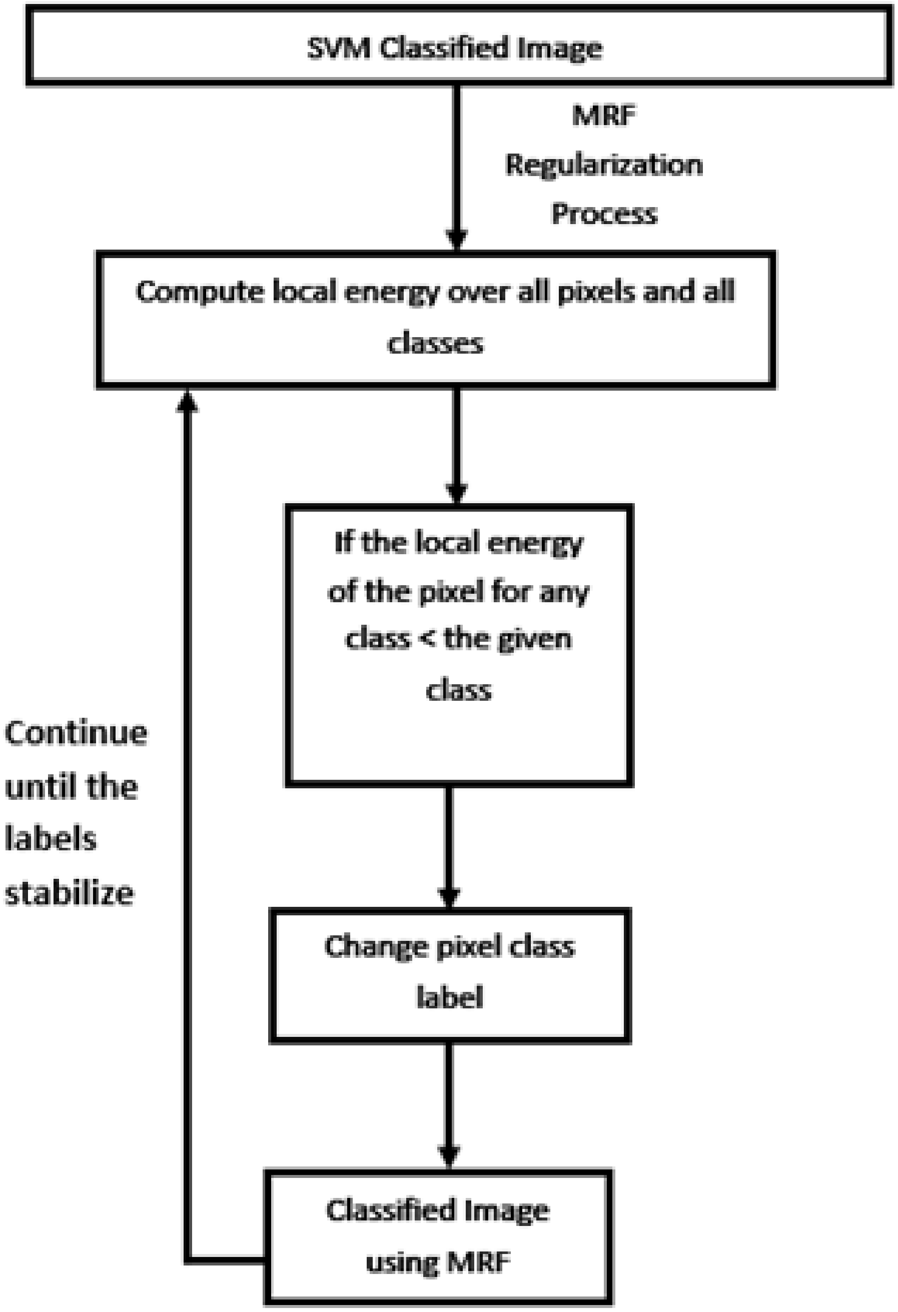


Figure 3.3: Flowchart of the MRF regularization method (Algorithm 1).

$$p = \exp^{(-\Delta U/T)}, \tag{3.5}$$

where T is a global control parameter called the temperature. This probability is used to determine if the new label is accepted. As the temperature decreases, it is less likely for the new label to be accepted if it does not improve local energy. T is varied during the Metropolis relaxation procedure [58]. At first, the value of T is kept high and is then gradually decreased after every l iterations. In this thesis, initially the temperature was set to $T = 1.5$, because using lower initial temperature value results in less computation time. For every new iteration l, the temperature is then recomputed as follows:

$$T^{new} = 0.98.T^{old}. \tag{3.6}$$

For each T value, the number of random pixels and class labels are equal to the total number of boundary pixels and the total number of classes, respectively, in the image. This approach is illustrated in Figure 3.4.

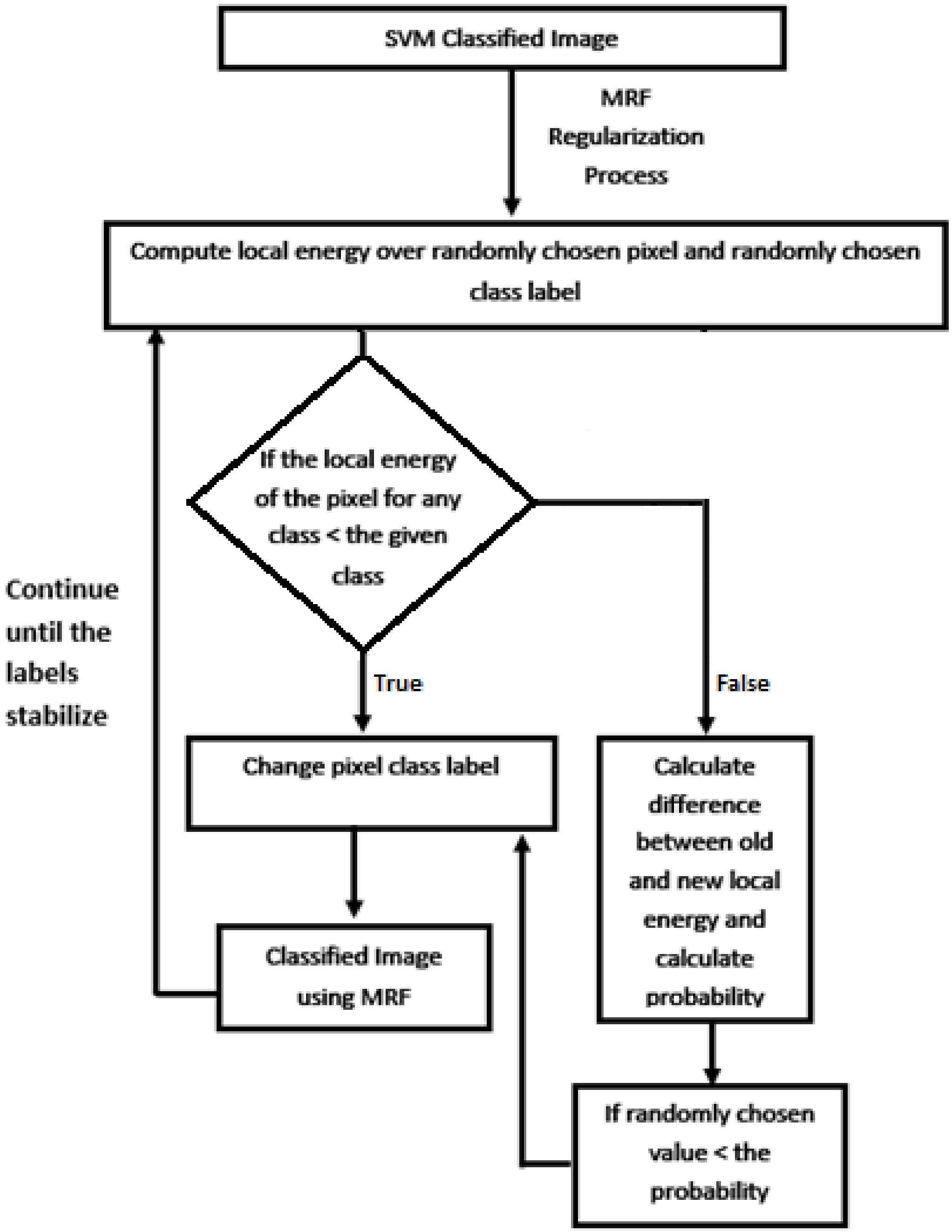


Figure 3.4: Flowchart of the MRF regularization method (Algorithm 2).

3.8 Band Extraction

Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are used as band extraction techniques in this thesis.

Principal Component Analysis (PCA) PCA is used to produce uncorrelated output components, to set apart noise components, and to reduce the dimensionality of datasets (Section 2.6.1). As hyperspectral data are often found highly correlated, PCA is one of the efficient methods that can produce uncorrelated output components. Band extraction using PCA is performed by the ENVI remote sensing software. In ENVI, PCA finds a new set of orthogonal axes that have their origin at the data mean and that are rotated so the data variance is maximized along the axes. As described in Section 2.6.1, Principal Components (PCs) are linear combinations of the original spectral bands and are uncorrelated.

The PCA transformation is mainly a computation of the covariance matrix. There is an option provided by ENVI for performing spectral and spatial subsetting. If there are any bad bands present in the dataset, spectral subsetting is used to exclude them from PCA. Spatial subsetting is controlled by the stats subset parameter. It is used to calculate the covariance matrix based on a spatial subset or the area under an ROI. In this thesis, the covariance matrix is computed from the entire ROI instead of from a spatial subset of the image. Afterwards, ENVI performs the PCA transformation. We generated a number (i.e., Total number of bands - 10) PCs. Initially, 10 of the most important PCs are selected for the experiments (e.g., SVM, MRF, UnECHO) in this thesis. All experiments are performed every time by repeatedly adding 10 PCs of the next most important PCs to the previous set of output PCs.

Independent Component Analysis (ICA) ENVI is used for the ICA transformation as well. As described in Section 2.6.1, ICA is a linear transformation of the input data. The

ENVI software performs the following steps:

- The sample data is centered (subtract the mean to create a zero mean signal) and whitened (transform the observed vector linearly to a new vector in order to obtain uncorrelated components and equal unity variances) using the mean, eigenvectors, and eigenvalues of the data, and the PC rotation is applied for data whitening [54].
- The dimension of the data is reduced by subsetting the whitened sample data when the number of output components is changed.
- Negentropy maximization is performed using the whitened sample to estimate the Independent Component (IC) transform matrix.
- The input image is transformed using the forward IC transform matrix.

Similar to the PCA, the forward ICA transform in ENVI can be customized by a number of options including spatial and spectral subsetting as for the PCA. There are a number of parameters, which control the number of iterations and the accuracy of the IC optimization process. The default values provided in ENVI are used. Among three contrast functions, Logcosh, Kurtosis and Gaussian, we select Logcosh because it is a good general-purpose contrast function and the coefficient value for this function is set 1.0 [54]. Similar to the PCA approach, 10 of the most important ICs are selected for the experiments at first. The experiments are performed every time by repeatedly adding 10 ICs of the next important ICs to the previous set of output ICs.

3.9 Band Selection

SVM-RFE, CFS and mRMR are used as band selection techniques in this thesis.

SVM-RFE The SVM-RFE method described in section 2.6.2 is used for selecting features or bands in this thesis. The main goal of this method is to find a good subset of bands in order to improve the classification accuracy. In a first step, SVM is trained with all the bands and all the training data. The weight values (w_i) are computed for each band in the training stage. If there are m classes, the number of weight values for each band is $(m-1)$. The ranking scores $(w_i)^2$ are then computed. After that two things are investigated. The first one is to find the maximum score for each band and the second one is to sum up all those scores for each band. For both of these cases, the feature with the smallest ranking score is eliminated. The SVM is then trained again with the rest of the surviving bands and the band with the smallest ranking score is eliminated. This process is repeated $(d-10)$ times where d is the total number of bands. The proposed method for the classification is examined at first for the 10 highest ranked bands and then by gradually adding 10 bands according to the rank of up to $(d-10)$.

CFS CFS is implemented according to the description of Section 2.6.2. Inter-correlation between bands, and correlation between a band and a class is considered by the CFS algorithm in order to find a best subset of bands. The heuristic merit of a band subset is calculated using equation (2.26) and evaluated based on a search algorithm. A standard hill climbing approach is used for searching. The current node is expanded and moved to the child node with maximum evaluation in the hill climbing approach. Search space operator is used to expand nodes by deleting one band at a time where the hill climbing approach starts working with a full set of bands. Deletion of bands is continued until a child node is not worse than its parents and finally, the remaining bands are considered as best subset of bands. This band subset is used for the experiments.

mRMR The mRMR approach is a combination of Max-Relevance and Min-Redundancy approaches and selects maximally relevant and minimally redundant features or bands. This approach is implemented based on the mutual information (Section 2.6.2). Mutual information between two variables is computed by (2.27) and used to calculate both criteria: Max-Relevance and Min-Redundancy. Max-Relevance, which uses Max-Dependency scheme finds a set of bands using (2.28) by taking mean of all mutual information between individual band and class, and therefore have greatest relevance or largest mutual information with the target class. These set of bands are not best in order to achieve good classification accuracy, because there might be redundancy among these. Hence, the condition of Min-Redundancy is considered and calculated using (2.29), which results mutually exclusive bands. Both Max-Relevance and Min-Redundancy criteria are combined and optimized using an incremental search condition (2.30) in order to add one band at each step with the selected bands. All experiments are performed using band subsets 10 at first and then adding 10 bands at a time of up to (d-10).

Chapter 4

Experiments for Hyperion Dataset

4.1 Hyperion Dataset

The EO-1 Hyperion and ground-reference data of an agricultural area near Lethbridge, Southern Alberta (4944'N, 11234'W) provided by the Agriculture and Agri-Food Canada (AAFC) was used in our experiment. The Hyperion data were acquired on July 2, 2005 and the spatial resolution is 30 m. The Hyperion data have 242 contiguous bands from 350 to 2582 nm with a 10-nm bandwidth. A subset of this data (Figure 4.1) is used in this experiment in order to reduce the computation time. This subset has dimensions of 200 pixels × 145 pixels × 160 bands. Ground-reference data was collected through a field survey and provides information of 10 different crop types and fellow for 53 fields. The crop types are alfalfa (Medicago sativa L.), barley (Hordeum vulgare L.), beans (Phaseolus vulgaris L.), canola (Brassica napus L.), corn (Zea mays L.), flax (Linum usitatissimum L.), grass (mixed grass species), potato (Solanum tuberosum L.), sugar beet (Beta vulgaris L.), and wheat (Triticum aestivum L.). In the chosen subset, six of these crop types are present as shown in Figure 4.2 and Table 4.1, and were used in this thesis. Figure 4.3 shows the spectral signature of the six ground-reference crop types (classes). The crop fields in this imagery were damaged by flooding, hailstorms, leaf diseases, and fertilizer leaching [2]. A few other, non-agricultural land-cover types are present in the image, including uncultivated grassland, roads, and water.

The preprocessing of the Hyperion data was performed by [45] in the Imaging Spectrometer Data Analysis System (ISDAS) [56]. The preprocessing steps are shown in Figure 4.4. At first, the right half of the image is shifted up by one-pixel in the Short-Wave Infrared Region (SWIR) band in order to correct the vertical offset. Next, the auto-destriping

Figure 4.1: Subset of the Hyperion image used for the experiments.

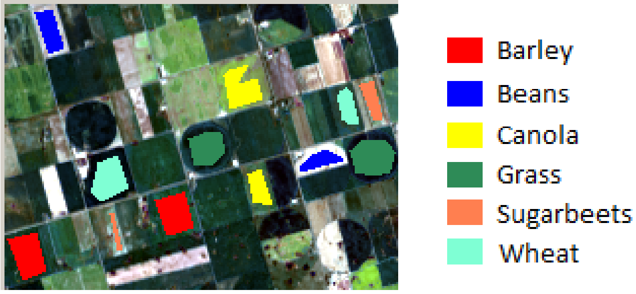


Figure 4.2: Regions of interest for the 6 crop types.

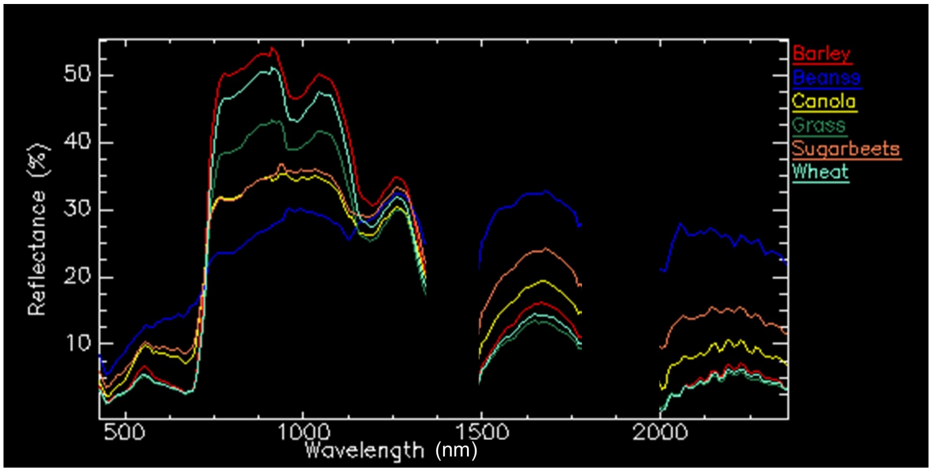


Figure 4.3: Spectral signatures of the 6 crop types.

Table 4.1: Regions of interest for the 6 crop types.

ROI Name	**Color**	**Pixels**
Barley (1 and 5)	Red	594
Beans (2 and 3)	Blue	323
Canola (6 and 7)	Yellow	418
Grass (2 and 3)	Sea Green	593
Sugarbeets (3 and 4)	Coral	178
Wheat (1 and 5)	Aquamarine	428

tool in ISDAS was used to remove striping artifacts and the mixed dropouts. An angular shift of 22 degrees between the Visible Near Infrared Region (VNIR) and SWIR datasets is then corrected by the Align Detector tool in ISDAS. According to the procedure of [34] bands 1–7 and 221–242 were removed because of insufficient signal and information in those bands. Bands 56 to 76 were deleted because this is where the two Hyperion sensors (VNIR and SWIR) overlap. Bands 120 to 133 and band 165 to 182 were removed because they carry no useful information. Solar radiation in this two regions (1346 - 1477nm and 1800 - 1972 nm) is absorbed by water vapor in the atmosphere (these are water absorption features).

The remaining bands (160 bands) are treated as follows. Random noise is removed and Signal-to-Noise Ratio (SNR) is increased by the Average Smooth tool in ISDAS. The atmospheric correction was performed by the central element (MODTRAN radiative transfer model) in ISDAS in order to convert Hyperion-corrected at-sensor radiance data to surface reflectance. Altitude of sensor, latitude and longitude of the image centre, ground elevation, atmospheric and aerosol model are used to perform the atmospheric correction. An across-track wavelength shift shows the form of spectral line curvature or smile/frown artifacts. Smile/frown artifacts cause slightly different band centre wavelengths from the leftmost to the rightmost column in a single band. This effect was detected before atmospheric cor-

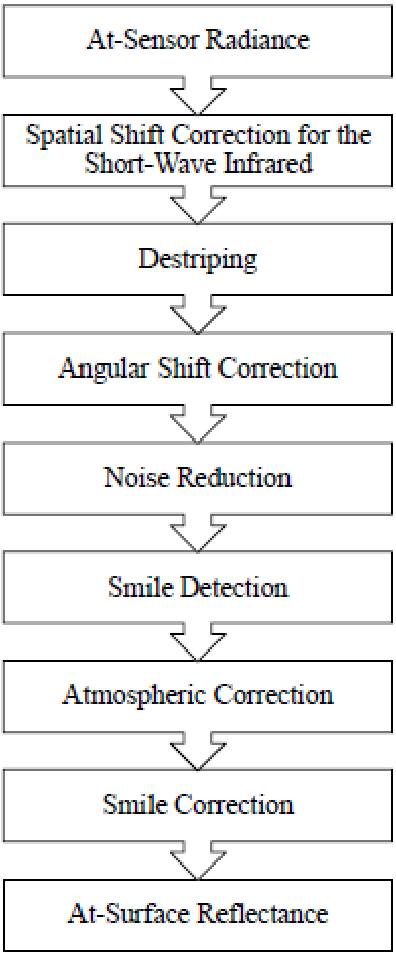


Figure 4.4: Preprocessing steps of the Hyperion data [45].

rection and removed after atmospheric correction. Finally, in order to remove remaining spikes, which may occur due to systematic calibration and atmospheric correction errors, ENVI's EFFORT (Empirical Flat Field Optimal Reflectance Transformation) polishing tool was used.

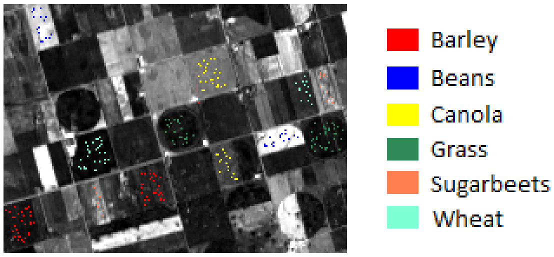


Figure 4.5: Random pixel selected from each class ROI.

4.2 Hardware and Software

The experiments were conducted on a laptop with an Intel Core i7–3517U 2.40–GHz processor and 6.00 GB of RAM, running under the Windows 8 operating system. The SVM classification, the feature extraction methods (PCA and ICA) and the accuracy assessment of all experiments were performed using ENVI version 4.7 [54]. The MRF and UnE-CHO methods, and SVM-RFE feature selection algorithm were implemented in MATLAB (Mathworks 2011) [35]. Finally, The CFS and mRMR feature selection algorithms were implemented in the C++ programming language.

4.3 SVM Experiment

As a supervised classification method, SVM needs appropriate training samples for classifying the data and 10 % of the pixels from each class ROI region of Hyperion data (barley, beans, canola, grass, sugarbeets, and wheat) were randomly selected to train the SVM classifier (Figure 4.5).

The SVM classified image of Hyperion data is shown in Figure 4.6. The accuracy assessment of the classification is carried out using the ground-reference ROI of the 6 classes in the subset scene excluding the random training samples (Figure 4.7). The overall accuracy of SVM classification is 89.35 %, which is quite satisfactory. Tables 4.2 and 4.3

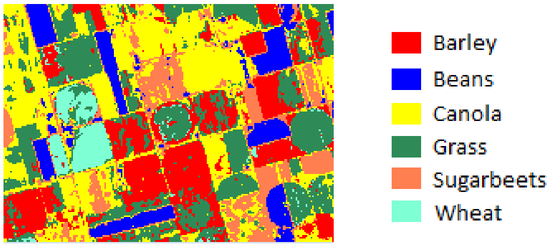


Figure 4.6: Resulting classification using the SVM.

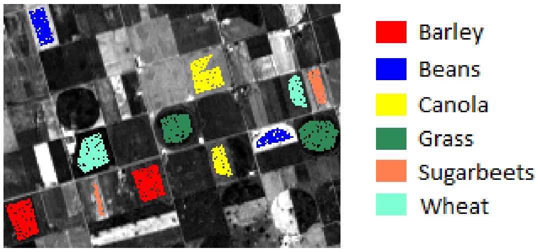


Figure 4.7: Testing ROIs without the training random samples.

show the confusion matrices, and producer and user accuracies of the SVM classification, respectively, in number of pixels and percentages for all classes. Among these classes all the bean pixels are classified correctly and 68.05 % of the wheat pixels are correctly classified, which was the lowest percentage.

Although the performance of SVM is good, there are errors when we compare the SVM classification result against the the ground-reference ROI (Figure 4.8). For example,

Table 4.2: Confusion matrix (pixels) of SVM experiment using Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	Total
Barley	**491**	0	19	15	0	7	532
Bean	0	**291**	0	0	0	0	291
Canola	3	0	**357**	0	19	4	383
Grass	41	0	0	**496**	0	112	649
Sugarbeets	0	0	0	0	**141**	0	141
Wheat	0	0	0	23	0	**262**	285
Total	535	291	376	534	160	385	**2281**

Table 4.3: Confusion matrix (percent) of SVM experiment using Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	User Acc. (%)
Barley	**91.78**	0.00	5.05	2.81	0.00	1.82	92.29
Bean	0.00	**100.00**	0.00	0.00	0.00	0.00	100.00
Canola	0.56	0.00	**94.95**	0.00	11.88	1.04	93.21
Grass	7.66	0.00	0.00	**92.88**	0.00	29.09	76.43
Sugarbeets	0.00	0.00	0.00	0.00	**88.13**	0.00	100.00
Wheat	0.00	0.00	0.00	4.31	0.00	**68.05**	91.93
Prod. Acc. (%)	91.78	100.00	94.95	92.88	88.13	68.05	

the entire wheat region (shown in the double circle in Figure 4.8) is classified as grass and barley. We can see that the spectral signature of grass, barley and wheat crops are very similar (Figure 4.3). The image data was acquired in early July and at this time barley and wheat crops look similar to grass. For this reason the wheat region is classified as grass and barley. Moreover, the two ground-reference wheat regions are actually two different types of wheat, which are trained as one crop type. Different crop fields of the same type of crop might have different phenological stages as well as flood damages. These factors can change the spectral characteristics of same crops in different fields. If the SVM is trained with both types of wheat separated into two classes, it can classify correctly both wheat regions. Figure 4.9 shows the SVM classification result where 7 classes (including both types of wheat) were used and the correctly classified wheat region is shown in circle. The overall accuracy then increases to 94.17%. Although the accuracy is better by training the SVM with more types of crops, the quality of the classifier might be lower because of overtraining. For example, the ground-reference information is only available for two types of wheat but there could be more types of wheat with more phenological variation despite having similar spectral characteristics. The success of a classifier depends on how better it performs using smaller number of ground-reference pixels for each class type. If we consider this factor, SVM produced better result for each crop type without using their subtypes except for the wheat crop.

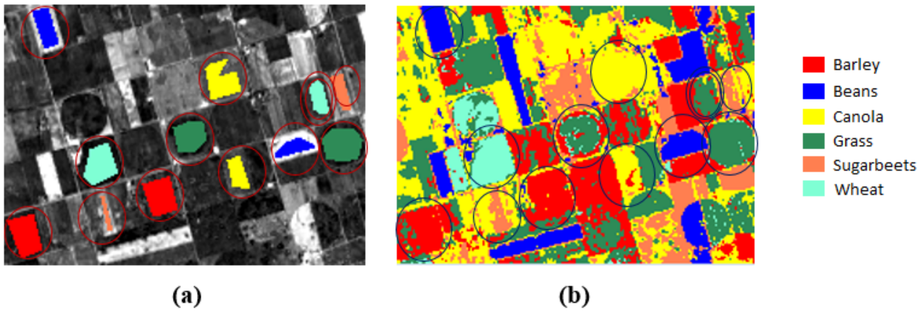


Figure 4.8: Visual comparison of SVM classification result against ground-reference ROI. (a) Ground-reference ROI . (b) Corresponding areas (circled) in the SVM classified image.

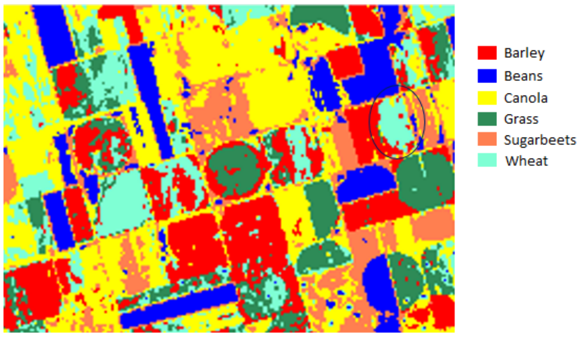


Figure 4.9: Result image using the SVM classification where both types of wheat pixels were used for training.

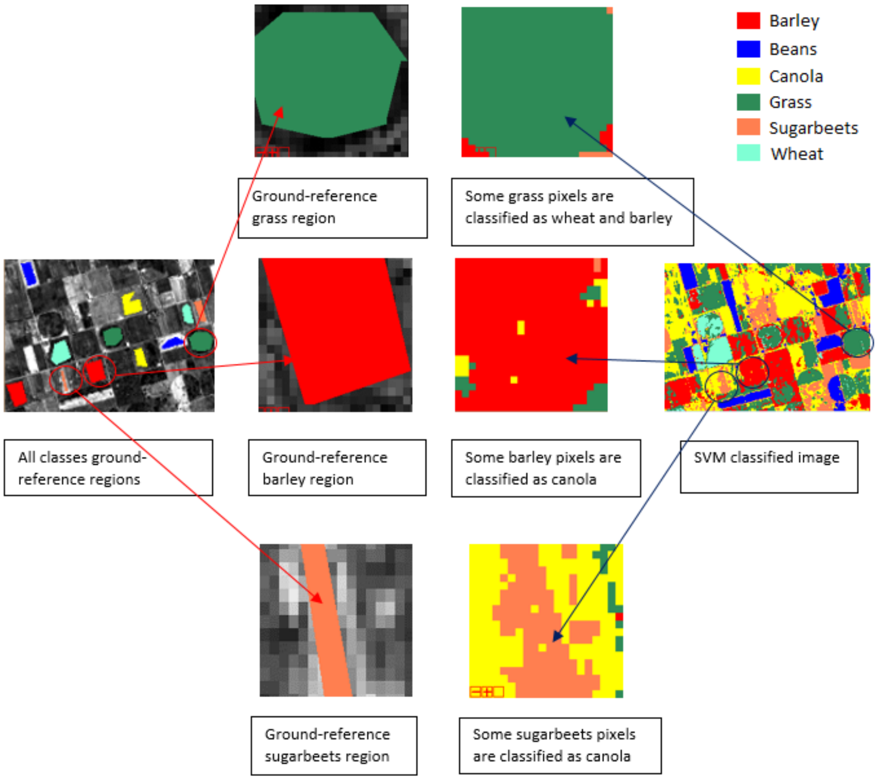


Figure 4.10: Examples of misclassified pixels of SVM.

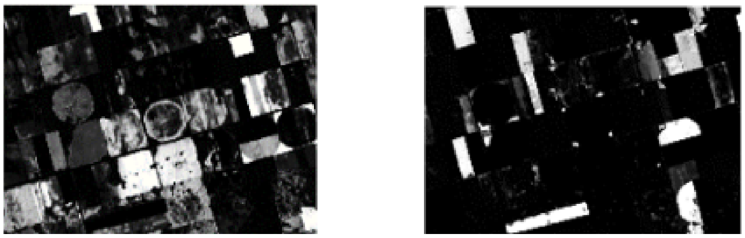
There is a lot of salt and pepper noise in the SVM classified image where pixels are classified differently from their neighbouring pixels. Some examples of misclassified pixels are shown in Figure 4.10 where grass is misclassified as wheat and barley, barley as canola, and sugarbeets as canola. The Hyperion data set covers an agricultural area and the regions of this data should be quite homogeneous, unless there is poor vegetation health due to natural disasters, fertilizer leaching, and leaf diseases. It is very unlikely to have one type of crop inside another crop field covering at least a 30 m × 30 m area (i.e., a pixel).

4.4 MRF Experiment on All Boundary Pixels

In this experiment, we apply only the MRF method to the boundary pixels after the SVM classification is done. The MRF regularization is applied to all boundary pixels (Algorithm 1 in Figure 3.3). In addition, this method is also applied to the entire image of the SVM classification for comparison. The probability values of each pixel for every class of the SVM classification obtained from the rule images, which were produced during SVM classification, are used in the MRF method.

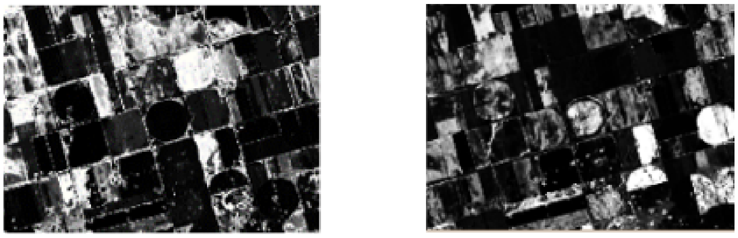
All the rule images of the SVM classification of the Hyperion dataset are shown in Figure 4.11. These rule images of each class give the probability of each pixel belonging to a class and are also used in the erosion technique (Section 3.5) to extract the interior pixels of regions of the SVM classified image. The resulting interior pixels of the regions are shown in Figure 4.12. The boundary pixels of the regions are obtained by subtracting the interior pixels from the SVM image (Figure 4.13).

Different values of β in equation (2.14) for local energy computation are tested to examine the sensitivity of this algorithm to this parameter value and the results are shown in Tables 4.4 and 4.5. Table 4.4 shows the overall accuracies and run time (in seconds) of the MRF classification applied to the boundary pixels of the SVM classification and Table 4.5 presents the overall accuracies and run time (in seconds) of the MRF classification applied to the entire image of the SVM classification for different β values. We used 15 iterations in the regularization process for both cases. From these tables we can see that increasing the value of β increases the accuracy. That means better accuracies are achieved by giving priority to the spatial correlation although after a certain point ($\beta = 0.80$) the accuracies do not improve. The issue of whether the improvement of the overall method is significant, will be examined later. We also examined different number of iterations for regularization in the MRF method and the results are shown in Tables 4.6 and 4.7. Table 4.6 shows the



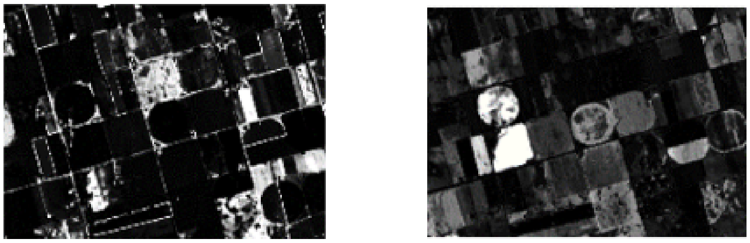
Barley

Beans



Canola

Grass



Sugarbeets

Wheat

Figure 4.11: SVM Rule images produced by ENVI for each crop type shown in gray scale using a probability range from 0 to 1.

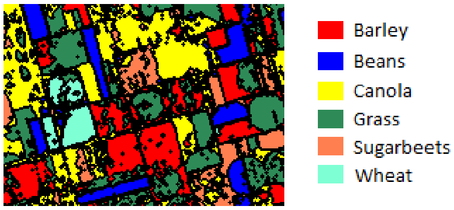


Figure 4.12: Interior pixels of regions of the SVM classified image.

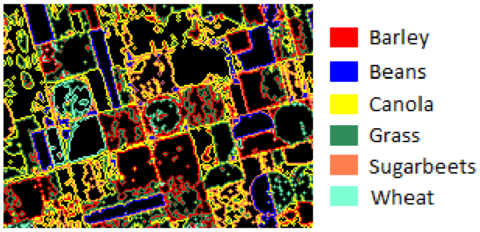


Figure 4.13: Boundary pixels of regions of the SVM classified image.

accuracies and run time (in seconds) of the MRF classification applied to the boundaries and Table 4.7 presents the accuracies and run time (in seconds) of the MRF classification applied to the entire image of the SVM classification. For both cases the value of β is set to 0.80. Table 4.7 shows that the accuracy improves initially when the number of iterations is increased, but does not improve for more than 15 iterations. In Table 4.6, we can see that the best accuracy is achieved using only 5 iterations. Based on this, we assume that using the value $\beta = 0.80$ for both cases, 5 iterations for the MRF on boundaries and 15 iterations for the MRF on the entire image gives the best accuracies for the Hyperion data. We use these values in our remaining experiments.

Table 4.4: Examination of the MRF method applied to the boundary pixels for different values of β.

Value of β	**Overall accuracy (%)**	**Run time (s)**
1	92.81	26.60
0.90	92.81	26.29
0.80	92.81	26.56
0.70	92.72	26.64
0.60	92.20	26.37
0.50	91.98	26.79
0.40	91.98	26.60
0.30	91.45	26.59
0.20	91.32	26.42
0.10	90.62	26.64

Table 4.5: Examination of the MRF method applied to the entire image for different values of β.

Value of β	**Overall accuracy (%)**	**Run time (s)**
1	93.42	55.73
0.90	93.42	55.62
0.80	93.42	55.75
0.70	93.34	55.48
0.60	92.24	55.40
0.50	92.02	55.22
0.40	92.02	55.70
0.30	91.45	55.18
0.20	91.32	55.02
0.10	90.62	55.48

The resulting image using the MRF method on the boundaries of the SVM classification is shown in Figure 4.14. The overall accuracy is 92.81 % and the processing time is 8.84 seconds. The overall accuracy is improved by 3.46 % compared to using only the SVM classification. This improvement is not so high numerically. However, if we compare this result visually with the SVM result, we can see that the MRF result is more homogeneous and contains less salt and pepper noise than the SVM result (Figure 4.15). The MRF method corrects many misclassified pixels. For example, all of the misclassified pixels by SVM shown in Figure 4.10 are corrected by MRF (Figure 4.16). The confusion matrices, producer accuracies and user accuracies of the MRF experiment are shown in Tables 4.8 (in pixels) and 4.9 (in percentages), respectively. It is noticeable from the diagonals of the confusion matrices that the number of correctly classified pixels is improved for most of the crop types such as barley, canola, grass and sugerbeets. The producer accuracies of these crop types are also increased, with the best improvement of 9.37 % for sugarbeets. Because SVM could not classify one entire wheat region, the MRF method cannot improve the accuracy of this region due to the fact that the MRF method uses the classified labels

Table 4.6: Examination of the MRF method applied to the boundaries for different number of iterations.

Number of iterations	**Overall accuracy (%)**	**Run time (s)**
2	91.80	3.72
5	92.81	8.84
10	92.76	17.71
15	92.81	26.55
20	92.76	34.92
25	92.81	44.25
30	92.76	55.54
35	92.81	61.60
40	92.76	65.35

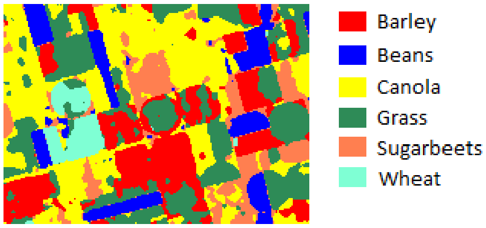


Figure 4.14: Result image using the MRF method applied only to the boundaries.

of the surrounding pixels to perform the correction. In this case, all surrounding pixels are misclassified.

Figure 4.17 shows the result of the boundary pixels only without interior regions of the MRF classification that applied only to the boundary pixels. Although there is no ground-reference data for most boundary pixels, some of them are artifacts due to salt and pepper noise introduced by the SVM classification. From the experimental results, we can conclude that these pixels are effectively corrected by the MRF method.

The resulting image using the MRF method on the entire image of the SVM classified result is shown in Figure 4.18 along with a visual comparison to the result of the MRF method applied to the boundaries. Both of these provide very similar results. The overall

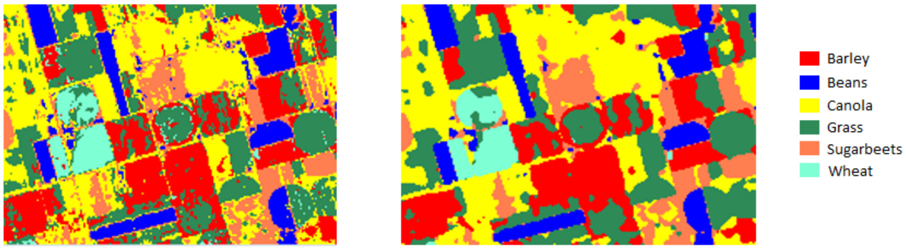


Figure 4.15: Visual comparison of results achieved with SVM (left), and MRF on boundaries (right).

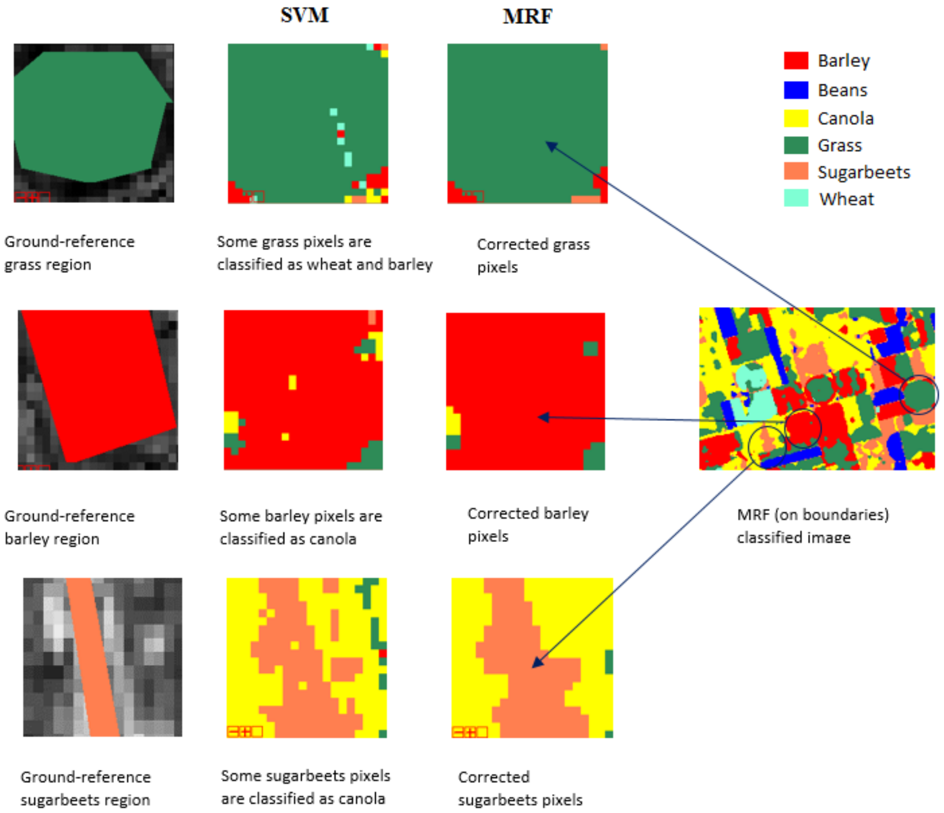


Figure 4.16: Examples of corrected misclassified pixels by MRF.

Table 4.7: Examination of the MRF method applied to the entire image for different number of iterations.

Number of iterations	**Overall accuracy (%)**	**Run time (s)**
2	91.80	7.52
5	92.94	18.48
10	93.16	36.90
15	93.42	54.52
20	93.42	72.32
25	93.42	92.18
30	93.42	109.75
35	93.42	126.56
40	93.42	144.95

Table 4.8: Confusion matrix (pixels) of the MRF experiment (on boundaries) of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	Total
Barley	**516**	0	13	3	0	0	532
Bean	0	**291**	0	0	0	0	291
Canola	0	0	**363**	0	4	0	367
Grass	19	0	0	**529**	0	123	671
Sugarbeets	0	0	0	0	**156**	0	156
Wheat	0	0	0	2	0	**262**	264
Total	535	291	376	534	160	385	**2281**

accuracy increased by only 0.61% to 93.42 % compared to the MRF boundary approach. The confusion matrices, producer accuracies and user accuracies of the MRF classification on the entire image are shown in Tables 4.10 (in pixels) and 4.11 (in percentages), respectively. It can be observed that there are only 13 more correct barley pixels and 1 more correct grass pixel in Table 4.10 compared to Table 4.8. The MRF classification on the entire image requires 55 seconds processing time whereas the MRF classification on the boundaries needs only 8 seconds. Therefore, it is quite acceptable to apply the MRF method only to the boundaries rather than the whole image even though the accuracy is slightly lower.

Table 4.9: Confusion matrix (percent) of the MRF experiment (on boundaries) of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	User Acc. (%)
Barley	**96.45**	0.00	3.46	0.56	0.00	0.00	96.99
Bean	0.00	**100.00**	0.00	0.00	0.00	0.00	100.00
Canola	0.00	0.00	**96.54**	0.00	2.50	0.00	98.91
Grass	3.55	0.00	0.00	**99.06**	0.00	31.95	78.84
Sugarbeets	0.00	0.00	0.00	0.00	**97.50**	0.00	100.00
Wheat	0.00	0.00	0.00	0.37	0.00	**68.05**	99.24
Prod. Acc. (%)	96.45	100.00	96.54	99.06	97.50	68.05	

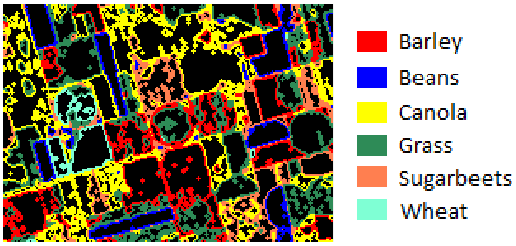


Figure 4.17: MRF (applied to boundaries) classified image showing the boundary pixels without the interior parts.

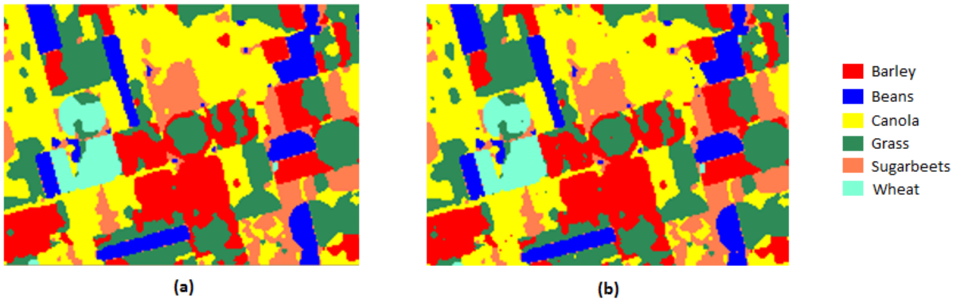


Figure 4.18: Visual comparison of MRF results: a) classified image using MRF on the entire image and b) classified image using MRF on the boundaries.

Table 4.10: Confusion matrix (pixels) of the MRF experiment on the entire image of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	Total
Barley	**529**	0	13	4	0	0	546
Bean	0	**291**	0	0	0	0	291
Canola	0	0	**363**	0	4	0	367
Grass	6	0	0	**530**	0	123	659
Sugarbeets	0	0	0	0	**156**	0	156
Wheat	0	0	0	0	0	**262**	262
Total	535	291	376	534	160	385	**2281**

Table 4.11: Confusion matrix (percent) of the MRF experiment on the entire image of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	User Acc. (%)
Barley	**98.88**	0.00	3.46	0.75	0.00	0.00	96.89
Bean	0.00	**100.00**	0.00	0.00	0.00	0.00	100.00
Canola	0.00	0.00	**96.54**	0.00	2.50	0.00	98.91
Grass	1.12	0.00	0.00	**99.25**	0.00	31.95	80.42
Sugarbeets	0.00	0.00	0.00	0.00	**97.50**	0.00	100.00
Wheat	0.00	0.00	0.00	0.00	0.00	**68.05**	100.00
Prod. Acc. (%)	98.88	100.00	96.54	99.25	97.50	68.05	

The statistical difference between the MRF method applied to the boundaries and the MRF method applied to the entire image is also evaluated using the McNemar's test [7]. It is a non-parametric test, which can be used to compare the accuracy of two classification algorithms and has been used in previous studies [7]. It is based on a 2×2 matrix with 4 possible outcomes in confusion matrices of two algorithms. The z score employed by the McNemar's test is calculated using the following formula:

$$z = \frac{(|N_{sf} - N_{fs}| - 1)}{\sqrt{N_{sf} + N_{fs}}}, \tag{4.1}$$

where, N_{sf} and N_{fs} are the numbers of pixels that one algorithm correctly classified and also the other algorithm failed to correctly classify. A value of $z = 0$ means the performance of two algorithms are similar and a larger value than +1.645 means that their performance differs significantly at a confidence level of 95 %.

Note that the z score is large in cases where one of N_{sf} or N_{fs} is close to zero, even if the other value is small. For example, if $N_{sf} = 5$ and $N_{fs} = 0$, the z score is 1.79, which indicates that the two algorithms are statistically different. The McNemar's test does not take into consideration whether $N_{sf} + N_{fs}$ represents a large portion of the sample data. Thus, this test may indicate a statistically significant difference even though the overall accuracies of the two algorithms are very close numerically.

The four possible results from the confusion matrices of the MRF method applied to the boundaries and to the entire image are shown in Table 4.12. According to the McNemar's test, the z score of the MRF method applied to the boundaries and the MRF method applied to the entire image is 3.47, which indicates that there is a significant difference between these algorithms even the difference of the overall accuracies is just 0.61 %.

Table 4.12: Results of the MRF method applied to boundaries and entire image (Method A = MRF on boundaries and Method B = MRF on entire image).

	Method A failed	**Method A succeeded**
Method B failed	150	0
Method B succeeded	14	2117

4.5 MRF on Randomly Selected Boundary Pixels

In this section, the MRF method is applied only on the boundary pixels of the SVM classification result. However, the energy minimization is performed over randomly selected pixels and randomly selected classes using Algorithm 2 in Figure 3.4. The result of this method is compared with the result of the MRF method applied to all boundary pixels (Algorithm 1 in Figure 3.3).

Different values of β and numbers of iterations (iteration l in Section 3.7) were tested in this method to determine the best values of these parameters, and the results are listed in Tables 4.13 and 4.14. Table 4.13 presents the overall accuracies (in percent) and run time (in seconds) for different values of β where the number of iterations is set to 20. Table 4.14 shows the overall accuracies (in percent) and run time (in seconds) for different numbers of iterations where $\beta = 0.80$. We found from these tables that increasing the values of β to more than 0.80 and increasing the number of iterations to more than 20 does not improve the overall accuracy. Moreover, we see that a large number of iterations increases the run time significantly. Consequently, we choose $\beta = 0.80$ and 20 iterations for our experiments in this section.

The result of the MRF method with energy minimization carried out on randomly selected boundary pixels is shown in Figure 4.19. The overall accuracy of this classification result is 91.02 % and the run time is 591 seconds. This method improves the accuracy of the SVM classification by 1.67 %. The confusion matrices of this classification are shown

Table 4.13: Examination of the MRF method (randomly selected pixels) applied to the boundaries of the SVM classification for different values of β.

Value of β	**Overall accuracy (%)**	**Run time (sec)**
1	90.84	696
0.90	90.66	680
0.80	91.02	591
0.70	89.43	653
0.60	87.33	603
0.50	87.42	579
0.40	86.19	457
0.30	85.18	238
0.20	83.95	512
0.10	83.65	475

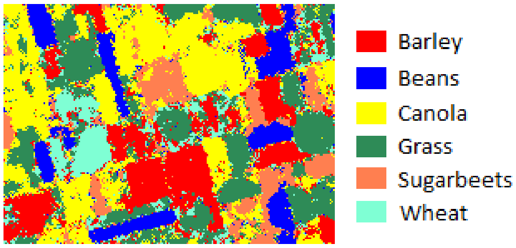


Figure 4.19: Result image using the MRF method on randomly selected boundary pixels.

in Tables 4.15 (in pixels) and 4.16 (in percentage along with the producer and user accuracies). These tables show that the accuracies of all classes including the wheat class are improved. Note that the MRF method applied on all boundary pixels and entire image (Section 4.4) could not improve the accuracy of the wheat class at all.

A visual comparison of the result from the MRF method applied to the randomly selected boundary pixels with the one from the MRF method applied to all boundary pixels is shown in Figure 4.20. It is obvious much of the salt and pepper noise are still present when the MRF method is only applied to the randomly selected boundary pixels. The accuracy of this classification is also 1.80 % lower than that of the MRF classification applied to

Table 4.14: Examination of the MRF method (randomly selected pixels) applied to the boundaries of the SVM classification for different number of iterations.

Number of iterations	**Overall accuracy (%)**	**Run time (sec)**
10	90.31	128
20	91.02	512
30	89.04	588
40	90.05	654
50	90.09	681
60	89.52	769
70	89.48	1023
80	90.53	874
90	89.65	995
100	90.84	1098

Table 4.15: Confusion matrix (pixels) of the MRF experiment applied to the randomly selected boundary pixels of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	Total
Barley	**497**	0	9	10	0	0	516
Bean	1	**291**	0	1	4	0	297
Canola	1	0	**364**	1	11	1	378
Grass	16	0	2	**515**	1	119	653
Sugarbeets	1	0	0	0	**144**	0	145
Wheat	19	0	1	7	0	**265**	292
Total	535	291	376	534	160	385	**2281**

all boundary pixels. The confusion matrices also show that the individual class accuracies are also higher than the method applied to randomly selected boundary pixels except for the wheat pixels. The z score of these two algorithms (Algorithm 1 and 2) is 4.39 which indicates that the classification accuracy using Algorithm 1 is significant. Moreover, the processing time of the MRF method applied to randomly selected boundary pixels is 591 seconds, which is 73 times longer than that of the MRF method applied to all boundary pixels where it takes only 8 seconds.

According to the above discussion we can conclude that the MRF method (Algorithm 1)

Table 4.16: Confusion matrix (percent) of the MRF experiment applied to the randomly selected boundary pixels of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	User Acc. (%)
Barley	**92.90**	0.00	2.39	1.87	0.00	0.00	96.32
Bean	0.19	**100.00**	0.00	0.19	2.50	0.00	97.98
Canola	0.19	0.00	**96.81**	0.19	6.88	0.26	96.30
Grass	2.99	0.00	0.53	**96.44**	0.63	30.91	78.87
Sugarbeets	0.19	0.00	0.00	0.00	**90.00**	0.00	99.31
Wheat	3.55	0.00	0.27	1.31	0.00	**68.83**	90.75
Prod. Acc. (%)	92.90	100.00	96.81	96.44	90.00	68.83	

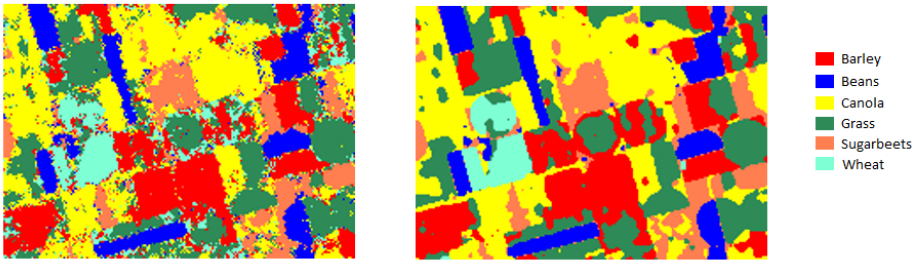


Figure 4.20: Visual comparison of the classification result of the MRF method applied to randomly selected boundary pixels (left) with the one of the MRF method applied to all boundary pixels (right).

applied to all boundary pixels is better than the MRF method applied to randomly selected boundary pixels (Algorithm 2) in terms of processing time and accuracy. For the MRF algorithm applied to randomly selected boundary pixels, it is unlikely to choose the right pixel and also the right class for correcting the label, because one pixel among all boundary pixels and also one class among all classes are randomly selected repeatedly. Sometimes, it might take many iterations to attempt to correct the proper pixel with the proper class, which is also time consuming.

4.6 UnECHO Experiment

The UnECHO method described in Section 3.6 is applied to the entire image of the SVM classification. It is tested for the following different types of distance measures: Euclidean distance (2.18), Mahalanobis distance (2.19), and Maximum likelihood distance (2.20). The overall accuracies of the UnECHO method for these distance measures are listed in Table 4.17. This method actually decreases the overall accuracy of the SVM classification for all distance measures. Among them, the choice of maximum likelihood distance decreases the accuracy (84.04 %) by 5.31 % (z score = 9.29), which is less than the other distance measures.

The result of the UnECHO method applied to the entire image of the SVM classification where the maximum likelihood distance is used, is shown in Figure 4.21 along with a visual comparison with the SVM classification result. We see that the result image of the UnECHO method is worse compared to the SVM classification. Most of the areas of the image are classified as canola (yellow). However, some of the misclassified sugarbeets and canola pixels using SVM are corrected by this method, but there are many pixels that remain misclassified. Figure 4.22 shows that many grass pixels are misclassified as canola, while misclassified barley pixels by SVM remain almost unchanged. However, some sug-

Table 4.17: Overall accuracies of the UnECHO experiment of Hyperion data for different distance measures and window sizes.

Name of distance measures and size of window	**Overall accuracy (%)**
Euclidean distance and 2×2 window size	82.24
Euclidean distance and 3×3 window size	82.46
Euclidean distance and 4×4 window size	82.81
Mahalanobis distance and 2×2 window size	81.76
Mahalanobis distance and 3×3 window size	81.24
Mahalanobis distance and 4×4 window size	82.90
Maximum Likelihood distance and 2×2 window size	83.91
Maximum Likelihood distance and 3×3 window size	84.04
Maximum Likelihood distance and 4×4 window size	83.34

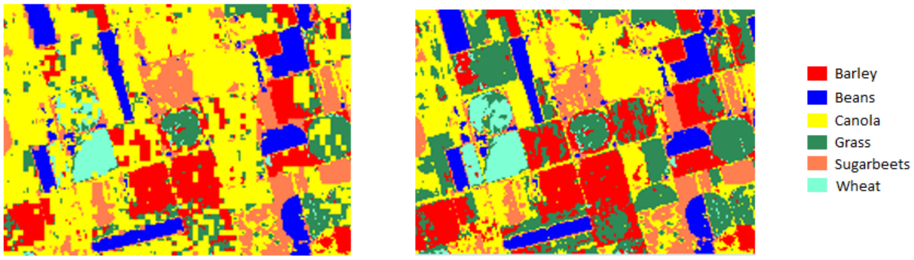


Figure 4.21: Visual comparison of the UnECHO classified image (left) and SVM classified image (right).

arbeets pixels misclassified by SVM (Figure 4.10) are corrected by the UnECHO method.

The confusion matrices, producer accuracies and user accuracies of the UnECHO experiment are shown in Tables 4.18 in pixels and 4.19 in percentages, respectively. These tables show that the accuracies of all classes decreased compared to using only SVM, except for the bean, canola and sugarbeets classes. All the bean pixels remain the same, while the accuracy of canola pixels and sugarbeet pixels is improved by 3.45 % and 6.25 %, respectively. The accuracy of grass pixels declined significantly (22.09 %) and, therefore, the overall accuracy is significantly decreased. The wheat region, which was misclassified as grass by the SVM is also misclassified as canola by the UnECHO method. This happens

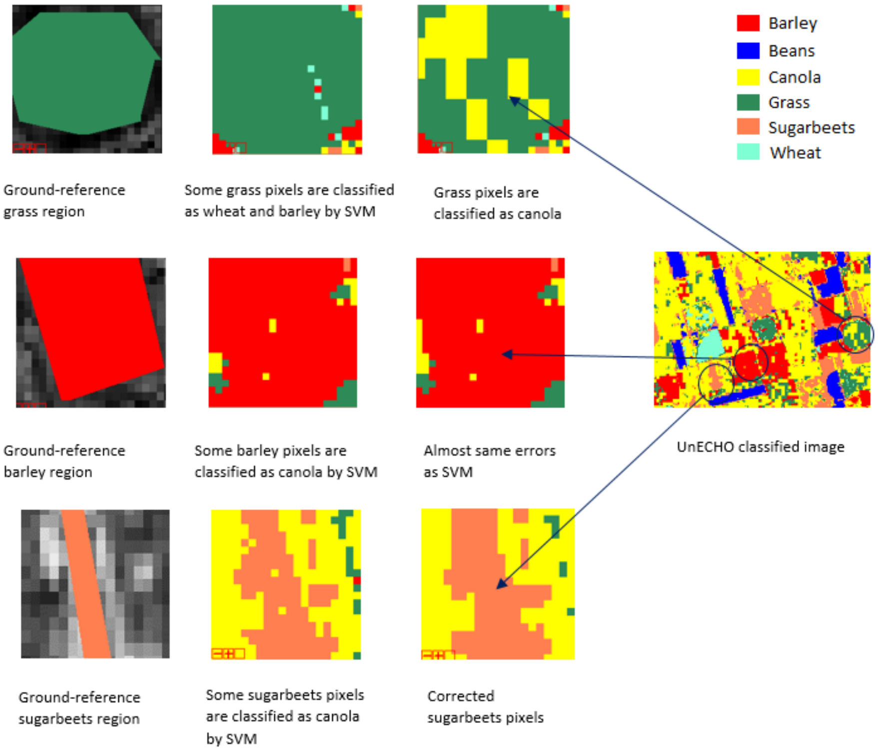


Figure 4.22: Example of corrected and uncorrected pixels by the UnECHO method.

Table 4.18: Confusion matrix (pixels) of the UnECHO experiment of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	Total
Barley	**467**	0	6	14	0	5	492
Bean	0	**291**	0	0	0	0	291
Canola	31	0	**370**	125	9	88	623
Grass	37	0	0	**378**	0	32	447
Sugarbeets	0	0	0	0	**151**	0	151
Wheat	0	0	0	17	0	**260**	277
Total	535	291	376	534	160	385	**2281**

Table 4.19: Confusion Matrix (Percent) of the UnECHO experiment of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	User Acc.
Barley	**87.29**	0.00	1.60	2.62	0.00	1.30	94.92
Bean	0.00	**100.00**	0.00	0.00	0.00	0.00	100.00
Canola	5.79	0.00	**98.40**	23.41	5.63	22.86	59.39
Grass	6.92	0.00	0.00	**70.79**	0.00	8.31	84.56
Sugarbeets	0.00	0.00	0.00	0.00	**94.38**	0.00	100.00
Wheat	0.00	0.00	0.00	3.18	0.00	**67.53**	93.86
Prod. Acc.	87.29	100.00	98.40	70.79	94.38	67.53	

because the amount of differences between the spectral signatures of wheat and grass is slightly higher than that between the spectral signatures of grass and canola (for example in the 700 nm–1100 nm range), and the UnECHO method chose the canola class instead of wheat class to correct the grass pixels (Figure 4.3).

4.7 Combination of MRF and UnECHO Results

In this section, labels extracted from the boundary pixels from the MRF classification result were combined with labels extracted from the interior pixels from the UnECHO classification to produce the final classification map. Using the UnECHO method, the whole image is divided into 3×3 windows, and each window may be classified as one class depending on the distance from the mean of each class and a computed threshold. For this reason, it

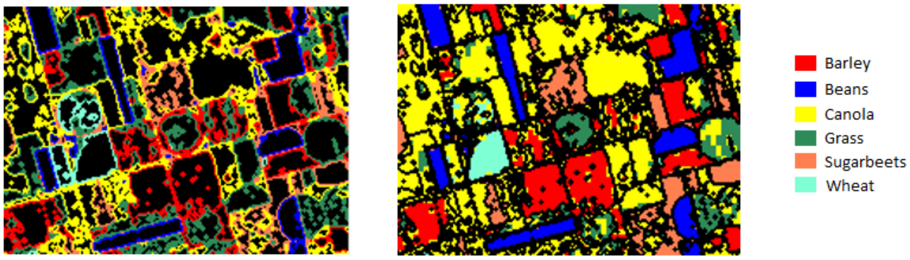


Figure 4.23: Boundary pixels of the MRF (left) and interior pixels of the UnECHO classification results (right).

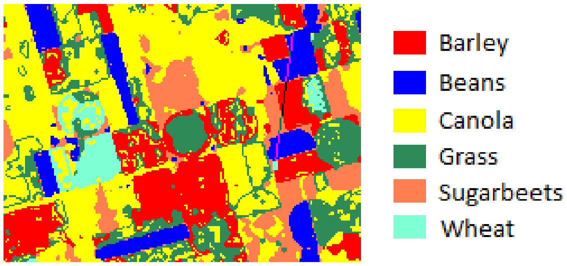


Figure 4.24: Combined image of the MRF and UnECHO results.

is not possible to apply the UnECHO method only on the interior parts of the SVM classification. However, we can extract the labels (obtained from the result of the UnECHO method) corresponding to the interior pixels and combine them with the labels attained on the boundary pixels using the MRF method. The boundary pixels of the MRF classification and the interior pixels of the UnECHO classified image are shown in Figure 4.23. The final result image is shown in Figure 4.24 and the accuracy is 87.59 %, which is 1.76 % lower (z score = 2.67) compared to the SVM classification. Although the MRF method improved the accuracy of the SVM classification, the UnECHO method decreased the accuracy significantly because of the grass pixels. For this reason, the accuracy of the final result decreased. The confusion matrices of the final result are shown in Tables 4.20 in pixels and 4.21 in percentages, respectively.

Table 4.20: Confusion matrix (pixels) of the final result of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	Total
Barley	**500**	0	9	3	0	0	512
Bean	0	**291**	0	0	0	0	291
Canola	16	0	**367**	105	4	70	562
Grass	19	0	0	**424**	0	55	498
Sugarbeets	0	0	0	0	**156**	0	156
Wheat	0	0	0	2	0	**260**	262
Total	535	291	376	534	160	385	**2281**

Table 4.21: Confusion matrix (percent) of the final result of Hyperion data.

	Barley	Bean	Canola	Grass	Sugarbeets	Wheat	User Acc.
Barley	**93.46**	0.00	2.39	0.56	0.00	0.00	97.66
Bean	0.00	**100.00**	0.00	0.00	0.00	0.00	100.00
Canola	2.99	0.00	**97.61**	19.66	2.50	18.18	65.30
Grass	3.55	0.00	0.00	**79.40**	0.00	14.29	85.14
Sugarbeets	0.00	0.00	0.00	0.00	**97.50**	0.00	100.00
Wheat	0.00	0.00	0.00	0.37	0.00	**67.53**	99.24
Prod. Acc.	93.46	100.00	97.61	79.40	97.50	67.53	

4.8 Experiments using Principal Component Analysis (PCA)

The Principal Component Analysis method is tested in this section as a band extraction method. As described in Section 2.6.1, PCs are uncorrelated. The Hyperion dataset containing the most important PCs is used as the input to the SVM classification to study the effects of the number of PCs on overall accuracies. The MRF and UnECHO methods are also applied to the SVM classification results derived from these PCs. Table 4.22 lists the overall accuracies of all these experiments for different number of PCs. The overall accuracies of the SVM classification is more than 99 % for all numbers of extracted PCs listed in the table. The wheat region, which was classified as grass by the SVM method using original all the bands, is now classified correctly by the SVM method using the PCs.

Figure 4.25 shows a visual comparison of the SVM and MRF classifications using 10 PCs against the ground reference ROI regions. All the regions of the ground reference ROIs are classified almost correctly by the SVM method, and the very few errors in the SVM classification are corrected by the MRF method. The PC band extraction approach improved the SVM results, because there is less redundancy of information among PCs and the first few PCs contain more than 90 % of the data variance. The rest of the components contain insignificant information and noise. Figure 4.26 shows the coefficient values of PCs for each ground truth crop type. As the values of PCA are very different for each crop type for the first few PCs, it is easy to distinguish the classes. The MRF results are the same for all numbers of PCs listed in Table 4.22 with 100 % accuracy giving slight improvements to the SVM classification. The UnECHO method performed worse than the other methods for all numbers of PCs as shown in Table 4.22.

Table 4.22: List of overall accuracies obtained by the SVM, MRF, and UnECHO methods using Hyperion dataset with different numbers of PCs.

Number of PCs	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	99.17	100.00	84.00
20	99.74	100.00	84.61
30	99.87	100.00	85.80
40	99.87	100.00	84.66
50	99.74	100.00	84.52
60	99.74	100.00	84.52
70	99.78	100.00	84.52
80	99.82	100.00	84.66
90	99.82	100.00	83.87
100	99.78	100.00	84.61
110	99.78	100.00	83.82
120	99.78	100.00	84.48
130	99.74	100.00	84.44
140	99.74	100.00	83.69
150	99.74	100.00	83.38

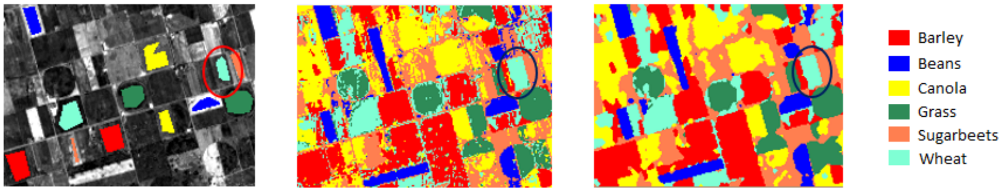


Figure 4.25: Visual comparison of SVM (middle) and MRF classification results (right) using 10 PCs with ground-reference ROIs (left).

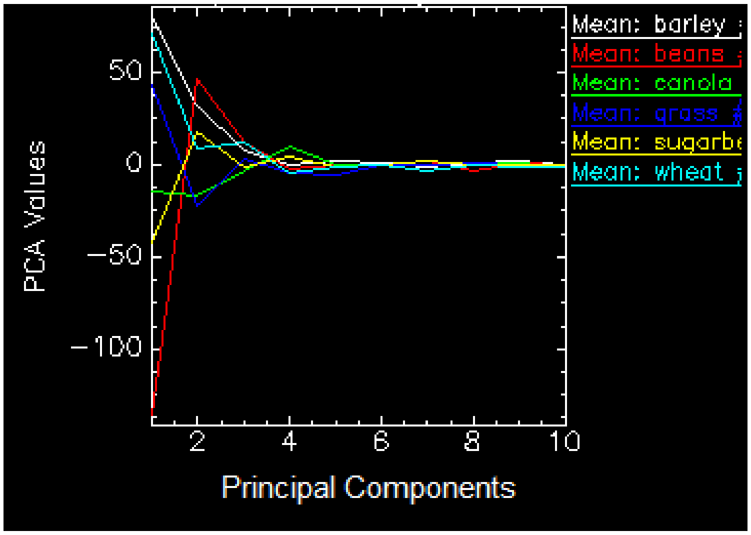


Figure 4.26: PC Coefficients for each ground-reference crop type.

We also experimented the SVM, MRF and UnECHO method for 4, 5, and 6 PCs, because the coefficient values of PCA are mostly different up to these numbers of PCs (Figure 4.26). The overall accuracies of these experiments are listed in Table 4.23. The improvement of the accuracy of the SVM and MRF method is only slightly better using the 4 PCs compared to the SVM and MRF results using original all bands. Although the accuracy of the SVM and MRF using 5 and 6 PCs is significant, the highest accuracy is achieved using all numbers of PCs starting from 10 as shown in Table 4.22.

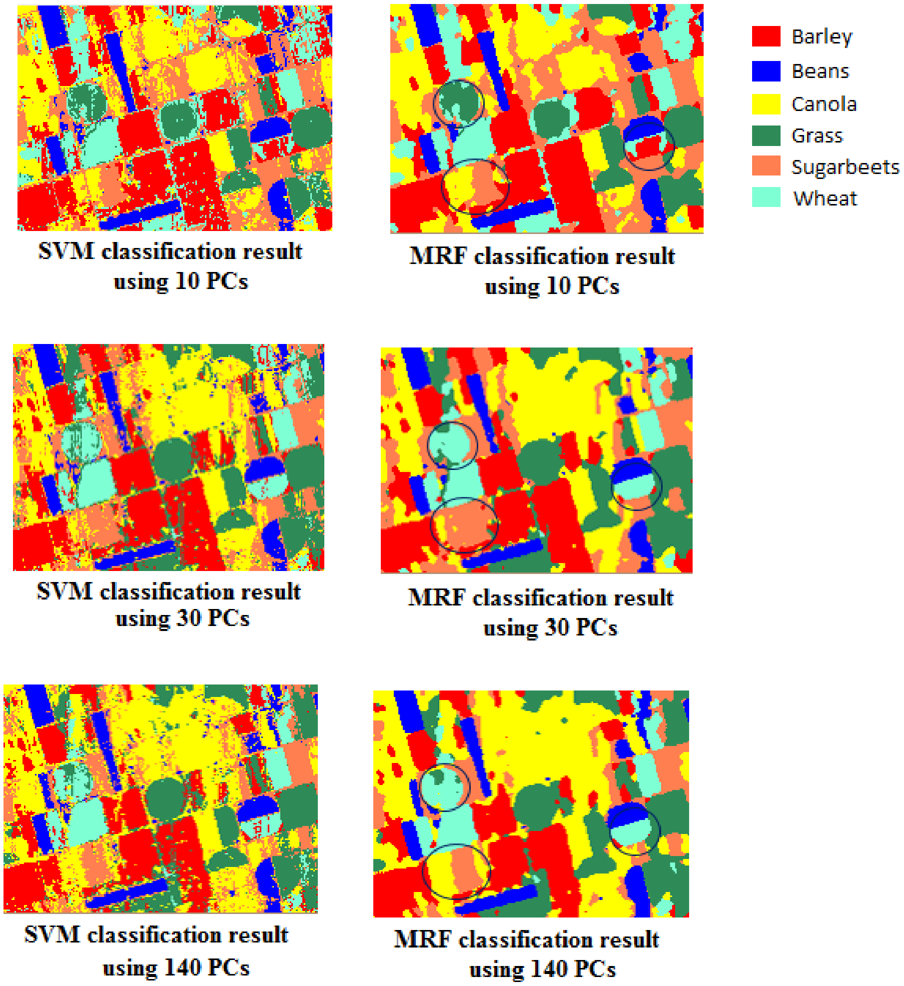
Table 4.23: List of overall accuracies obtained by the SVM, MRF, and UnECHO methods for the Hyperion dataset using 4, 5 and 6 PCs.

Number of PCs	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
4	91.32	93.47	81.37
5	97.94	99.82	82.90
6	98.07	99.78	83.12

The overall accuracies of the SVM and MRF methods are improved significantly using different number of PCs compared to using all the bands (e.g., z score = 14.52 of the SVM method using 10 PCs and all bands). Even though the overall accuracies of the SVM and MRF methods using different number of PCs are statistically similar (e.g., z score = 0.89 between the SVM methods using 20 PCs and 30 PCs, which indicates a insignificant difference) compared to each other using available ground-reference information, visual comparison of individual crop fields of the entire images shows better accuracy improvement using 10 PCs. When 20 or more PCs are used, many regions are classified as one crop type (i.e., canola). Figure 4.27 shows the SVM and MRF classification results using 10, 30, and 140 PCs. We see that classifications of SVM and MRF using the 10 PCs are more reasonable than those using 30 and 140 PCs, because result images derived from more than 20 PCs consist of an increasing area of canola regions. In the MRF classification using the 140 PCs, it is clear that several canola fields are contiguous with each other and there is no boundary between these crop fields, which is not desirable. In contrast, some separate regions are more homogeneous outside the ground-reference ROIs in the MRF classification using 140 PCs than in the MRF image using the 10 PCs (shown in circles in the MRF images of Figure 4.27), which may be correct. Therefore, it is not clear which one is true due to the lack of ground reference data for those crop fields. However, based on the general knowledge about individual fields, the results using the 10 PCs are better, because the fields in the result images are separate, homogeneous and also have good accuracy with available ground-reference.

4.9 Experiments using Independent Component Analysis (ICA)

The Independent Component Analysis (ICA) is tested on the Hyperion dataset. The overall accuracies of the SVM, MRF, and UnECHO experiments using different number of Independent Component (IC) are reported in Table 4.24. The overall accuracies of the SVM and MRF experiments using different number of ICs show similar trends as the PCA (i.e., z score = -0.71 between the SVM methods using 20 PCs and 20 ICs). However, the overall accuracies of the UnECHO results for some different number of ICs are better compared to using the PCs (Section 4.8). Like PCA, all ground-reference regions including the wheat region are classified quite correctly by the SVM method with very few salt and pepper noise. The overall accuracies of the MRF methods are 100 % for almost all number of ICs with insignificant improvement of the SVM results based on the ground-reference regions. However, the SVM and MRF classification results using 10 ICs are visually better, because most of the regions are either canola or barley in the classification results using more than 10 ICs. The classification results of SVM and MRF using 10, 40 and 120 ICs are pre-



SVM classification result using 10 PCs

MRF classification result using 10 PCs

SVM classification result using 30 PCs

MRF classification result using 30 PCs

SVM classification result using 140 PCs

MRF classification result using 140 PCs

Figure 4.27: The SVM and MRF classification results using 10, 30, and 140 PCs.

sented in Figure 4.28. The classification results for 10 ICs are more realistic than the others because most of the regions should not be only barley or canola without any boundaries between crop fields like the images using 40 or 120 ICs.

Table 4.24: List of overall accuracies obtained by the SVM, MRF, UnECHO methods for Hyperion dataset with different numbers of ICs.

Number of ICs	SVM (%)	MRF (%)	UnECHO (%)
10	99.39	99.91	82.55
20	99.74	100.00	93.34
30	99.74	99.96	87.29
40	99.82	100.00	85.14
50	99.74	100.00	85.93
60	99.74	100.00	94.30
70	99.78	100.00	89.08
80	99.82	100.00	86.85
90	99.82	100.00	87.64
100	99.87	100.00	88.95
110	99.82	100.00	84.92
120	99.74	100.00	84.17
130	99.87	100.00	88.69
140	99.87	100.00	94.34
150	99.82	100.00	93.47

4.10 Experiments using SVM-RFE Bands

In this section, the SVM-RFE method, which was described in Section 2.6.2, is tested using the Hyperion dataset as a feature selection technique. The SVM is trained with 160 bands and 6 classes with 10 % randomly selected pixels. In each iteration (150 iterations for Hyperion data) one band is eliminated according to the ranking score. The weight vector of SVM is used as a ranking criterion, and the band with the smallest ranking score is eliminated. After eliminating each band, the SVM is trained again with the selected bands

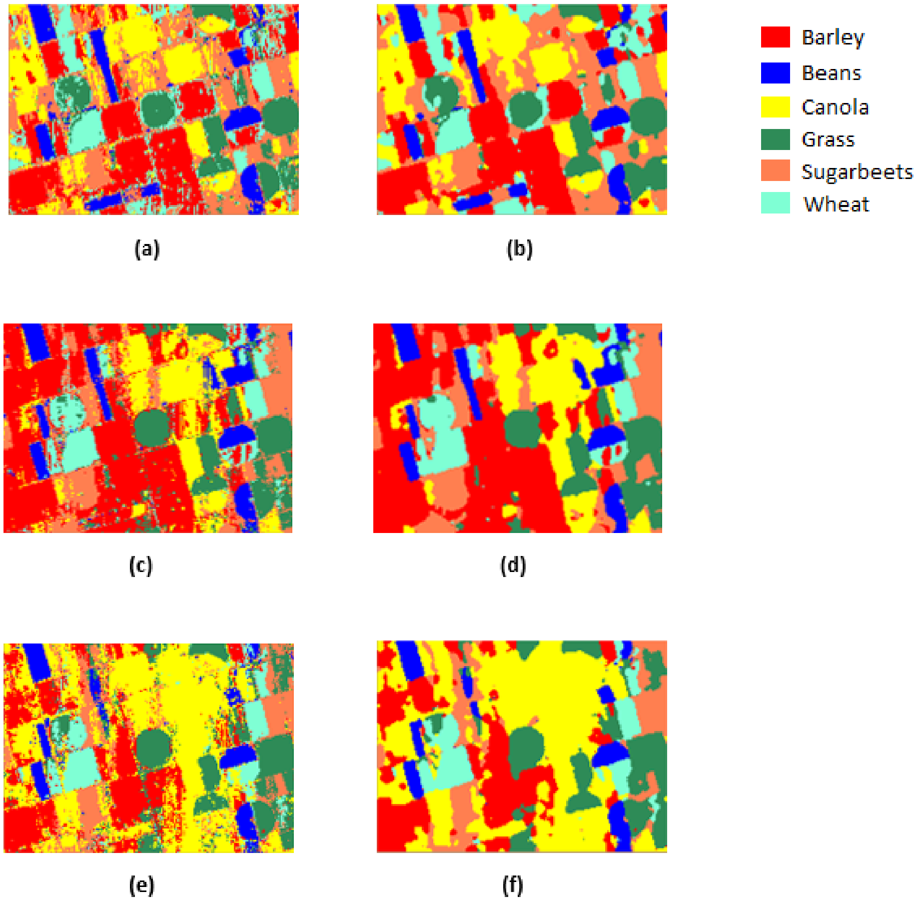


Figure 4.28: The SVM and MRF result images: (a) SVM classification result using 10 ICs, (b) MRF classification result using 10 ICs, (c) SVM classification result using 40 ICs, (d) MRF classification result using 40 ICs, (e) SVM classification result using 120 ICs, and (f) MRF classification result using 120 ICs.

containing pixel values and classes and, thus, a subset of best selected bands is obtained.

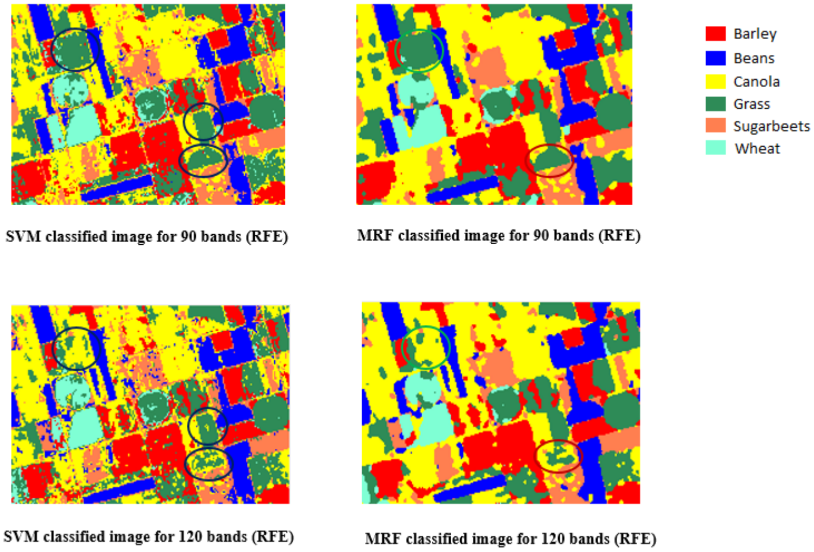
The Hyperion dataset with different numbers of selected bands acquired from the SVM-RFE algorithm is tested for the experiments. Each of the SVM, MRF and UnECHO methods were applied on the dataset containing different number of selected bands (10, 20, ... ,150). The resulting overall accuracies are shown in Table 4.25. We know from the SVM experiment (Section 4.3) that the accuracy of the SVM using 160 bands is 89.35 %. From the table, we see that at first after eliminating 10 bands (with 150 bands) the accuracy of SVM is improved slightly by 0.13 %. The accuracies of SVM are higher for 60–150 bands but lower for 10–40 bands. The best accuracy of SVM is achieved using 90 selected bands and the improvement is 0.39 % (z score = 0.40) compared to the SVM using all the bands. This result is improved to 92.99 % by the MRF method. Therefore, the accuracy is improved by 0.18 % compared to the accuracy (92.81 %) of the MRF result using the all bands. The best accuracy of 93.2% of the MRF method is attained using 120 selected bands. The SVM and MRF classification results using the 90 and 120 selected bands are shown in Figure 4.29. Although the accuracy of the MRF classification using 120 selected bands is the highest, the SVM classification result using 90 selected bands is more homogeneous, and contain less salt and pepper noise than the SVM classification result using the 120 selected bands (shown in circles) if we look at the outside of the ground reference regions. That means reducing the number of bands decreases some of these classification errors accordingly. In addition, there is no significant numerical difference between the classification results using the 90 and 120 selected bands. However, there are some notable qualitative differences especially in the outside of ground-reference ROIs of the MRF images. For example, one region (shown in the green circle in the MRF images) is classified as grass using the 90 band set. On the other hand, this region is classified as canola using the 120 band set. Because of the unavailability of the ground-reference of this region, it is hard to say which one is correct. However, some fields are completely homogeneous in the

MRF image using the 90 selected bands compared to the MRF image using 120 selected bands (shown in the red circle in the MRF images). Because these data cover agricultural areas, it is acceptable to be classified as homogeneous areas. Moreover, the McNemar's test (z score = 6.25 between the SVM methods using 90 and 120 selected bands) also shows a significant difference.

Table 4.25: Overall accuracies of the SVM, MRF, and UnECHO methods obtained from the Hyperion dataset with SVM-RFE selected bands.

Number of bands	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	70.19	75.19	71.77
20	73.04	77.86	70.45
30	80.49	85.53	73.69
40	86.32	89.08	75.49
50	88.21	90.36	78.61
60	89.65	92.42	81.76
70	89.22	92.55	80.93
80	89.52	92.85	82.20
90	89.74	92.99	82.03
100	89.61	92.99	81.46
110	89.57	92.94	81.94
120	89.61	93.20	82.33
130	89.47	93.03	82.68
140	89.61	92.72	84.13
150	89.48	92.68	84.88

For the case of the UnECHO method all the listed accuracies are lower than the accuracy of the experiment with 160 bands except for 140 and 150 selected bands. The accuracy using the 150 band set is 84.88 % which is a 0.84 % improvement compared to the result with 160 band set. Again the UnECHO experiment could not improve the SVM classification results.



SVM classified image for 90 bands (RFE)

MRF classified image for 90 bands (RFE)

SVM classified image for 120 bands (RFE)

MRF classified image for 120 bands (RFE)

Figure 4.29: The SVM and MRF result images for the 90 and 120 bands selected by SVM-RFE.

4.11 Experiments using CFS Bands

The Correlation-based Feature Selection (CFS) technique is applied to select a best band subset from the 160 bands of Hyperion data. In the CFS algorithm, features or bands are selected based on the correlation with the class and the other bands as described in Section 2.6.2. Noisy and redundant bands are identified and removed, and the remaining subset of bands are correlated with the class, but not correlated with each other. The SVM method is applied to the resulting band subset selected by the CFS algorithm from the Hyperion dataset. Afterwards, the MRF and UnECHO methods are applied on the SVM classification results.

The CFS algorithm selected 18 bands from the Hyperion dataset. The overall accuracies obtained from the SVM, MRF, and UnECHO methods are presented in Table 4.26. The overall accuracies of the SVM and MRF classifications are improved by 1.97 % (z score = 4.00) and 1.31 % (z score = 2.82), respectively, compared to the SVM and MRF classifications using all bands. The accuracy of UnECHO classification is not improved compared to the SVM and MRF classifications.

Table 4.26: Overall accuracies obtained from the the SVM, MRF, UnECHO methods for the Hyperion dataset with the CFS selected bands.

SVM (%)	**MRF (%)**	**UnECHO (%)**
91.32	94.13	84.00

The result images of the SVM and MRF classifications derived from the Hyperion dataset with CFS selected bands are shown in Figure 4.30 along with a visual comparison of the SVM and MRF classifications without band selection. Some regions are classified as barley and grass by the SVM classification when all bands are considered. However, some pixels in those regions are classified as wheat by the SVM classification using the CFS selected bands. The wheat region, which was totally misclassified as grass by the SVM is partly classified correctly when the CFS band set is used. The number of misclassified pixels is less when CFS is used. The MRF method corrected these misclassified pixels and improved the accuracy. Therefore, noisy and redundant bands are removed by the CFS method and, hence, experimental results provided by the SVM and MRF methods for the Hyperion dataset using CFS, show better accuracies.

4.12 Experiments using mRMR Bands

In this section, we apply the SVM, MRF and UnECHO methods on the Hyperion dataset with the selected bands provided by the Minimum-Redundancy–Maximum-Relevance (mRMR) feature selection method (Section 2.6.2).

The overall accuracies for different number of mRMR selected bands are shown in Table 4.27. The accuracies of the SVM, MRF and UnECHO classification results are not

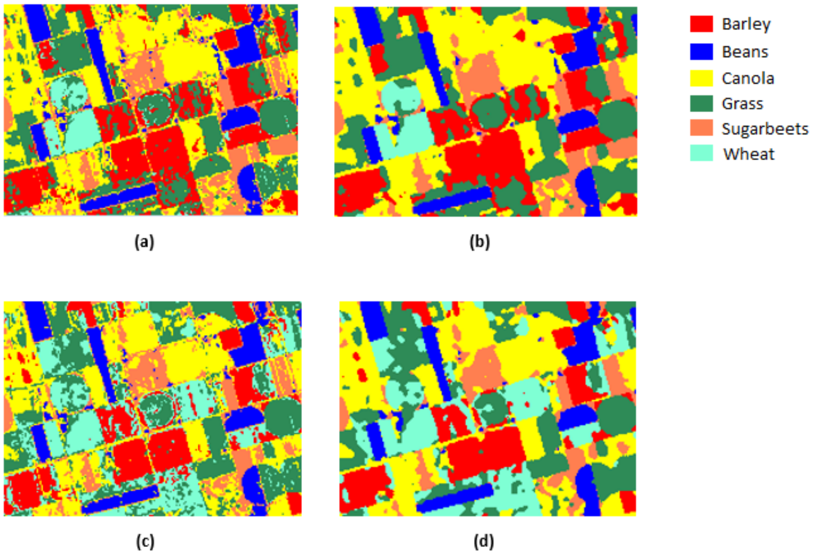


Figure 4.30: The SVM and MRF classification results: (a) SVM result image for 160 bands, (b) MRF result image for 160 bands, (c) SVM result image for CFS selected bands, and (d) MRF result image for CFS selected bands.

improved for almost all the cases when mRMR is used for band selection purposes. Among the listed experiments, only the accuracy of the SVM classification using 150 mRMR selected bands is improved slightly (0.04 %) compared to the SVM classification using all the bands (z score = 0), which is negligible. The accuracy of the subsequent MRF classification is lower than the accuracy of the MRF classification using all bands. The accuracies of the SVM and MRF classifications using the 150 mRMR selected bands are the highest among all number of bands listed in the table. The second highest accuracy of MRF method is obtained using the 20 mRMR selected bands. The classification results of SVM and MRF using both 20 and 150 mRMR selected bands in Figure 4.31, show that many regions are classified as barley using the 20 mRMR selected bands. Many regions in the SVM and MRF classification results using 20 mRMR selected bands are classified as one type of homogeneous crop field (shown in circles). Visually these regions remain distinct in the SVM and MRF classification results using 150 mRMR selected bands. On the other hand, the numerical classification accuracy using 150 mRMR selected bands are similar to the classification results using all the bands (Figure 4.6). Based on our knowledge of the scene and accuracy, we can conclude that the best result is attained for the SVM with 150 mRMR selected bands.

Table 4.27: List of overall accuracies obtained by the SVM, MRF, and UnECHO methods for Hyperion dataset using different number of mRMR bands.

Number of mRMR Bands	SVM (%)	MRF (%)	UnECHO (%)
10	86.54	90.39	74.09
20	87.06	92.15	74.44
30	83.99	88.20	71.15
40	81.06	83.29	68.47
50	86.01	90.97	73.70
60	85.97	91.31	73.47
70	84.74	90.75	72.03
80	84.39	89.96	72.07
90	85.92	88.95	81.37
100	86.27	90.66	81.14
110	87.11	91.58	81.24
120	87.81	91.05	83.03
130	87.46	90.57	80.97
140	88.68	92.10	83.60
150	89.39	92.72	84.26

4.13 Summary of Experiments without Feature Reduction

Table 4.28 shows the best accuracies obtained using the SVM, MRF, UnECHO and proposed methods. After the SVM experiment we see that there are many pixels misclassified, and salt and pepper noise is present in the classification result. To reduce these noise and improve the accuracy, the MRF and UnECHO methods are chosen as spatial classification techniques. Boundary and interior pixels of the SVM classification result are separated us-

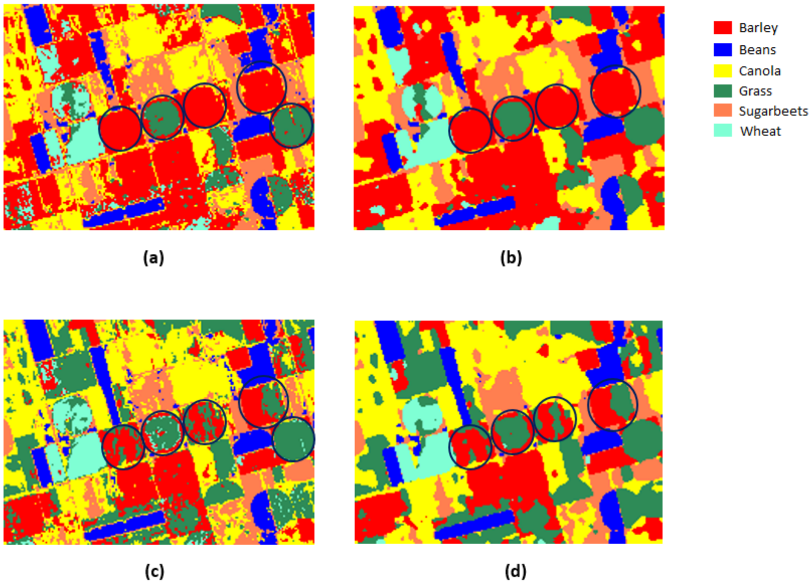


Figure 4.31: The SVM and MRF classification results using mRMR selected bands, (a) SVM classification result using 20 bands, (b) MRF classification result using 20 bands, (c) SVM classification result using 150 bands, and (d) MRF classification result using 150 bands.

ing the erosion technique. We tried two algorithms of the MRF method where Algorithm 1 computes the local energy over all boundary pixels and Algorithm 2 computes the local energy over randomly selected boundary pixels. We conclude that Algorithm 2 actually resulted in lower accuracy and higher computation time compared to the Algorithm 1. The highest accuracy acquired from the UnECHO method by using the maximum likelihood distance measure is lower and reduces the accuracy of the SVM classification significantly even though some of the misclassified pixels are corrected. Finally, the accuracy of the combination of the MRF method on the boundaries and the UnECHO method on the interior pixels did not improve the SVM classification.

To conclude about the experiments using the Hyperion dataset without feature selection or extraction, we can say that according to the accuracy and computation time, SVM+MRF (Algorithm 1) for only boundary pixels is the best combination compared to the other classification approaches.

Table 4.28: Overall accuracies obtained by SVM, MRF, and UnECHO experiments using Hyperion dataset without Feature Reduction.

SVM (%)	**SVM + MRF (%)**	**SVM + UnECHO (%)**	**SVM + MRF + UnECHO (%)**
89.35	92.81	84.04	87.59

4.14 Summary of Experiments with Feature Reduction

The accuracies of the best results achieved from experiments after applying different individual feature reduction techniques are summarized in Table 4.29 including the number of component or band subsets used for the classification. These results are taken considering not only the accuracy but also the visual examination.

The accuracies improved very significantly using the SVM and MRF methods after applying both feature extraction techniques, PCA and ICA. The accuracies of the SVM and MRF classification results are 99.17 % and 100.00 %, respectively using 10 PCs. The accuracies of the SVM and MRF classification results are 99.39 % and 99.91 %, respectively using 10 ICs. We concluded after visual comparison that the results using both 10 PCs and 10 ICs are the best.

Among all feature selection techniques, CFS with 18 bands provided the best accuracy. The improvements of the accuracies of the SVM and MRF methods are not significant when the SVM-RFE technique is used with 90 bands compared to using all bands. The improvement of the SVM classification is very insignificant using 150 mRMR selected bands compared to the SVM using all the bands. There is no improvement of the MRF method using 150 mRMR selected bands compared to the MRF method using all the bands.

The UnECHO method could not improve the SVM result for any of the feature reduction techniques and also for any of the number of bands. Finally, we can say that for the Hyperion dataset, PCA and ICA feature extraction techniques provided the best results among all feature reduction techniques.

Table 4.29: List of overall accuracies obtained by SVM, MRF, UnECHO experiments for Hyperion dataset with Feature Reduction.

Feature Reduction Techniques	**SVM (%)**	**SVM + MRF (%)**	**SVM + UnECHO (%)**
PCA (10 bands)	99.17	100.00	84.00
ICA (10 bands)	99.39	99.91	82.55
SVM-RFE (90 bands)	89.74	92.99	82.03
CFS (18 bands)	91.32	94.13	84.00
mRMR (150 bands)	89.39	92.72	84.26

Chapter 5

Experiments for Other Datasets

5.1 AVIRIS Dataset

Another agricultural hyperspectral dataset used in this thesis is an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) dataset collected over northwest Indiana's Indian Pines test site in June 1992 [32, 33, 37, 40]. This dataset has been widely used for classification studies in the remote sensing community, because the scene with ground-reference information is publicly available from the Purdue University's MultiSpec site [37]. The scene is a subset of a larger image with 145×145 pixels and 200 spectral bands between 0.4 and 2.45 μm with 20-m spatial resolution and calibrated radiance values [38]. A total of 20 bands (104–108, 150–163, 220) are excluded from the original 220 bands due to noise and water absorption phenomena [33]. The scene mostly contains agricultural areas where crops (e.g., corn and soybeans) were in early growth stage with few forested areas and natural perennial vegetation, and very few impervious surfaces or infrastructure. The ground-reference data of the scene contain 16 classes with a total of 10,366 labeled samples.

5.2 Experiments for the AVIRIS Dataset without Feature Reduction

In this section, experiments are conducted with the SVM, MRF and UnECHO classifiers on the Indian Pines dataset without applying any feature or band reduction techniques. The Indian Pines scene with false-color composite and the scene with the 16 classes (all

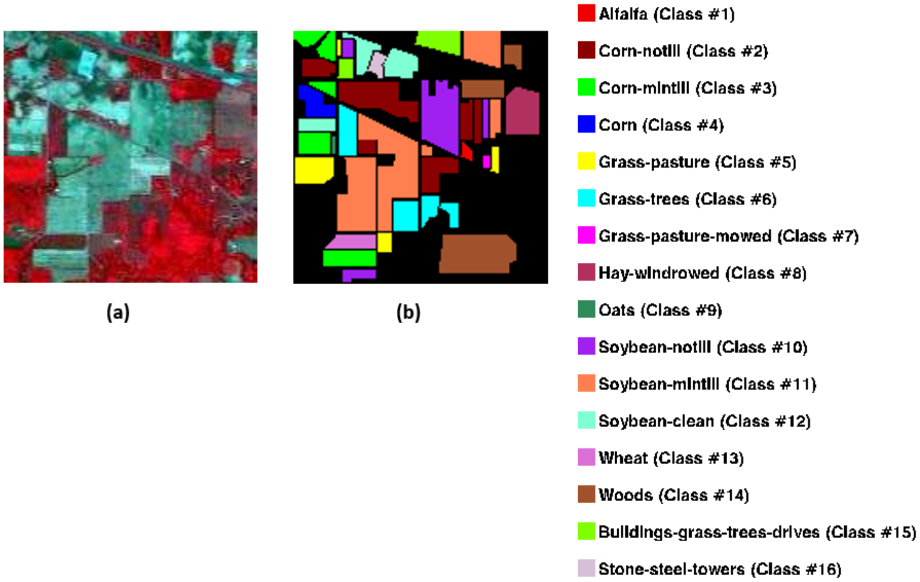

(a) (b)

Figure 5.1: AVIRIS Indian Pines dataset [32]; (a) false-color composite (R: NIR, G: Red, B: Blue), and (b) ground-reference of 16 classes.

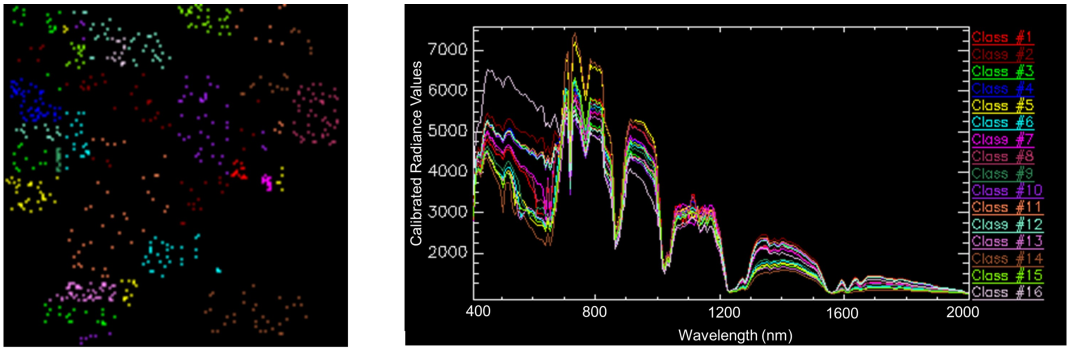
classes) ground-reference data is shown in Figure 5.1. The class names of the 16 classes, the corresponding colour of these classes, the total number of ground-reference pixels for every class and the total number of training pixels selected for each class are reported in Table 5.1. The training pixels are selected randomly and the number of training pixels for each class are the same as in a previous study [33]. The accuracy assessment of all classification is done with the remaining pixels of the ground-reference dataset.

The randomly selected 16-class training samples and their corresponding spectral signatures used to train the SVM are shown in Figure 5.2. The overall accuracy of the SVM experiment is 62.54 % when all default parameters values are used (Section 4.3). Because the accuracy is not high, different values of the penalty parameter C and kernel parameter γ are tried to improve the overall classification accuracy of the SVM experiment. A common method to obtain good parameter values is a grid search [25]. Table 5.2 shows that increasing the value of C increases the overall accuracy up to the value of $C = 2000$ and

Table 5.1: Ground-reference classes for the Indian Pines scene and their respective total number of pixels, number of training samples and associated colour.

Class No.	Class Name	Colour	# of pixels	# of training pixels
Class 1	Alfalfa	Red 1	46	15
Class 2	Corn-notill	Red 2	1428	50
Class 3	Corn-mintill	Green	830	50
Class 4	Corn	Blue	237	50
Class 5	Grass-pasture	Yellow	483	50
Class 6	Grass-trees	Cyan	730	50
Class 7	Grass-pasture-mowed	Magenta	28	15
Class 8	Hay-windrowed	Maroon	478	50
Class 9	Oats	Sea Green	20	15
Class 10	Soybean-notill	Purple	972	50
Class 11	Soybean-mintill	Coral	2455	50
Class 12	Soybean-clean	Aquamarine	593	50
Class 13	Wheat	Orchid	205	50
Class 14	Woods	Sienna	1265	50
Class 15	Buildings-Grass-Trees-Drives	Chartreuse	386	50
Class 16	Stone-Steel-Towers	Thistle	93	15

after that the overall accuracy started decreasing. We also selected different values of γ and found that changing γ value for any particular C value does not affect the overall accuracy at all. Therefore, we found that the value of $C = 2000$ and the default γ value (i.e., 1/total number of bands) are the best parameters values for this dataset. Accordingly the overall accuracy of the SVM experiment is 72.84 % with an improvement of 10.30 % using the value of $C = 2000$. The overall accuracy is not particularly high but the improvement is significant (z score = 22.39). If we look at the confusion matrix in Table 5.3, we see that the producer accuracies of most of the classes (10 classes) are above 80 %. Among these, Hay-windrowed (class 8) obtained the highest accuracy (99.07 %). The producer accuracies of only 2 classes are less than 60 % and the class with the least producer accuracy (51.43 %) is soybean-mintill (class 11). The spectral signatures of the classes with poor producer accuracies are very similar to the spectral signatures of other classes. For example, the



(a) **(b)**

Figure 5.2: AVIRIS Indian Pines data set; (a) randomly selected training samples and (b) spectral signature of the 16 classes.

soybean-mintill (class 11) is mostly labeled as Corn-mintill (class 3) and Soybean-notill (class 10), because the spectral signatures of these three classes are very similar as shown in Figure 5.3. Figure 5.4 shows the result image of the SVM experiment along with the ground reference image. It is noticeable that a significant amount of salt and pepper noise are present in the result image compared to the ground-reference image even with good producer accuracies of the individual classes.

After the SVM classification, the MRF method is applied on the classification of the SVM. We applied the MRF (Algorithm 1) just on the boundaries of the SVM classification in this section after extracting the boundary pixels using the erosion technique just as we did in Section 4.4 for the Hyperion dataset. For this experiment, 5 iterations are used and the value of the β parameter is set to 0.80 as this value and this number of iteration showed the best performance based on the accuracy and execution time.

The overall accuracy of the MRF experiment is 83.34 % which improved the SVM result by 10.50 %. This is a significant improvement (z score = 26.28). The individual class accuracies are also improved as shown in the confusion matrix of Table 5.4. The highest accuracy improvement is 29.04% for Alfalfas (class 1). In Figure 5.5, the MRF classified

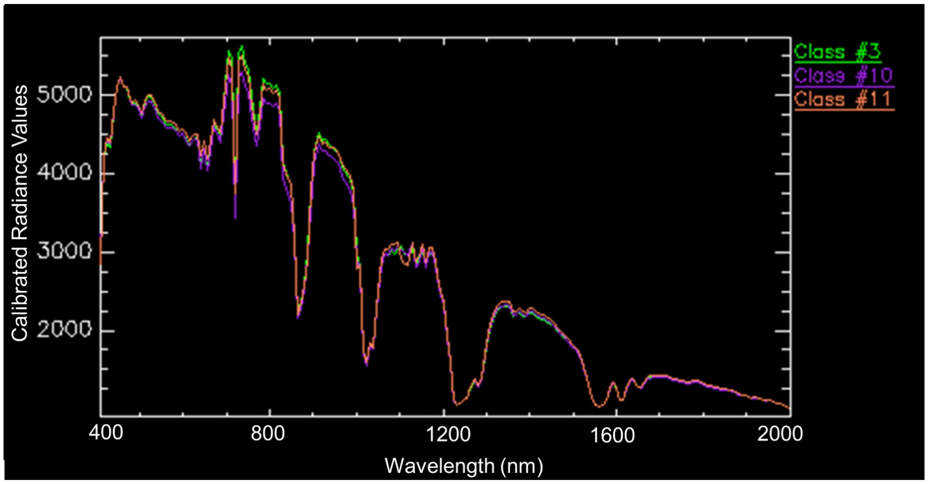


Figure 5.3: Spectral signatures of classes 3, 10 and 11.

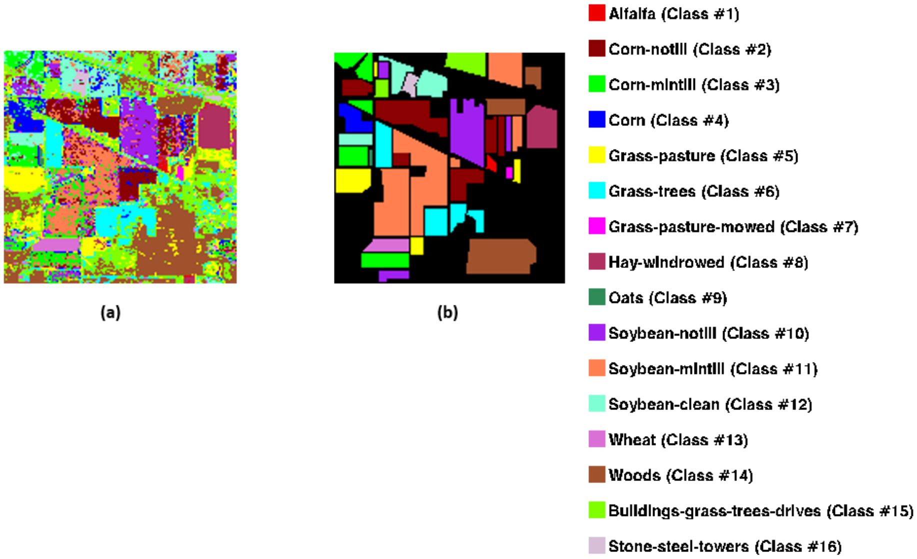


Figure 5.4: (a) The SVM result image of AVIRIS Indian Pines data set, and (b) ground-reference image of AVIRIS Indian Pines dataset.

Table 5.2: Overall accuracies of the SVM experiment for different values of the penalty parameter C.

Value of C	**Overall accuracy (%)**
20	53.90
60	59.71
100	62.54
200	67.37
400	70.66
700	71.67
1000	72.07
1500	72.44
1700	72.72
2000	72.84
2200	72.82
2500	72.83
3000	72.79
5000	72.22
50000	72.06
500000	72.06

image shows more homogeneous regions and almost no salt and pepper noise compared to the SVM classified image.

The UnECHO method applied on the Hyperion data in Section 4.6 was unable to improve the accuracy of the SVM result. In this section, we evaluated the performance of this method for different distance measures and different window sizes. The overall accuracies of this method are listed in the Table 5.5 for different distance measures and also for different window sizes. None of the experiment could exceed the overall accuracy of the SVM classification. Among these experiments, the UnECHO with Euclidean distance measure using a 4×4 window size achieved the highest accuracy and is almost the same as the overall accuracy of the SVM classification.

Table 5.3: Confusion matrix in percentage of the SVM experiment (PA = Producer Accuracy and UA = User Accuracy).

Class	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**UA**
1	**58.06**	0.00	0.00	0.00	0.69	0.00	0.00	0.47	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	75.00
2	0.00	**63.13**	5.64	1.60	0.00	0.00	0.00	0.00	0.00	5.86	10.40	1.47	0.00	0.00	0.00	5.13	70.56
3	0.00	3.70	**67.56**	3.74	0.00	0.00	0.00	0.00	0.00	3.69	14.26	2.21	0.00	0.00	0.00	0.00	54.11
4	0.00	3.77	4.49	**90.37**	2.31	1.18	0.00	0.00	0.00	2.60	2.37	8.84	0.65	0.00	0.30	0.00	41.73
5	3.23	0.22	0.00	0.00	**89.15**	1.62	7.69	0.23	0.00	0.22	0.71	0.18	0.00	4.44	3.57	0.00	78.94
6	0.00	0.07	0.00	1.07	0.69	**94.41**	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.33	5.36	0.00	94.97
7	0.00	0.07	0.00	0.00	0.46	0.00	**84.62**	0.23	0.00	0.22	0.04	0.00	0.00	0.00	0.00	0.00	61.11
8	38.71	0.00	0.00	0.53	0.69	0.00	7.69	**99.07**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	96.15
9	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	**80.00**	0.11	0.00	0.00	0.65	0.00	1.19	0.00	36.36
10	0.00	12.26	0.38	0.00	0.46	0.00	0.00	0.00	0.00	**75.49**	11.39	0.92	0.00	0.00	0.00	1.28	60.52
11	0.00	13.50	8.59	0.00	0.00	0.00	0.00	0.00	0.00	9.00	**51.43**	3.13	0.00	0.00	0.00	0.00	77.80
12	0.00	3.12	13.33	2.67	0.23	0.00	0.00	0.00	0.00	2.49	8.94	**82.69**	0.00	0.00	0.00	11.54	52.89
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.00	0.00	0.00	0.00	**98.71**	0.08	2.08	0.00	94.44
14	0.00	0.00	0.00	0.00	1.39	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**91.69**	21.73	0.00	93.14
15	0.00	0.07	0.00	0.00	3.93	2.35	0.00	0.00	0.00	0.22	0.21	0.37	0.00	3.46	**65.77**	0.00	72.22
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	**82.05**	98.46
PA	58.06	63.13	67.56	90.37	89.15	94.41	84.62	99.07	80.00	75.49	51.43	82.69	98.71	91.69	65.77	82.05	

Table 5.4: Confusion matrix in percentage of the MRF experiment (PA = Producer Accuracy and UA = User Accuracy).

Class	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**UA**
1	**87.10**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
2	0.00	**73.00**	1.41	0.00	0.00	0.00	0.00	0.00	0.00	0.43	5.20	0.00	0.00	0.00	0.00	0.00	87.78
3	0.00	2.25	**83.85**	0.00	0.23	0.00	0.00	0.00	0.00	1.95	10.56	0.00	0.65	0.00	0.00	0.00	68.20
4	0.00	1.96	3.97	**100.00**	0.46	0.00	0.00	0.00	0.00	2.39	1.58	5.89	0.00	0.00	0.00	0.00	55.16
5	0.00	0.00	0.00	0.00	**95.15**	0.44	7.69	0.00	0.00	0.11	0.46	0.00	0.00	1.07	0.00	0.00	93.42
6	0.00	0.15	0.00	0.00	0.46	**99.26**	0.00	0.00	0.00	0.22	0.17	0.00	0.00	0.08	0.30	0.00	98.25
7	0.00	0.00	0.00	0.00	0.00	0.00	**92.31**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
8	12.90	0.00	0.00	0.00	0.00	0.00	0.00	**100.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	99.07
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**100.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
10	0.00	11.25	0.13	0.00	0.00	0.00	0.00	0.00	0.00	**90.56**	6.32	0.00	0.00	0.00	0.00	0.00	73.05
11	0.00	9.22	2.56	0.00	0.00	0.00	0.00	0.00	0.00	3.25	**65.65**	0.00	0.00	0.00	0.00	0.00	89.92
12	0.00	2.18	8.08	0.00	0.00	0.00	0.00	0.00	0.00	0.76	9.60	**93.37**	0.00	0.00	0.00	17.95	59.51
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**99.35**	0.08	0.60	0.00	98.09
14	0.00	0.00	0.00	0.00	1.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**96.63**	17.86	0.00	94.75
15	0.00	0.00	0.00	0.00	2.54	0.29	0.00	0.00	0.00	0.33	0.46	0.55	0.00	2.14	**81.25**	0.00	82.98
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	**82.05**	98.46
PA	87.10	73.00	83.85	100.00	95.15	99.26	92.31	100.00	100.00	90.56	65.65	93.37	99.35	96.63	81.25	82.05	

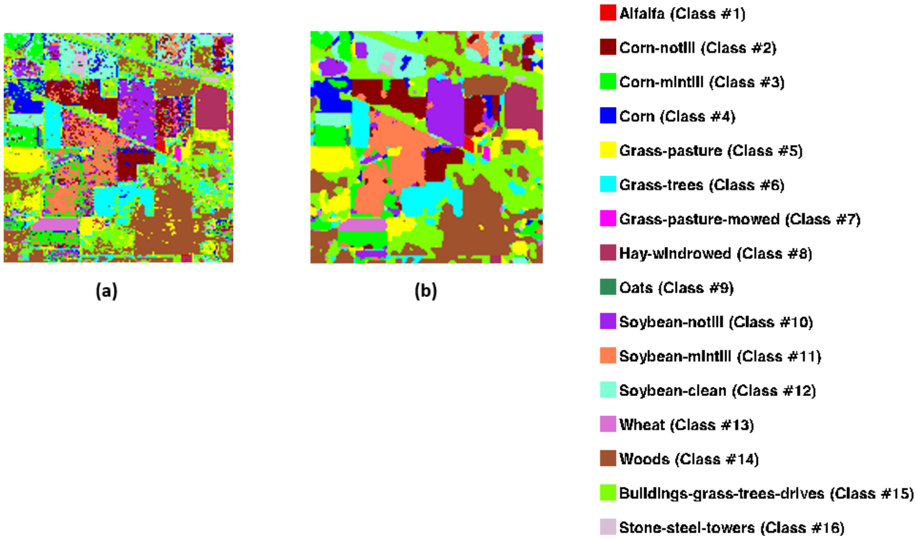


Figure 5.5: Classification maps of AVIRIS Indian Pines dataset; (a) SVM classified image and (b) MRF classified image.

5.3 Experiments for the AVIRIS Dataset with Feature Reduction

All feature reduction techniques (i.e., PCA, ICA, SVM-RFE, CFS, and mRMR) that were applied to the Hyperion dataset (Sections 4.8 to 4.12) are also applied to the AVIRIS dataset in this section for only 10, 20, 30, 40, 60, 80, 100, 140, and 180 bands. The grid search method is used in the SVM experiment for some of the band subsets of each feature reduction technique. It is found that changing the parameters values in the SVM experiment (i.e., different values of C and γ) with these feature reduction techniques shows very insignificant improvement in accuracies compared to the results with the corresponding band subsets where the same parameters values are used that were selected as best in Section 5.2 (i.e., $C = 2000$ and γ =1/total number of bands). Therefore, we used the same parameters values and choices for the SVM, MRF and UnECHO experiments as in Section 5.2.

At first, the SVM, MRF and UnECHO methods are applied for different number of PCs.

Table 5.5: Overall accuracies of the UnECHO experiment for different distance measures and window sizes.

Name of distance measures and size of window	**Overall accuracy (%)**
Euclidean distance and 2×2 window size	70.96
Euclidean distance and 3×3 window size	71.13
Euclidean distance and 4×4 window size	71.58
Mahalanobis distance and 2×2 window size	70.85
Mahalanobis distance and 3×3 window size	71.04
Mahalanobis distance and 4×4 window size	71.62
Maximum Likelihood distance and 2×2 window size	35.42
Maximum Likelihood distance and 3×3 window size	42.14
Maximum Likelihood distance and 4×4 window size	52.97

The overall accuracies of all experimental results are listed in Table 5.6. None of the results of SVM using PCs exceeded the accuracy of the SVM method using all the bands. The highest accuracy of SVM is achieved for the 40 PCs and it is lower by 4.32 % compared to the SVM method using all the bands. However, for the MRF method, accuracies for some PCs (i.e., 40, 60 and 100 PCs) exceeded the accuracy of the MRF result using all the original bands. Among these, the 60 PCs shows the highest accuracy that is improved by 1.11 % compared to the MRF method using all original bands. The improvement is 16.12 % compared to the SVM method using the 60 PCs, which is a significant improvement. Figure 5.6 shows the visual comparison of the SVM and MRF classification results using the 60 PCs together with the SVM and MRF result images using the all original bands along with the ground-reference data. We can see that the SVM result image using 60 PCs contains more salt and pepper noise than the SVM result image using all original bands and the overall accuracy of the SVM method using PCs is lower than the overall accuracy of the SVM method using all the bands, because PCs are noisy except the first few bands. The salt and pepper noise and misclassified pixels in the SVM result image for the 60 PCs are randomly distributed, but in the SVM result image for original bands mostly occurred as clusters. As the MRF method classify pixels to the class of its neighbouring pixels, it is easy for this method to correct the misclassified pixels, which are randomly distributed without clusters. For example, some pixels of the region of class 11 (Figure 5.6) are misclassified as class 5 (mostly clustered) in the SVM classified image using all original bands. However, in the SVM classified image using the PCs the misclassified pixels are not clustered. The MRF method could correct all the misclassified pixels of the SVM with the PCs, but could not correct all the misclassified pixels of the SVM using the original bands. The UnECHO method showed similar results as in the previous experiments. However, the accuracy for the experiment using the 140 PCs is improved by 0.43 % which is not a significant improvement.

Table 5.6: List of overall accuracies obtained by the SVM, MRF, and UnECHO experiments for AVIRIS data with different number of PCs.

Number of PCs	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	64.08	73.59	63.75
20	67.80	79.60	67.73
30	66.59	80.86	65.89
40	68.52	83.76	68.08
60	68.33	84.45	67.87
80	66.66	83.17	65.65
100	64.83	83.51	64.25
140	59.87	81.85	60.30
180	55.75	79.03	55.06

Table 5.7 lists the overall accuracies of all experimental results (SVM, MRF and UnECHO) for the different number of ICs. It shows almost similar trends as the PCA. The highest accuracy of SVM is obtained using the 20 ICs. The MRF method is improved by 0.23 % for 40 ICs compared to the MRF result using the original bands and 17.40 % compared to the SVM result using 40 ICs, which is also a significant improvement and the reason is the same as for the PCA.

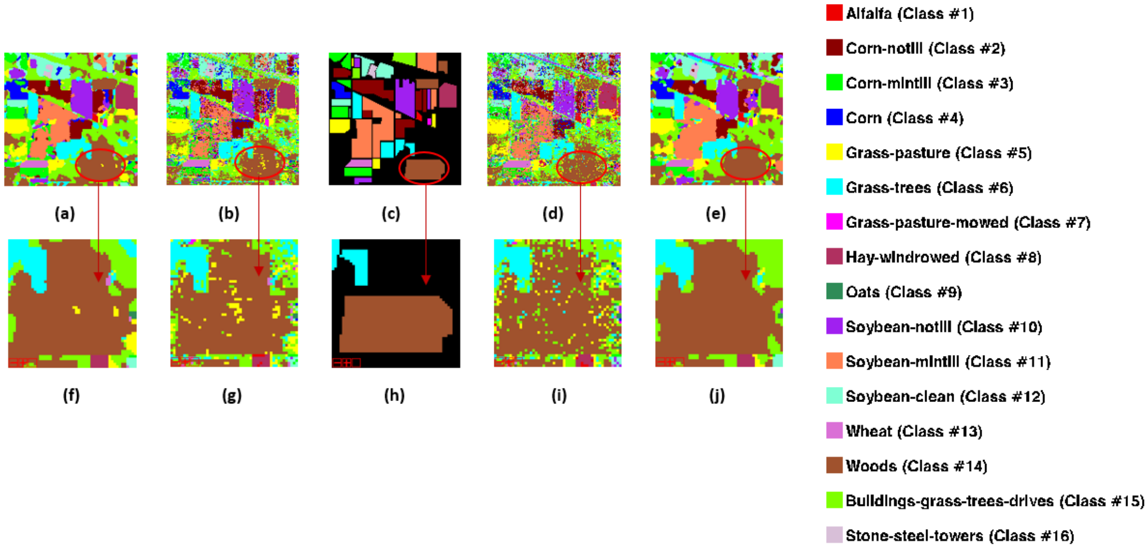


Figure 5.6: (a) MRF classification using all the original bands, (b) SVM classification using all the original bands, (c) ground-reference data, (d) SVM classification using the 60 PCs, (e) MRF classification using the 60 PCs, (f) misclassified pixels (class 5) in the region (class 11) of the MRF classification using all the bands, (g) misclassified pixels mostly clustered based on the region (class 11) of the SVM classification using all the bands, (h) ground-reference region of class 11, (i) misclassified pixels mostly scattered in the the region (class 11) of the SVM classification using all the bands, and (j) all misclassified pixels corrected in the region (class 11) of the MRF classification using the 60 PCs

Table 5.7: Overall accuracies obtained by the SVM, MRF, and UnECHO experiments for the AVIRIS data using different number of ICs.

Number of ICs	SVM (%)	MRF (%)	UnECHO (%)
10	63.40	72.97	63.39
20	68.18	80.58	68.86
30	66.87	81.82	67.54
40	66.17	83.57	66.88
60	63.78	83.22	64.30
80	62.75	82.69	63.45
100	59.51	78.88	59.92
140	58.58	80.82	59.78
180	53.98	79.15	54.25

The overall accuracies of the SVM, MRF, and UnECHO experiments obtained from the AVIRIS dataset for the different numbers of SVM-RFE selected bands are reported in Table 5.8. All the accuracies of the SVM result for the SVM-RFE selected bands 10 to 60 are above the accuracy of SVM result using all the bands. The highest accuracy improved by 1.69 % for the 20 and 30 SVM-RFE selected bands compared to the SVM result using all original bands. The accuracy of MRF method improved by 0.02 % for only the 60 SVM-RFE selected bands compared to the MRF result using all the bands.

The CFS method selects 35 bands from the dataset. From the Table 5.9 we see that none of the experiments (SVM, MRF, UnECHO) using the 35 CFS selected bands could improve the accuracies of the experiments using all the bands.

Using 80, 100, 140, 180 mRMR selected bands for the SVM classification improved the accuracies noticeably as shown in Table 5.10. The accuracies of the corresponding MRF results except for the 100 mRMR selected bands are also improved. The highest accuracy improvements of SVM and MRF results are 3.40 % and 1.75 % for 140 mRMR selected bands compared to the accuracies of SVM and MRF results using all the bands. These accuracies of SVM and MRF results are the highest among all feature reduction

Table 5.8: Overall accuracies of the SVM, MRF, and UnECHO experiments obtained from the AVIRIS data using SVM-RFE selected bands.

Number of bands	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	74.10	83.17	73.47
20	74.53	82.19	73.42
30	74.53	82.58	73.21
40	74.25	82.81	73.55
60	73.56	83.36	72.70
80	69.85	78.69	69.72
100	68.08	77.13	67.84
140	65.48	75.26	64.75
180	57.30	65.29	55.17

Table 5.9: Overall accuracies obtained by the SVM, MRF, and UnECHO experiments for the AVIRIS data using the CFS selected bands.

SVM (%)	**MRF (%)**	**UnECHO (%)**
71.26	82.10	71.06

band subsets (i.e., PCA, ICA, SVM-RFE, CFS, and mRMR). The result images of the SVM and MRF are shown in Figure 5.7 along with a visual comparison with the SVM and MRF classifications using all bands. Most regions are less noisy in the images for the 140 mRMR selected bands than the images for all original bands (e.g., class 11 region is shown in circle).

We also experimented with the SVM method for different number of PCs and ICs of the data where all default parameters values were used. After that all the subsequent MRF and UnECHO methods were tested for those number of PCs and ICs. A summary of the overall accuracies of the SVM, MRF and UnECHO experiments are shown in Tables 5.11 and 5.12, respectively for different numbers of PCs and ICs. In both of these tables, none of the accuracies of the SVM experiments are better than the accuracy of SVM using all bands where all default values were set. However, it can be seen that most of the accuracies

Table 5.10: Overall accuracies obtained by the SVM, MRF, and UnECHO experiments for the AVIRIS data using different number of mRMR bands.

Number of mRMR Bands	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	46.91	55.69	46.07
20	65.41	76.05	63.60
30	66.44	76.49	64.96
40	69.04	78.63	66.72
60	71.47	81.20	69.52
80	74.08	83.45	72.78
100	74.11	83.04	72.19
140	76.24	85.09	73.56
180	74.36	83.87	73.04

of the MRF experiments for both PCs and ICs are more than 70 % where the accuracy of the MRF experiment using all the bands was 68.85 %. Among all PCs and ICs, the highest accuarcy of the MRF method is 79.82 % using the 140 ICs. The improvement is 20.64 % compared to the SVM result using 140 ICs and 10.97 % compared to the MRF result using all the bands. This is the most significant improvement and the reason of this MRF improvement is similar to the case for the 60 PCs where the value of $C = 2000$ is set.

Table 5.11: Overall accuracies obtained by the SVM (with default parameter values), MRF, and UnECHO experiments for the AVIRIS data using different number of PCs.

Number of PCs	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	55.80	63.19	57.03
20	61.62	72.28	62.04
30	61.07	72.64	61.15
40	62.38	74.20	61.63
60	61.76	74.67	60.91
80	61.75	73.78	60.34
100	60.75	73.80	59.13
140	59.59	75.84	59.12
180	56.89	77.14	56.69

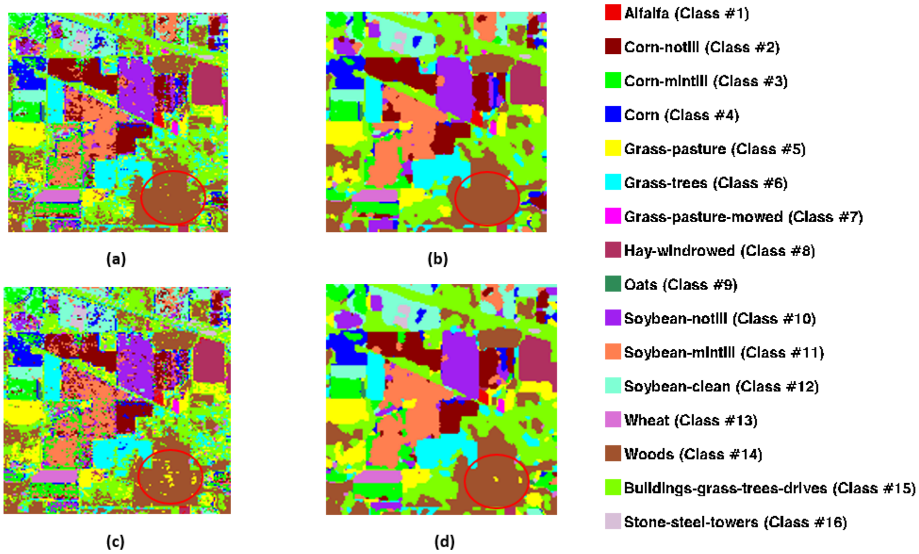


Figure 5.7: Classification maps of the AVIRIS Indian Pines dataset: (a) SVM classification for 140 mRMR selected bands, (b) MRF classification for 140 mRMR selected bands, (c) SVM classification for all original bands, and (d) MRF classification for all original bands.

All experimental results that are presented so far for the AVIRIS dataset can be grouped according to whether feature reduction is used, and whether the dafault parameter values are used in the SVM training. According to the accuracy assessment, the best result is achieved by searching for the best parameters values of SVM. However, it is computationally very intensive to find the best parameters values of SVM. Sometimes 100 SVM experiments might be required to determine the best values of parameters, which is very time consuming. Instead, if we apply SVM classification with all default parameter values of the data with ICs and then apply the MRF method, we can obtain accuracies much closer to those obtained by using the best parameters values. Therefore, one may choose to use the SVM method (with default parameters values) in combination with the MRF method, because of the reduced processing time even if the result is slightly less accurate.

Table 5.12: Overall accuracies obtained by the SVM (with default parameter values), MRF, and UnECHO experiments for the AVIRIS data using different number of ICs.

Number of ICs	SVM (%)	MRF (%)	UnECHO (%)
10	56.15	63.19	56.53
20	60.13	70.63	59.95
30	58.71	68.55	58.01
40	58.65	69.75	58.47
60	57.95	73.07	58.28
80	59.80	75.51	58.92
100	58.69	71.96	57.86
140	59.18	79.82	58.34
180	54.27	77.52	54.18

5.4 Summary of Experiments using the AVIRIS Dataset

The accuracy of the SVM classification using the AVIRIS dataset was not satisfactory (62.54 %) with all default parameter value settings in the SVM experiment. Hence, the grid search technique (testing accuracy using different values of the SVM parameters) was applied in the SVM experiment and the best accuracy (72.84 %) was achieved using $C = 2000$. The accuracy improvement was significant (10.30 %) with satisfactory individual class producer accuracies. The overall accuracy of the subsequent MRF method was 83.34 % with an improvement of 10.50 % compared to the accuracy of the SVM classification result using best parameters values.

The best parameters values obtained from the SVM experiment using all bands were also applied to all experiments for all feature reduction techniques. All the accuracies of the SVM classification using different PCs were less than the accuracy of the SVM classification using all bands. However, the MRF method using 40, 60, and 100 PCs improved the accuracy of the MRF method using all the bands. The improvement is 1.11 % using 60 PCs compared to the MRF using all the bands and 16.12 % compared to the SVM using 60

PCs. The experiments using different ICs showed similar accuracies as the PCs. The MRF using 40 ICs showed an improvement of 0.23 % compared to the MRF using all the bands and 17.40 % compared to the SVM using 40 ICs. Among all experiments using different SVM-RFE selected bands, the SVM using 20 and 30 SVM-RFE selected bands showed the highest improvement (1.69 %) compared to the SVM using all the bands. MRF showed an improvement of 0.02 % using 60 selected bands compared to the MRF using all the bands. The SVM and MRF experiments using CFS selected bands was unable to improve the accuracy compared to the SVM and MRF using all the bands. mRMR showed the best result among all feature reduction techniques. The accuracy of the SVM and MRF result using 140 selected bands improved by 3.40 % and 1.75 % respectively, compared to the accuracy of the SVM and MRF classification using all the bands.

All the experiments were conducted using different PCs and ICs where all default parameters values were set in the SVM experiment. The accuracies were not improved by the SVM method compare to the SVM using all the bands where default parameters values were used. However, the MRF method improved the accuracies of the corresponding SVM classification significantly, which are closer to the accuracy of the MRF result where best parameters values were used in the SVM experiment. The highest accuracy, 79.82 % was achieved by the MRF method using 140 ICs with an improvement of 20.64 % compared to the SVM method using the same number of ICs and 10.97 % compared to the MRF using all bands where all default parameters values were used. We concluded that it is beneficial to apply the SVM method with default parameters values with a combination of the MRF method using PCA/ICA instead of applying the SVM with grid search and the MRF method.

5.5 ROSIS Dataset

An urban hyperspectral dataset acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) over the University of Pavia, Pavia, Italy was used for this experiment [32, 40]. The flight was operated by the German Aerospace Centre under the European Union's HySens project. This airborne hyperspectral dataset with ground-reference information is publicly available from the Purdue Universitys MultiSpec website [32]. The scene was already atmospherically corrected [15] and has 610×340 pixels with 103 bands covering the spectral range from 0.43 to 0.86 μm with 1.3 m spatial resolution. Ground-reference data include 9 classes with 42,776 samples that comprise urban properties and few soil and vegetation areas.

5.6 Experiments for the ROSIS Dataset without Feature Reduction

In this section, we describe experiments using the SVM, MRF and UnECHO methods on the ROSIS dataset without feature reduction. A false-colour composite of the dataset together with the ground-reference of the 9 classes and the randomly selected training samples used to train the SVM are shown in Figure 5.8. The name of all classes of this dataset, the corresponding colours, the number of total samples of each class present in the ground-reference data, and the number of training samples selected for the SVM experiment are reported in Table 5.13. The training pixels are selected randomly and the number of the training pixels is the same as used in a previous work [32].

The accuracy of the SVM classification using all default parameters values is 76.24 %. The accuracy of the subsequent MRF classification is 81.37 % with an improvement of 5.13 %. Because the accuracy is not satisfactory, different values of the penalty parameters

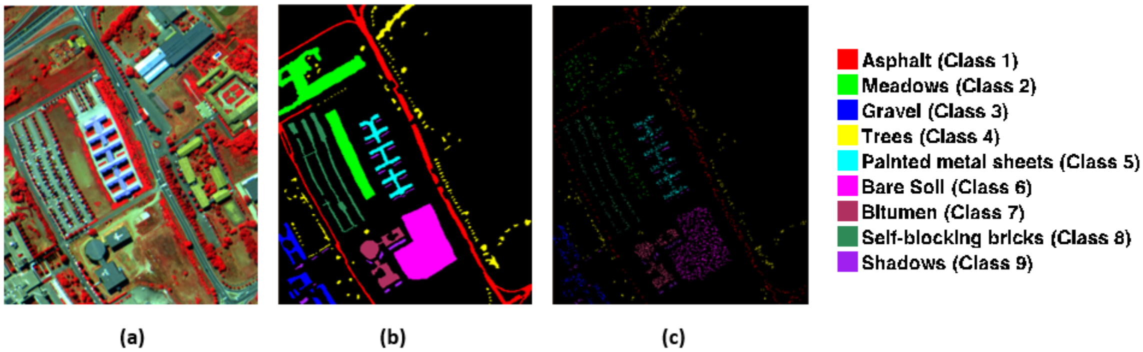


Figure 5.8: ROSIS University dataset: (a) false-color composite (R: NIR, G: Red, B: Blue), (b) ground-reference of the 9 classes, and (c) training samples.

Table 5.13: Ground-reference classes for the ROSIS dataset and their respective total number of samples, number of training samples and colour.

Class No.	**Class Name**	**Colour**	**Pixels**	**Training pixels**
Class 1	Asphalt	Red	6631	548
Class 2	Meadows	Green	18649	540
Class 3	Gravel	Blue	2099	392
Class 4	Trees	Yellow	3064	524
Class 5	Painted metal sheets	Cyan	1345	265
Class 6	Bare Soil	Magenta	5029	532
Class 7	Bitumen	Maroon	1330	375
Class 8	Self-Blocking Bricks	Sea Green	3682	514
Class 9	Shadows	Purple	947	231

C and kernel parameter γ are selected in a grid search for the SVM classification to obtain better accuracy. The highest accuracy, 91.81 % is achieved using the value of the penalty parameters $C = 27000$ and the kernel parameter $\gamma = 0.01$, which lead to an improvement of 15.57 % compared to the SVM experiment with default parameter value settings. This is a very significant improvement (z score = 66.09). The producer accuracies of most of the classes (Table 5.14) are also over 90 %. It is lowest for class 3 (83.13 %). There are 15.35 % of pixels in class 3 misclassified as class 8. On the other hand, there are 10.26 % of pixels in class 8 misclassified as class 3. The reason behind this is the similar spectral

Table 5.14: Confusion matrix of the SVM classification in percent for the ROSIS data (PA = Producer Accuracy, UA = User Accuracy).

Class	1	2	3	4	5	6	7	8	9	UA (%)
1	**90.63**	0.03	1.05	0.08	0.00	0.11	10.16	1.55	0.14	96.89
2	0.21	**92.45**	0.47	2.24	0.09	7.61	0.00	0.22	0.00	97.51
3	1.46	0.01	**83.13**	0.00	0.00	0.00	0.00	10.26	0.00	77.37
4	0.00	1.26	0.00	**97.52**	0.00	0.31	0.00	0.00	0.00	91.07
5	0.00	0.00	0.00	0.04	**99.35**	0.00	0.00	0.00	0.00	99.91
6	0.15	6.07	0.00	0.12	0.09	**91.35**	0.00	0.38	0.00	78.50
7	4.36	0.00	0.00	0.00	0.28	0.00	**89.74**	0.16	0.00	75.84
8	3.17	0.18	15.35	0.00	0.19	0.62	0.10	**87.44**	0.00	84.25
9	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**99.86**	99.86
PA (%)	90.63	92.45	83.13	97.52	99.35	91.35	89.74	87.44	99.86	

signature of class 3 and class 8 (Figure 5.9).

The MRF method improved the accuracy to 97.35 % with an improvement of 5.54 % (z score = 42.55). The confusion matrix of the MRF experiment is shown in Table 5.15. The producer accuracies of all classes are very high and improved compared to the SVM classification result. The SVM and MRF classification results are shown in Figure 5.10 along with the ground-reference data. It is clear that most of the ground-reference regions are classified correctly by the SVM with a small amount of salt and pepper noise and the MRF method corrected those noise and provided homogeneous regions.

The UnECHO method could not improve the accuracy of the SVM for any of distance measures and window sizes. The highest accuracy is 87.34 % with the Euclidean distance measure and a 3×3 window size.

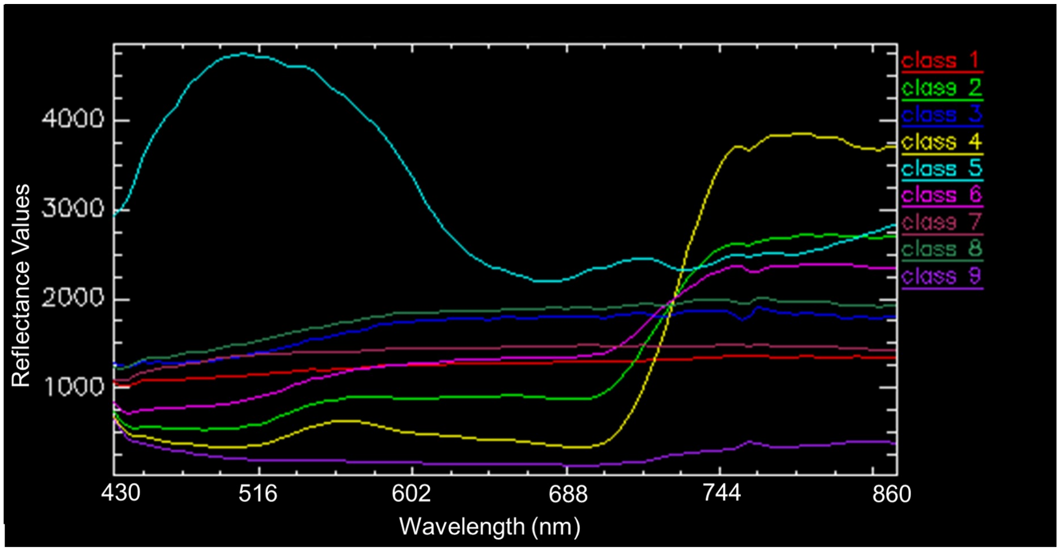


Figure 5.9: Spectral signatures of the 9 classes.

Table 5.15: Confusion matrix of the MRF classification in percent for the ROSIS data (PA = Producer Accuracy, UA = User Accuracy).

Class	1	2	3	4	5	6	7	8	9	UA (%)
1	**96.89**	0.00	0.64	0.08	0.00	0.00	3.77	0.47	0.00	98.93
2	0.00	**97.98**	0.18	0.67	0.00	1.42	0.00	0.19	0.00	99.50
3	0.35	0.00	**88.52**	0.00	0.00	0.00	0.00	4.04	0.00	91.02
4	0.00	0.13	0.00	**99.06**	0.00	0.00	0.00	0.03	0.00	99.02
5	0.00	0.00	0.00	0.00	**99.91**	0.00	0.00	0.00	0.00	100.00
6	0.05	1.88	0.00	0.20	0.09	**98.58**	0.00	0.16	0.00	92.59
7	0.15	0.00	0.00	0.00	0.00	0.00	**96.23**	0.00	0.00	99.03
8	2.56	0.00	10.66	0.00	0.00	0.00	0.00	**95.11**	0.00	89.91
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**100.00**	100.00
PA (%)	96.89	97.98	88.52	99.06	99.91	98.58	96.23	95.11	100.00	

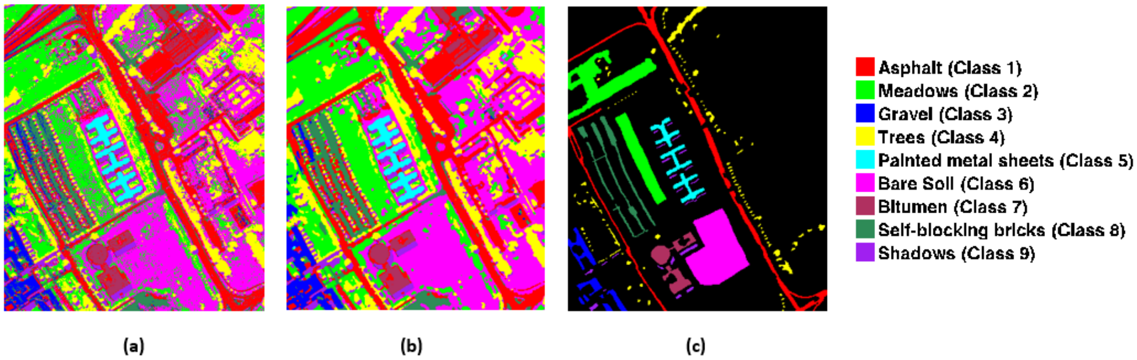


Figure 5.10: (a) SVM classified image, (b) MRF classified image, and (c) ground-reference data.

5.7 Experiments for the ROSIS Dataset with Feature Reduction

In this section, the SVM, MRF and UnECHO methods are applied to the ROSIS dataset using different band subsets with 10, 20, 30, 40, 60 and 80 bands after applying all feature reduction techniques one at a time just as we did to the AVIRIS dataset (Section 5.3). Because there is very insignificant improvement using the grid search method in the SVM experiment with feature reduction, we used the same parameter values as in Section 5.6 for all feature reduction techniques. The default values are then used only for the PCs and ICs.

The accuracies of all experimental results with best parameters values in the SVM experiment using different PCs are listed in Table 5.16. All accuracies are lower than the accuracies of experiments using all bands. The highest accuracies are achieved by the SVM and MRF methods using 20 PCs. The accuracy of the SVM method declined the accuracy of the SVM method using all bands by 3.40 % and the accuracy of the MRF method declined the accuracy of the MRF method using all bands by 2.53 %. The accuracies of all methods shown in Table 5.17 for the different ICs are lower than the accuracies achieved using PCs. Among all the accuracies listed in the Table 5.18 for different SVM-

RFE band subsets, the highest accuracy is achieved by the SVM and MRF methods using 10 SVM-RFE selected bands which are very close to the accuracy of the SVM and MRF method using all bands. However, the accuracy of the SVM method is declined by 0.66 % compared to the accuracy of the SVM method using all bands and the accuracy of the MRF method is declined by 0.26 % compared to the accuracy of the MRF method using all bands. The accuracies of all experiments using CFS selected bands (31 bands) shown in Table 5.19. The accuracies are not improved compared to the accuracies using all bands. Table 5.20 shows the accuracies of all experiments using different mRMR selected bands, which are also less than the accuracies using all bands.

Table 5.16: Overall accuracies obtained by the SVM, MRF, and UnECHO experiments for the ROSIS data using different number of PCs.

Number of PCs	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	84.00	92.54	80.17
20	88.41	94.81	85.50
30	88.03	94.46	85.40
40	87.87	94.52	85.35
60	87.63	94.18	84.93
80	87.21	94.10	84.03

Table 5.17: Overall accuracies obtained by the SVM, MRF, and UnECHO experiments for the ROSIS data using different number of ICs.

Number of ICs	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	71.28	88.05	73.24
20	78.29	91.80	78.42
30	79.56	92.74	79.81
40	80.75	93.36	80.18
60	81.24	93.18	80.51
80	84.12	94.45	82.05

Tables 5.21 and 5.22 represent the accuracies of all experiments using different PCs and

Table 5.18: Overall accuracies of the SVM, MRF, and UnECHO experiments obtained from the ROSIS data with SVM-RFE selected bands.

Number of bands	SVM (%)	MRF (%)	UnECHO (%)
10	91.15	97.09	87.94
20	90.81	96.83	87.65
30	89.83	96.14	86.79
40	89.48	95.84	86.32
60	86.34	94.05	83.59
80	79.62	89.95	79.04

Table 5.19: Overall accuracies obtained by the SVM, MRF, and UnECHO experiment for the ROSIS data with the CFS selected bands.

SVM (%)	MRF (%)	UnECHO (%)
90.06	96.69	86.65

ICs where all default values were set in the SVM experiment. Among all the accuracies, the accuracies of the SVM and MRF methods using 30 ICs are highest. The accuracy of the SVM method is improved by 4.96 % compared to the SVM classification using all bands where all default parameters values were used. The subsequent MRF method improved the accuracy by 7.78 % compared to the SVM classification using 30 ICs and 7.61 compared to the MRF classification using all the bands. Although the accuracy of the MRF method (88.98 %) is not close to the accuracy of the MRF classification (97.35 %) using best parameters values in the SVM experiments, it is satisfactory. Sometimes it will be time efficient to apply the SVM and MRF method using PCs/ICs with default parameter value settings in the SVM experiment instead of applying the SVM with grid search and MRF method.

Table 5.20: Overall accuracies obtained by the SVM, MRF, UnECHO experiments for the ROSIS data using different number of mRMR selected bands

Number of mRMR Bands	SVM (%)	MRF (%)	UnECHO (%)
10	64.82	76.17	66.47
20	75.10	83.47	72.17
30	82.90	90.56	80.09
40	84.91	91.87	82.41
60	88.33	95.38	85.14
80	90.19	96.90	86.65

Table 5.21: Overall accuracies obtained by the SVM (with default parameter values), MRF, and UnECHO experiment for the ROSIS data using different number of PCs

Number of PCs	SVM (%)	MRF (%)	UnECHO (%)
10	75.65	80.99	69.04
20	77.19	82.55	72.56
30	75.58	80.76	71.02
40	74.96	80.20	68.56
60	73.29	78.24	65.98
80	71.95	76.92	64.34

5.8 Summary of Experiments using the ROSIS Dataset

The SVM experiment was conducted with the grid search technique (accuracy was 91.81 % with a significant improvement of 15.57 %) using the ROSIS dataset because the accuracy (76.24 %) was not satisfactory using default parameter settings in the SVM experiment. The subsequent MRF method improved the accuracy from 91.81 % to 97.35 %. This is also a satisfactory improvement (5.54 %). None of the feature reduction techniques (i.e., PCA, ICA, SVM-RFE, CFS, mRMR) improved the accuracy of the experiments using all the original bands. The accuracies of the experiments using the 30 ICs with default parameter settings in the SVM experiment were quite satisfactory. The accuracy was improved to 88.98 % using the MRF method. Therefore, considering the computational time, some-

Table 5.22: Overall accuracies obtained by the SVM (with default parameter values), MRF, and UnECHO experiments for the ROSIS data using different number of ICs

Number of ICs	**SVM (%)**	**MRF (%)**	**UnECHO (%)**
10	77.43	85.36	72.21
20	79.46	85.60	73.98
30	81.20	88.98	76.63
40	76.52	86.12	73.46
60	79.82	88.27	74.83
80	77.65	87.17	72.32

times it is better to apply the SVM and MRF method using PCs/ICs with default parameters values rather than the SVM method with grid search in a combination of the MRF method.

Chapter 6

Conclusion And Future work

6.1 Summary and Conclusion

In this thesis, a new spectral-spatial or contextual classification approach is proposed. It is a combination of one spectral and two spatial methods. The main objective of this thesis is to improve classification accuracy. However, computation time is also considered. We examined and compared individual spectral-spatial classification methods with each other and also with the spectral method alone. The proposed method was investigated for different sets of hyperspectral data. We also tried different types of feature reduction techniques to select features or bands in order to improve the accuracy and computation time.

The SVM classification method was chosen as the spectral method. Next, a spatial method using MRF was applied to boundary pixels of the SVM classification where the boundary pixels were extracted using the erosion technique. We also used another spatial method called the UnECHO method. It was applied on the SVM classification and the resulting labeled pixels were extracted from the interior parts. The labels of the boundary pixels extracted from the MRF classification and the interior pixels from the UnECHO classification were then combined to produce the final classification map. Three different hyperspectral datasets were used. The first one is a Hyperion dataset (agriculture), the second one is an AVIRIS dataset (agriculture), and the third one is a ROSIS dataset (urban). We tried two types of feature reduction techniques: feature extraction and feature selection for all datasets. PCA and ICA were applied as feature extraction techniques and SVM-RFE, CFS and mRMR were applied as feature selection techniques. The five main findings that have been made in this thesis are as follows:

SVM + MRF effect: For the Hyperion dataset, the MRF method applied only to the boundaries of the SVM classification improved the accuracy from 89.35% to 92.81%, when the MRF regularization algorithm minimizes the local energy of all boundary pixels. The improvement was not substantial with the small amount of ground-reference information available. However, it was noticeable in the visual comparison that the MRF method corrected most of the salt and pepper noise of the SVM classification and made regions more homogeneous. This method improved the accuracy of the SVM classification of the AVIRIS dataset significantly (10.50 %) with almost no salt and pepper noise and more homogeneous regions. The accuracy of the SVM classification was already above 90% for the ROSIS dataset and the MRF method improved the accuracy by 5.54 %, which was a satisfactory improvement. Therefore, we can conclude that the MRF method is an effective method when applied to the boundary pixels of the SVM classification because the performance of this method was very good for all datasets. The main drawback of this method is that it will not be able to improve the accuracy of a non-homogeneous datasets.

SVM + UnECHO effect: The accuracy of the SVM classification was not improved by the UnECHO method for any of the datasets and for any cases. We can conclude that this method does not provide better accuracy in combination with supervised classification techniques such as SVM.

Feature reduction effect: The SVM, MRF and UnECHO methods were applied for different numbers of bands of the datasets and compared after applying all 5 individual feature reduction techniques. The accuracies of the SVM and MRF classifications improved significantly for the Hyperion dataset to almost 100% for all numbers of PCs. However, after visual comparison of the results for three different PCs (i.e., 10, 30 and 140 PCs) we concluded that the results using 10 PCs were the best. In case of the ICA, we obtained similar accuracies as with the PCA. The results using the 10 ICs were considered to be the best result after visually comparing it with two other different numbers of ICs (i.e., 40 and 120 ICs). Therefore, experiments of SVM and MRF methods with both PCA and ICA showed excellent performance for the Hyperion dataset. Among all feature selection techniques, CFS showed the highest performance (accuracy of the MRF classification is 94.13 %) for this dataset although the improvement was not significant.

All feature reduction techniques were applied to the AVIRIS dataset for all experiments using different numbers of PCs/ICs where the best parameter values for SVM were found with a grid search. Although the accuracies of the SVM classification were lower using all numbers of PCs compared to the accuracy of the SVM classification using all bands, the MRF method improved the accuracies for three different numbers of PCs compared to the MRF method with all bands. Among these, the highest improvement was achieved with 60 PCs, but this was a very slight improvement. The ICA method showed similar results. The mRMR method showed the highest accuracy improvement using 140 bands among all feature reduction techniques for this dataset, however this was also not a significant improvement. All the experiments were considered with different numbers of PCs and ICs where all default values were used in the SVM experiment. Although the accuracies of the SVM classifications were not better than the accuracy of the SVM classification with all bands, the MRF method improved all accuracies of the SVM classifications significantly for almost all numbers of PCs and ICs. The highest accuracy was close to the accuracy of the MRF classification using all bands where the best parameter values were found with a grid search in the SVM experiment. Therefore, for a large dataset, PCA and ICA may be used with the SVM and MRF method to reduce running time without significantly reducing accuracy. Although it is slightly more accurate, it is very time consuming to use the grid search for the best SVM training parameters when all bands are used.

None of the feature reduction techniques could improve the accuracy for the ROSIS dataset when the grid search was used to obtain the best parameters for the SVM experiment. However, PCA and ICA experiments using default parameter values in the SVM experiment showed significant accuracy improvement of the SVM classifications using the MRF method. Although the highest accuracy of the MRF classification using the 30 ICs was still lower than the accuracy of the MRF classification with all bands, it is less computationally intensive to apply the SVM and MRF method with the ICs with all default parameter settings.

Among all feature reduction techniques, the experiments with PCA and ICA work very well when all default parameter values are used. However, these techniques are unable to improve the accuracy when experiments are conducted with the grid search result.

Computation time and accuracy trade-off: After applying the SVM method to the Hyperion dataset using all the bands, two types of MRF algorithms (Algorithm 1: for all boundary pixels and Algorithm 2: for randomly selected pixels) were applied to the SVM classification results, and compared with each other for this dataset only. At first, the MRF method with Algorithm 1 was applied to the entire image and also to the boundaries of the SVM classification, and they were compared with each other based on the accuracy and computation time. Experimental results with best parameter values showed that Algorithm 1 applied to the entire image, provided slightly more accurate result, but took 7 times longer in execution time than when it was applied to only the boundaries. Therefore, it is worthwhile to apply the MRF method for all pixels only to the boundaries instead of the entire image even with a slightly less accurate result. The MRF method for randomly selected pixels was then applied only to the boundaries of the SVM classification and compared with the MRF method for all pixels applied to the boundary pixels. The accuracy of the MRF method for randomly selected pixels was lower and this method took 73 times more computation time than the MRF method for all pixels. The MRF method for randomly selected pixels has been used in previous studies in order to improve the accuracy, but the experimental results of this thesis showed that the MRF method for all pixels is better than MRF method for randomly selected pixels in terms of both accuracy and computation time.

Grid search vs. no grid search: Because the accuracy of the SVM method with all bands for the AVIRIS dataset was not satisfactory using the default SVM training parameters values in the experiment, the grid search technique was applied to select the best training parameters for this experiment. The accuracy of this method improved significantly (10.30 %) using the best parameter values from the grid search result. For the ROSIS dataset the accuracy of the SVM method was improved by 15.57 % after applying the grid search technique. Hence, we can say that grid search technique is a very effective way of improving the accuracy of the SVM method when the accuracy of the SVM method is not satisfactory using default values.

SVM + MRF + UnECHO effect: The final classification map resulting from the combination of the MRF and UnECHO results was shown only for the Hyperion dataset. The accuracy of the final result did not improve mainly because of the UnECHO method.

6.2 Future Work

In the future, the following investigations can be carried out to examine the effectiveness of spectral-spatial methods used in this thesis.

Multispectral datasets: In this thesis, the proposed method is applied only to different hyperspectral datasets. All methods that are used in this thesis can be used for different multispectral datasets.

Other Spectral Methods: Only one method (i.e., SVM) is used as a spectral class classifier in this study. Other methods (e.g., Maximum Likelihood and Random Forest) can be tested and compare with the SVM method. Spatial methods such as MRF and UnECHO can also be applied on the result image derived from other spectral methods to examine the results for accuracy improvement.

Spatial Methods for the Interior Pixels: Different spatial techniques such as partitioning homogenous regions based on majority vote (i.e., adaptive neighbourhood) can be applied to the interior parts of spectrally classified imagery, because the UnECHO method cannot improve the accuracy of the SVM classification result.

Segmentation Process: Segmentation process which partitioned the image into multiple segments may classify the hyperspectral image more accurately, because this method labels pixels based on the locating objects and boundaries.

Spectral Unmixing for the Boundary Pixels: Sub-pixel level information extraction technique, such as spectral unmixing can be applied to the hyperspectral data. Spatial-Spectral Endmember Extraction (SSEE) algorithm can be applied to extract endmember (i.e., pure pixel) spectrum for each class. Dominant abundances (i.e., proportion of endmembers) from spectral unmixing results can be assigned to particular classes of the boundary pixels to improve the accuracy.

Class Mean Spectra from Cluster: The class mean spectra is required in the UnECHO method. It can be acquired from cluster or a contiguous group of pixels of ground-reference regions or by applying endmember extraction methods in order to examine the result of the UnECHO method.

Modify Energy Minimization Equation: The energy minimization equation of the MRF method with randomly selected pixels can be modified in order to improve the accuracy.

Other Feature Reduction Techniques: Other feature reduction techniques (e.g., MNF) may improve the accuracy of the methods used in this thesis.

Bibliography

[1] National Aeronautics and Space Administration. Landsat 7. `http://landsat.gsfc.nasa.gov/about/landsat7.html`.

[2] Alberta Agriculture and Rural Development. Crop Conditions as of June 23, 2005. Alberta Agriculture and Rural Development. Accessed on October 6, 2012. `http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/sdd10016`.

[3] A. Ahmad and S. Quegan. Analysis of Maximum Likelihood classification technique on Landsat 5 TM satellite data of tropical land covers. *2012 IEEE International Conference on Control System, Computing and Engineering*, pages 280–285, Nov 2012.

[4] M. Ahmad, U. I. Haq, Q. Mushtaq, and M. Sohaib. A New Statistical Approach for Band Clustering and Band Selection Using K-Means Clustering. *IACSIT International Journal of Engineering and Technology*, 3(6):606–614, December 2011.

[5] M. Ahmad, S. Lee, U. I. Haq, and Q. Mushtaq. Article: Hyperspectral Remote Sensing: Dimensional Reduction and End member Extraction. *International Journal of Soft Computing and Engineering (IJSCE)*, 2:170–175, May 2012.

[6] M. A. Ashraf, M. J. Maah, and I. Yusoff. *Introduction to Remote Sensing of Biomass, Biomass and Remote Sensing of Biomass*. Dr. Islam Atazadeh (Ed.), InTech, DOI: 10.5772/16462, 2011.

[7] J. C. Bansal. Advances in Intelligent Systems and Computing. *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, 201, 2013.

[8] R. Congalton. Exploring and Evaluating the Consequences of Vector-to-Raster and Raster-to-Vector Conversion. *Photogrammetric Engineering and Remote Sensing*, 63(4):425–434, April 1997.

[9] Satellite Imaging Corporation. SPOT-6 SATELLITE SENSOR. `http://www.satimagingcorp.com/satellite-sensors/spot-6.html`.

[10] N. Cristianini and J Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods.* Cambridge, U.K.: Cambridge Univ. Press, 2000.

[11] M. Dalponte, L. Bruzzone, L. Vescovo, and D. Gianelle. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment*, 113:2345–2355, November 2009.

[12] N. Efford. *Digital Image Processing: A Practical Introduction Using Java*. Pearson Education Limited, Harlow, England., 2000.

[13] S. A. El_Rahman, W. A. Aliady, and N. I. Alrashed. Supervised Classification Approaches to Analyze Hyperspectral Dataset. *I.J. Image, Graphics and Signal Processing*, pages 42–48, April 2015.

[14] F. Evans. An investigation into the use of maximum likelihood classifiers, decision trees, neural networks and conditional probabilistic networks for mapping and predicting salinity,. *M.Sc. thesis, School of Computing, Curtin University of Technology*, March 1998.

[15] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson. Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. . *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3804–3814, 2008.

[16] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Kernel Principal Component Analysis for Feature Reduction in Hyperspectrale Images Analysis. *Proceedings of the 7th Signal Processing Symposium*, pages 238–241, June 2006.

[17] R Fisher and Brown N. Erosion. `http://homepages.inf.ed.ac.uk/rbf/HIPR2/erode.htm`.

[18] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice-Hall, Inc., 2007.

[19] J. A. Gualtieri and S. Chettri. Support vector machines for classification of hyperspectral data. *In Proc. IGARSS*, 2:813–815, 2000.

[20] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson. Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification. *IEEE Transactions on Image Processing*, 17(4):622–629, April 2008.

[21] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene Selection for Cancer Classication using Support Vector Machines. *Machine Learning*, 46:389–422, 2002.

[22] M. Hall. Correlation Based Feature Selection for Machine Learning. *Doctoral Dissertation, University of Waikato, Department of Computer Science*, 1999.

[23] M. Hall. Correlation-Based Feature selection for discrete and numeric class machine learning. *In: Langley P, et al., eds. Proc. of the 17th Intl Conf. Machine Learning. San Francisco: Morgan Kaufmann Publishers*, pages 359–366, 2000.

[24] M. E. Hodgson, J. R. Jensen, J. A. Tullis, K. D. Riordan, and C. M. Archer. Synergistic Use of Lidar and Color Aerial Photography for Mapping Urban Parcel Imperviousness. *Photogrammetric Engineering and Remote Sensing*, 69(9):973–980, 2003.

[25] C. W. Hsu, C. C. Chang, and C. J. Lin. A practical guide to support vector classification. *Technical Report, Department of Computer Science and Information Engineering, University of National Taiwan, Taipei*, pages 1–12, 2003.

[26] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[27] A. Hyvarinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.

[28] R. Jensen. *Introductory digital image processing*. Pearson Prentice Hall, 2005.

[29] O. Jimenez, L. Rivera-Medina, E. Rodriguez-Diaz, E. Arzuaga-Cruz, and M. Ramirez-Velez. Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):844–851, April 2005.

[30] G. Joseph. *Fundamentals of Remote Sensing*. Universities Press,India, 2005.

[31] P. Julrode, S. Supratid, and U. Suksawatchon. A performance comparison of using PCA-based feature reduction and ant colony optimization with soft clustering approaches. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1391–1395, October 2012.

[32] M. Khodadadzadeh, A. Plaza, H. Ghassemian, and M. Bioucas-Dias. Spectral-Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization. *IEEE Transactions on Geoscience and Remote Sensing*, pages 6298–6314, October 2014.

[33] M. Khodadadzadeh, R. Rajabi, and H. Ghassemian. A novel approach for spectral-spatial classification of hyperspectral data based on SVM-MRF method. *IEEE Transactions on Geoscience and Remote Sensing*, pages 24–29, July 2011 doi:10.1109/IGARSS.2011.6049493.

[34] K. S. Khurshid, K. Staenz, L. Sun, R. Neville, H. P. White, A. Bannari, C. M. Champagne, and R. Hitchcock. Preprocessing of EO-1 Hyperion data. *Canadian Journal of Remote Sensing*, 32(2):84–97, 2006.

[35] Mathworks. Global Optimization Toolbox: User's Guide (r2011b), 2011.

[36] F. Melgani and L. Bruzzone. Classication of hyperspectral remote-sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, Aug 2004.

[37] Multispec. Hyperspectral Images. `https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html`.

[38] Multispec. Information on 220 Channel AVIRIS Data Set. `https://engineering.purdue.edu/~biehl/MultiSpec/aviris_documentation.html`.

[39] I. Nugraha, C. H. Ayuningtyas, and G. A. P. Saptawati. Performance analysis of relief and mRMR algorithm combination for selecting features in lupus Genome-Wide Association Study. *International Conference on Electrical Engineering and Informatics*, pages 1–5, July 2011.

[40] University of the Basque Country. Hyperspectral Remote Sensing Scenes. `http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes&redirect=no`.

[41] University of Tokyo. Remote sensing of environment and disaster. `http://stlab.iis.u-tokyo.ac.jp/~wataru/lecture/rsgis/rsnote/cp11/cp11-7.htm`.

[42] National Consortium on Remote Sensing in Transportation Environmental Assessment. Land Cover Classification Comparison with Different Resolution Imagery. `http://wwwghcc.msfc.nasa.gov/land/ncrst/multiclas.html`.

[43] M. Pal and G. M. Foody. Feature Selection for Classification of Hyperspectral Data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2297–2307, May 2010.

[44] M. Pal and P. M. Mather. Support vector classiers for land cover classication. *Map India, Image Processing and Interpretation*, 2003.

[45] I. Parshakov. Automatic class labeling of classified imagery using a hyperspectral library,. *M.Sc. thesis, University of Lethbridge*, December 2012.

[46] D. R. Peddle, P. M. Teillet, and M. A. Wulder. Radiometric image processing. In. *Remote Sensing of Forest Environments: Concepts and Case Studies*, 2003. eds. Wulder, M. A. and Franklin, S. E. 553. Boston, Massachusetts, USA: Kluwar Academic Publishers.

[47] H. Peng, F. Long, and C. Ding. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, August 2005.

[48] G. P. Petropoulos, C. Kalaitzidis, and K. P. Vadrevu. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers and Geosciences*, 41:99–107, April 2012.

[49] J. C. Platt. *Advances in Kernel Methods - Support Vector Learning*. MIT Press, January 1998.

[50] C. Rodarmel and J. Shan. Principal Component Analysis for Hyperspectral Image Classification. *Surveying and Land Information Systems*, 62(2):115–123, 2002.

[51] J. Schott. Remote sensing : the image chain approach. *New York: Oxford University Press*, page 666, 2007.

[52] G. A. Shaw and H. K. Burke. Spectral Imaging for Remote Sensing. *Lincoln Laboratory Journal*, 14(1):3–28, 2003.

[53] R. B. Smith. *Introduction to Hyperspectral Imaging*. MicroImages, Inc., 2012.

[54] ENVI software. Software - ITT Visual Information Solutions. Accessed Jan 07, 2011. `http://www.exelisvis.com/ProductServices/ENVI.aspx`.

[55] K. Staenz, J. Secker, B. C. Gao, C. Davis, and C. Nadeau. Radiative transfer codes applied to hyperspectral data for the retrieval of surface reflectance. *ISPRS Journal of Photogrammetry and Remote Sensing*, (57):194–203, 2002.

[56] K. Staenz, T. Szeredi, and J. Schwarz. A system for processing/analyzing hyperspectral data. *Canadian Journal of Remote Sensing*, 24(2):99–113, 1998.

[57] United States Geological Survey. Sensors - Hyperion. `http://eo1.usgs.gov/sensors/hyperion`.

[58] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson. SVM and MRF-Based Method for Accurate Classification of Hyperspectral Images. *IEEE Geoscience and Remote Sensing Letters*, 7(4):736–740, October 2010.

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

[60] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.

[61] Y. Wei and X. Wu. A New Fuzzy SVM based on the Posterior Probability Weighting Membership. *Journal of Computers*, 7(6), 2012.

[62] T. F. Wu, C. J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res*, pages 975–1005, Dec 2004.