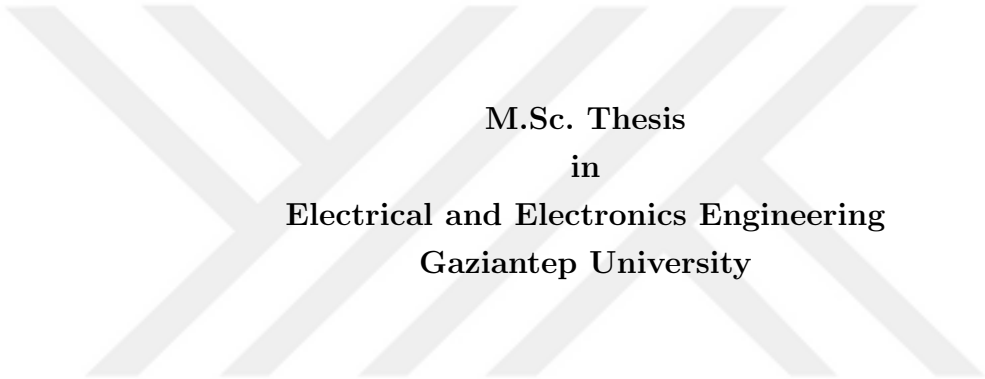


**DEVELOPMENT OF HYPERSPECTRAL
IMAGE CLASSIFICATION ALGORITHMS
FOR UNMANNED AERIAL VEHICLES**



**M.Sc. Thesis
in
Electrical and Electronics Engineering
Gaziantep University**

**Supervisor
Asst. Prof. Dr. Taner İNCE**

**by
Tuğcan DÜNDAR**

May 2019



© 2019 [Tuğcan DÜNDAR]

REPUBLIC OF TURKEY
GAZİANTEP UNIVERSITY
GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES
ELECTRICAL AND ELECTRONICS ENGINEERING

Name of the Thesis : Development of Hyperspectral Image Classification Algorithms for Unmanned Aerial Vehicles
Name of the Student : Tuğcan DÜNDAR
Exam Date : 24.05.2019

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Ahmet Necmeddin YAZICI
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Ergun ERÇELEBİ
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Taner İNCE
Supervisor

Examining Committee Members:

Signature

Prof. Dr. Gölge ÖĞÜCÜ YETKİN

.....

Asst. Prof. Dr. Hasari KARCI

.....

Asst. Prof. Dr. Taner İNCE

.....

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Tuğcan DÜNDAR

ABSTRACT

DEVELOPMENT OF HYPERSPECTRAL IMAGE CLASSIFICATION ALGORITHMS FOR UNMANNED AERIAL VEHICLES

DÜNDAR, Tuğcan

M.Sc. in Electrical and Electronics Engineering

Supervisor: Asst. Prof. Dr. Taner İNCE

May 2019

63 pages

Hyperspectral imaging is the measurement of light spectrum reflected from objects in many narrow wavelengths. High spectral resolution in hyperspectral images (HSIs) allows identification and discrimination of the land-cover materials. Therefore, hyperspectral imaging is used in various fields such as military, surveillance, mineralogy and agriculture. In these fields, classification of the pixels of an HSI is studied extensively. Sparse representation based classifiers have been a powerful tool for the classification purposes. These classifiers use the idea that the spectral pixels can be represented by only a few samples with same class label in a training dictionary. Recent studies have shown that use of spatial information in HSI in addition to spectral information increases the classification performance. In the scope of this thesis, two new sparse representation based classification methods which use both spectral and spatial information have been developed in order to increase the success of classification process. In the first study, multiscale superpixels (MSSs) are utilized to acquire spatial information in a local area using different region scales. Pixels in these areas are jointly classified by sparse representation classifier and then classification maps are formed. Guided filter (GF) is applied on these classification maps to improve the misclassifications near the edges. In the second study, the neighbor pixels having similar spectral characteristics with the test pixel are selected by spectral matching methods and others are ignored. To verify the feasibility of the proposed methods, the performance are evaluated over two widely used hyperspectral data sets. Experimental results demonstrate that the proposed algorithms exhibits good performance compared with other related methods in the literature.

Key words: Hyperspectral image, Sparse representation, Spatial-spectral, Classification.

ÖZET

İNSANSIZ HAVA ARAÇLARI İÇİN HİPERSPEKTRAL GÖRÜNTÜ SINIFLANDIRMA ALGORİTMALARININ GELİŞTİRİLMESİ

DÜNDAR, Tuğcan

Yüksek Lisans Tezi, Elektrik ve Elektronik Mühendisliği Bölümü

Tez Yöneticisi: Dr. Öğr. Üyesi Taner İNCE

Mayıs 2019

63 sayfa

Hiperspektral görüntüleme birçok dar dalgaboyunda nesnelerden yansıyan ışık spektrumunun ölçümüdür. Hiperspektral görüntülerdeki (HSI) yüksek spektral çözünürlük, arazi örtüsü maddelerinin tanımlanmasını ve ayırt edilmesini sağlar. Bu nedenle de hiperspektral görüntüleme askeri, gözetleme, mineraloji ve tarım gibi çeşitli alanlarda kullanılmaktadır. Bu alanlarda, bir HSI'nin piksellerinin sınıflandırılması yoğun olarak çalışılmaktadır. Seyrek gösterim tabanlı sınıflandırıcılar, sınıflandırma amaçları için güçlü bir araç olmuştur. Bu sınıflandırıcılar, spektral piksellerin bir eğitim sözlüğünde aynı sınıf etiketine sahip birkaç örnekle gösterilebileceği fikrini kullanır. Son çalışmalar, HSI'de uzamsal bilginin spektral bilgiye ek olarak kullanılmasının sınıflandırma performansını arttırdığını göstermiştir. Bu tez kapsamında, sınıflandırma işleminin başarısını arttırmak için, hem spektral hem de uzamsal bilgiyi kullanan iki yeni seyrek gösterim tabanlı sınıflandırma yöntemi geliştirilmiştir. İlk çalışmada, farklı bölge ölçeklerini kullanarak yerel bir alanda uzamsal bilgi elde etmek için çok ölçekli süperpikseller (MSS'ler) kullanılmıştır. Bu alanlardaki pikseller, seyrek gösterim sınıflandırıcısı ile ortaklaşa sınıflandırılır ve daha sonra sınıflandırma haritaları oluşturulur. Kenarlardaki yanlış sınıflandırmaları düzeltmek için bu sınıflandırma haritalarına rehberli filtre (RF) uygulanır. İkinci çalışmada, test pikseli ile benzer spektral özelliklere sahip olan komşu pikseller, spektral eşleştirme yöntemleri ile seçilmekte ve diğerleri gözardı edilmektedir. Önerilen yöntemlerin uygulanabilirliğini doğrulamak için, performans yaygın olarak kullanılan iki hiperspektral veri setinde değerlendirilmiştir. Deneysel sonuçlar, önerilen algoritmaların literatürdeki diğer ilgili yöntemlerle karşılaştırıldığında iyi bir performans sergilediğini göstermektedir.

Anahtar kelimeler: Hiperspektral görüntü, Seyrek gösterim, Uzamsal-spektral, Sınıflandırma.



I dedicate this work to my family

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Asst. Prof. Dr. Taner İNCE for his excellent guidance, caring, patience, encouragements and insight throughout this study.

I would also like to thank my family for supporting and encouraging me with their best wishes.

This work is supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK), through project 215E179. I would also like to thank TÜBİTAK for providing financial support throughout my graduate study.

TABLE OF CONTENTS

	Page
ABSTRACT	v
ÖZET	vi
ACKNOWLEDGEMENTS	viii
TABLE OF CONTENTS	ix
LIST OF TABLES	xii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xvi
CHAPTER 1: INTRODUCTION	1
1.1 Hyperspectral Imaging	1
1.2 Related Works	4
1.3 Contribution of the Thesis	6
1.4 Outline of the Thesis	7
CHAPTER 2: THEORETICAL BACKGROUNDS OF METH-	
ODS USED IN THE PROPOSED ALGORITHMS	8
2.1 Sparse Representation Classifier	8
2.2 Joint Sparse Representation Classifier	10
2.3 Principal Component Analysis	11
2.4 Entropy Rate Superpixels Segmentation	13
CHAPTER 3: DATA SETS AND PERFORMANCE EVALUA-	
TION METRICS	15
3.1 Data Sets	15
3.1.1 Indian Pines	15
3.1.2 University of Pavia	16
3.2 Performance Evaluation Metrics	18

CHAPTER 4: SPARSE REPRESENTATION-BASED HYPER-SPECTRAL IMAGE CLASSIFICATION USING MULTISCALE SUPERPIXELS AND GUIDED FILTER	20
4.1 Methodology and Formulation of the Proposed Method	21
4.1.1 Forming Different-scaled Superpixel Segmentations	21
4.1.2 Sparse Representation Based Multiscale Superpixels	21
4.1.3 Guided Filter	22
4.1.4 Problem Formulation	23
4.1.5 Determination of Class Label	24
4.2 Simulation Setup and Experimental Results	24
4.2.1 Simulation Results on Indian Pines	25
4.2.2 Simulation Results on University of Pavia	30
CHAPTER 5: SPARSE REPRESENTATION-BASED CLASSIFICATION OF HYPERSPECTRAL IMAGES USING SPECTRAL MATCHING METHODS	38
5.1 Spectral Matching Methods	38
5.1.1 Euclidean Distance	38
5.1.2 Spectral Angle Mapper	39
5.1.3 Pearson Correlation Coefficient	39
5.2 Formulation of the Proposed 3SM-JRC Method	40
5.3 Simulation Setup and Experimental Results	42
5.3.1 Simulation Results on Indian Pines	42
5.3.2 Simulation Results on University of Pavia	46
CHAPTER 6: CONCLUSION AND FUTURE WORK	51
6.1 Conclusion	51
6.2 Future Work	52
REFERENCES	53
APPENDIX	60
APPENDIX A SELECTION OF THE REGION SCALES	

FOR THE PROPOSED MSS–GF ALGORITHM	60
A.1 University of Pavia	60
A.2 Indian Pines	61
CURRICULUM VITAE	63



LIST OF TABLES

	Page
Table 3.1 The class names and the number of samples of the Indian Pines.	16
Table 3.2 The class names and the number of samples of the University of Pavia.	17
Table 3.3 An example of two class confusion matrix table.	18
Table 4.1 Number of superpixels in each region scale for the Indian Pines.	25
Table 4.2 Classification accuracies (%) of Indian Pines dataset obtained by SVM, EPF, JSRC, MSS–MV and the proposed MSS–GF method.	28
Table 4.3 Number of superpixels in each region scale for the University of Pavia.	31
Table 4.4 Classification accuracies (%) of University of Pavia dataset obtained by SVM, EPF, JSRC, MSS–MV and the proposed MSS–GF method.	34
Table 5.1 The number of train and test samples for each class used in the experiments for the Indian Pines data set.	43
Table 5.2 Classification accuracies (in %) obtained by the SVM, SRC, JSRC, NLW–JSRC and the proposed 3SM–JSRC methods for the Indian Pines.	45
Table 5.3 The number of train and test samples for each class used in the experiments for the University of Pavia data set. . .	47

Table 5.4	Classification accuracies (in %) obtained by the SVM, SRC, JSRC, NLW-JSRC and the proposed 3SM-JSRC methods for the University of Pavia.	49
Table A.1	Classification accuracies (%) of different region scales for the University of Pavia.	60
Table A.2	Classification accuracies (%) of different region scales for the Indian Pines.	61



LIST OF FIGURES

	Page
Figure 1.1 Hyperspectral remote sensing	2
Figure 1.2 Hyperspectral data cube	2
Figure 2.1 Steps of the principal component analysis	14
Figure 3.1 Indian Pines; (a) False color image, (b) Ground-truth map and (c) Class names.	16
Figure 3.2 University of Pavia; (a) False color image, (b) Ground- truth map and (c) Class names.	17
Figure 4.1 Effect of region scales and sparsity-level on the Indian Pines; (a) OA and (b) AA.	26
Figure 4.2 Superpixel segmentation maps of the Indian Pines for the region scales (a) 3×3 , (b) 5×5 , (c) 7×7 , (d) 9×9 and (e) 11×11	27
Figure 4.3 Effect of guided filter radius r for the Indian Pines.	27
Figure 4.4 Effect of different training percents on (a) OA and (b) AA using SVM, JSRC, EPF, MSS-MV and proposed MSS- GF method for the Indian Pines.	29
Figure 4.5 Indian Pines; (a) Guidance image, (b) Ground truth. The classification maps obtained by (c) SVM, (d) JSRC, (e) EPF, (f) MSS-MV and (g) MSS-GF.	30
Figure 4.6 Single-scale and multiscale superpixels comparison for the Indian Pines.	31
Figure 4.7 Effect of region scales and sparsity-level on the University of Pavia; (a) OA and (b) AA.	32

Figure 4.8	Superpixel segmentation maps of the University of Pavia for the region scales (a) 7×7 , (b) 9×9 , (c) 11×11 , (d) 13×13 , (e) 15×15 and (f) 17×17	33
Figure 4.9	Effect of guided filter radius r for the University of Pavia.	34
Figure 4.10	Effect of different training percents on (a) OA and (b) AA using SVM, JSRC, EPF, MSS-MV and proposed MSS-GF method for the University of Pavia.	35
Figure 4.11	University of Pavia; (a) Guidance image, (b) Ground truth. The classification maps obtained by (c) SVM, (d) JSRC, (e) EPF, (f) MSS-MV and (g) MSS-GF.	36
Figure 4.12	Single-scale and multiscale superpixels comparison for the University of Pavia.	37
Figure 5.1	Selection of the neighbor pixels within the window created around the test pixel by the methods ED, SAM and PCC.	41
Figure 5.2	The results obtained using different region scales and sparsity levels for the Indian Pines data set; (a) OA, (b) AA.	44
Figure 5.3	Indian Pines data set: (a) Ground truth; Classification maps and OA results obtained by the (b) SVM, (c) SRC, (d) JSRC, (e) NLW-JSRC, (f) 3SM-JSRC methods.	46
Figure 5.4	The results obtained using different region scales and sparsity levels for the University of Pavia data set; (a) OA, (b) AA.	48
Figure 5.5	University of Pavia data set: (a) Ground truth; Classification maps and OA results obtained by the (b) SVM, (c) SRC, (d) JSRC, (e) NLW-JSRC, (f) 3SM-JSRC methods.	50

LIST OF ABBREVIATIONS

AA	Average Accuracy
AVIRIS	Airborne Visible Infrared Imaging Spectrometer
CA	Class Accuracy
CC-JSRC	Correlation Coefficient Joint Sparse Representation Classifier
DT	Decision Trees
ED	Euclidean Distance
EPF	Edge Preserving Filtering
ERS	Entropy Rate Superpixels
GF	Guided Filter
HSI	Hyperspectral Image
JSRC	Joint Sparse Representation Classifier
KNN	k -Nearest Neighbor
MASR	Multiscale Adaptive Sparse Representation
MSS-GF	Multi-scale Superpixels Guided Filter
MSS-MV	Multi-scale Superpixels Majority Voting
MV	Majority Voting
NLS-SCC	Non-local Similarity-based Sparse Coding Classifier
NLW-JSRC	Non-local Weighted Joint Sparse Representation Classifier
OA	Overall Accuracy
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient

RBFB	Radial Basis Function
RF	Random Forests
RGB	Red–Green–Blue
ROSI	Reflective Optics System Imaging Spectrometer
SAJSRC	Shape Adaptive Joint Sparse Representation Classifier
SAM	Spectral Angle Mapper
SBDSM	Superpixel–Based Discriminative Sparse Model
SMTJSRC	Superpixel Multi Task Joint Sparse Representation Classifier
SOMP	Simultaneous Orthogonal Matching Pursuit
SRC	Sparse Representation Classifier
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle

CHAPTER 1

INTRODUCTION

In this chapter, we first give a brief information about hyperspectral imaging and its applications. Then, related works about HSI classification in the literature are discussed. After that, we explain how this thesis contributes to the literature in terms of HSI classification. At the end, we present the outline of the thesis.

1.1 Hyperspectral Imaging

All substances in the nature absorb, emit and reflect electromagnetic energy at certain wavelengths due to the characteristics of the molecular structures [1]. Hyperspectral imaging uses this basic principle. Unlike the classical red–green–blue (RGB) imaging technique, which covers only the visible portion of the electromagnetic spectrum, hyperspectral imaging covers a broad range of this spectrum. It starts from the visible region ($0.4\text{--}0.7\ \mu\text{m}$) and goes up to near-infrared ($\sim 2.4\ \mu\text{m}$) [2]. In addition to the wide spectrum range, HSIs contain lots of narrow-placed spectral bands which is usually more than 100. To cover these spectrums, expensive hyperspectral sensors are used. They are generally mounted on an aircraft or an unmanned aerial vehicle (UAV) to collect hyperspectral data [3]. The sensor measures the light reflected from the objects on the ground over a wide spectral bands. These measured values by the sensor are called radiance and these radiance characteristics against the wavelengths are called spectral signatures of the objects. This property can be used to classify or identify land-cover materials on the ground. Figure 1.1 shows a visual about the hyperspectral imaging.

A pixel of an HSI can be thought as a vector whose length is equal to the number of spectral bands [4]. Since HSI contains lots of pixels, it can be visualized as a data cube which is shown in Figure 1.2. The front face represents spatial dimension and the depth represent spectral dimension. Due to the high spectral resolution, different materials can be separated from each other. So, hyperspectral

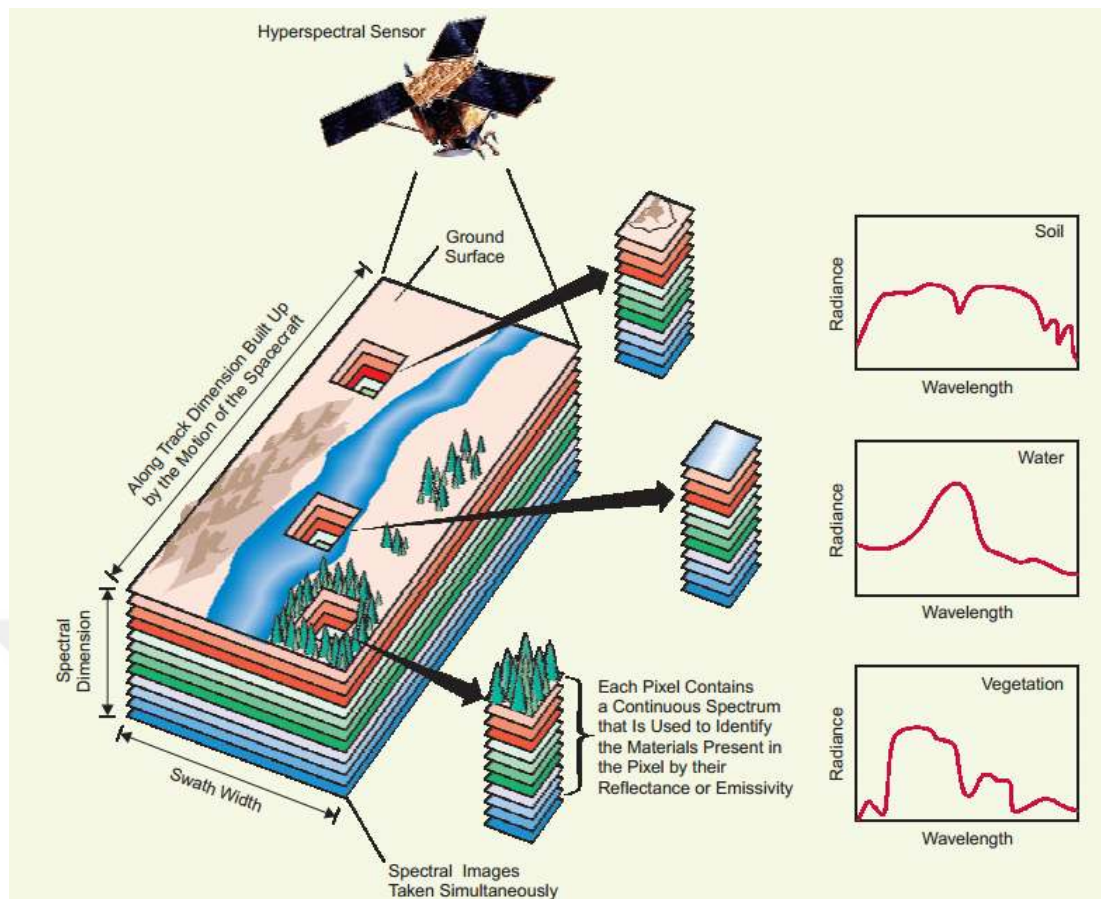


Figure 1.1 Hyperspectral remote sensing [2].

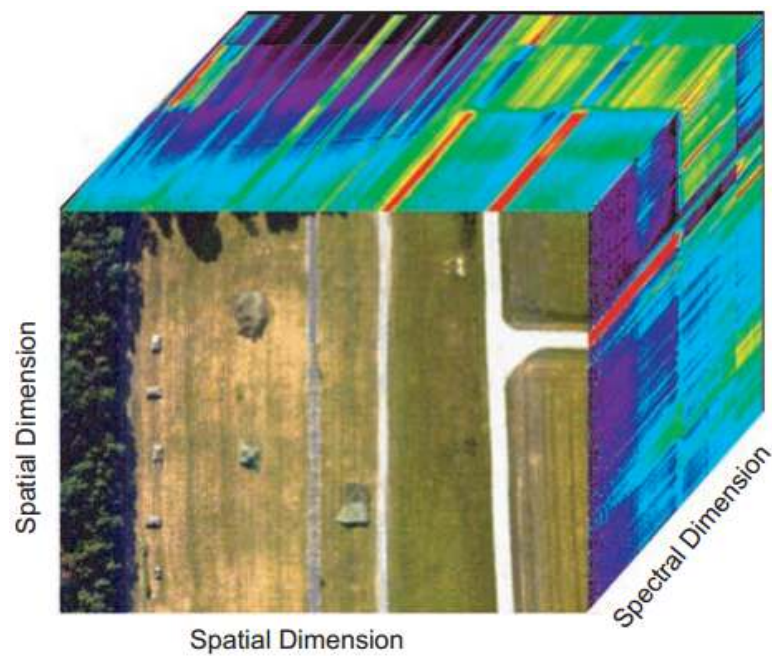


Figure 1.2 Hyperspectral data cube [2].

imaging is used in many applications such as military, agriculture, environmental monitoring, mineral exploration, security and defense [5, 6]. In these applications, topics such as classification [7–9], target detection [10–12], anomaly detection [13–15] and unmixing [16–18] are studied extensively by researchers. These topics can be briefly summarized as follows:

Classification: An HSI is composed of different land-cover regions. It means that image has different classes and so each pixel should belong to a specific class. The aim of the classification is to give a unique label to each pixel using an appropriate classification algorithm [19].

Target detection: The aim is to detect man-made substances in natural environments for the purposes such as illegal planting, defense and rescue activities [1].

Anomaly detection: Determination of changes after natural disasters. These type of pixels have different spectral characteristic than their local background pixels [2].

Unmixing: Due to the atmospheric effects and sensor with low spatial resolution, some pixels consist of spectral signatures of different materials. These pixels are called mixed pixels. If a mixed pixel occur in the image, it is not possible to identify the materials inside directly from the measured spectral signatures. So, unmixing is applied on the image to identify these materials [20].

Development of HSI classification algorithms has been the most popular one among these topics in recent years [21]. Classification can be divided into three categories [19]. The first one is called unsupervised classification which predicts the labels of pixels using a software without any pre-defined samples (also called training samples). The software clusters the pixels using their spectral behaviours. The second one is called semi-supervised classification which utilizes both a few labeled training samples and unlabeled samples to determine the labels. The third one is called supervised classification which predicts the label of the pixels using pre-defined training samples with their class labels.

According to the usage of information in HSI, classifiers can be divided into two categories called spectral classifiers (or pixel-wise classifiers) and spatial-spectral classifiers [22]. Spectral classifiers only use the information of a single pixel and ignores the spatial information in HSI. On the other hand, spatial-spectral classifiers incorporate the spatial and spectral information in HSI. Spatial information means that spectral characteristics of the neighbor pixels are similar and so they

most likely belong to the same class. Spatial information of an HSI can be incorporated into a classification problem by pre-processing (such as mathematical morphology [23]) or post-processing operations (such as morphological watershed transformation [23]) [20].

1.2 Related Works

HSI classification has become very popular among researchers in recent years. In the first studies, only spectral information was used in the classification process. For example, a machine learning algorithm called support vector machine (SVM) was used in [24–26] to classify test pixels. SVM was first studied for binary classification problem [27]. Given a set of training samples, SVM finds an optimal hyperplane that maximizes the margin between classes. If the data are not linearly separable, SVM uses some kernel tricks to project the data to a higher-dimensional space where the data is linearly separable here. When it comes to separate a data with multiclass, SVM uses two procedures called one versus one and one versus all strategies [27]. In the one versus one method, SVM trains a binary classifier for each distinct pair of labels. If the data have C different classes in total, then $\frac{C(C-1)}{2}$ classifiers will be used in the one versus one method. On the other hand, SVM trains a binary classifier for each class in the one versus all method. So, C binary classifiers are used to separate C distinct classes. Another machine learning algorithm to be used to classify HSI is the k -nearest neighbor (KNN) [28]. This algorithm seeks k -nearest training samples of the test sample because it assumes that similar pixels are found in nearby locations. After finding k -nearest training samples, label of the test pixel is assigned as the most repetitive class among the labels of those k -nearest training samples. Decision trees (DT) was proposed in [29] to classify HSI into different classes and have investigated in researches such as [30, 31]. DT classifier divides the classification problem into sub-stages and then performs decision process. DT classify samples by sorting their feature values. Each node represents a feature of a sample and each branch symbolizes a value that the node can support. A sample can be classified by beginning at the root node and then sorted based on their feature values. Random forests (RF) was also introduced for the classification of HSI in [32, 33]. It can be thought as a special case of DT. However, RF is formed by many classifiers so it is an ensemble classifier. Ensemble classifiers means that the result of the classification process is determined by majority voting. Since outputs of the many classifiers are combined, RF provides better classification performance than DT.

In recent years, sparse representation based approaches have begun to be applied to HSI processing in the research areas such as denoising [34, 35], unmixing [36, 37], target detection [38, 39] due to the performance in the areas such as signal and image processing [40, 41]. In addition, sparse representation has been also used for the classification purposes of HSIs [42–45]. Chen *et. al* investigated sparse representation classifier (SRC) in [42] to assign a unique label to each test pixel. SRC assumes that a test pixel can be linearly represented by a few training samples in a structured dictionary which consists of training samples from each class.

All the methods mentioned so far have used only the spectral information in HSI. Recent studies have shown that use of spatial information in addition to spectral information improve the performance of the classifiers in terms of classification accuracy [42, 46–52]. Sparse representation based HSI classification methods using spatial information have been studied quite a lot in recent years. For example, [42] proposed joint sparse representation classifier (JSRC) which constructs a fixed-size square window around the test pixel to obtain spatial information. JSRC assumes that neighbor pixels share a set of common atoms from the structured dictionary and so they can be represented simultaneously. However, the optimal window size may change from pixel to pixel because some of the pixels may have large neighbor areas and the others may have small neighbor areas. Also, pixels from other classes can also be found in the window, which will reduce the classification results. To overcome these problems, different methods have been proposed [53–59]. In [53] (multiscale adaptive sparse representation, MASR), different region scales were constructed around the test pixel to adaptively utilize the spatial information because different regions complements the shortcomings of each others. In [54] (shape-adaptive joint sparse representation classifier, SAJSRC), similar neighbor pixels around the test pixel were selected using shape-adaptive algorithm [60] and shape-adaptive regions were constructed for each test pixel. Then, pixels in these regions were represented simultaneously. Although the proposed MASR and SAJSRC methods provide better classification accuracies than the JSRC method, their computation time is longer compared to JSRC. This is expected result since they classify only the center pixel (or test pixel). In order to overcome this problem, superpixel-based methods have been proposed [55, 56]. A superpixel is a group of pixels that have similar spectral characteristics. Its shape and size vary for distinct spatial structure in HSI. In [55] (superpixel based discriminative sparse model, SBDSM), HSI was divided into many superpixels and each superpixel was represented simultaneously and then labeled at the same time instead of a single pixel. In [56] (superpixel multitask

joint sparse representation classifier, SMTJSRC), authors exploited superpixel segmentation and multiple feature fusion and then labeled each superpixel. In [57] (non-local weighted joint sparse representation classifier, NLW-JSRC), different weights were assigned to each pixel located around the test pixel using a non-local weighting mechanism studied in [61–63]. This method compares the similarities between the test pixel and its neighbors. Also, it compares the geometrical shape of the neighborhood area. In this way, neighbor pixels similar to the test pixel have larger weight than dissimilar ones and so will contribute more during classification. In [58] (correlation coefficient joint sparse representation classifier, CC-JSRC), authors calculated the similarities between the training samples and test sample using correlation coefficient and assigned these weights to the representation residual. In this way, training samples belonging to the same class with the test sample will have more weight and this process will improve the accuracy results of the JSRC method. In [59] (non-local similarity-based sparse coding classifier, NLS-SCC), like the NLW-JSRC algorithm, a non-local weighting was used to assign different weights to each neighbor pixel around the test pixel. The non-local weights were calculated by the use of spectral angle mapper measurement which calculates the angle between two observations.

1.3 Contribution of the Thesis

In the scope of this thesis, two sparse representation based supervised HSI classification algorithms, which use both spatial and spectral information, have been proposed. The first algorithm uses multiscale superpixels to utilize spatial information effectively and then the created classification map is enhanced using an edge filtering process. In this way, instead of labeling one pixel at a time, pixels in a superpixels are labeled together. Also, spatial information is fully considered by the use of multiscale superpixels and guided filter. The second algorithm selects the neighbors of the test sample in a fixed-size square window using some spectral matching methods. In this way, pixels with different spectral characteristics from the test pixel are discarded and the remaining ones are jointly represented with the test pixel. The proposed methods have improved the results of the some existing classifiers and have brought a new perspective to the literature.

1.4 Outline of the Thesis

This chapter presents the basics of hyperspectral imaging and literature survey about HSI classification. In Chapter 2, theoretical backgrounds and details of the SRC and JSRC algorithms are presented. In addition, some methods used in the proposed algorithms such as principal component analysis (PCA) and entropy rate superpixel (ERS) are also described in this section. Chapter 3 gives information about data sets used in the experiments. Chapter 4 and 5 explains the proposed algorithms in detail, respectively. Their mathematical backgrounds are first described and then results of the experiments are given. Chapter 6 concludes the thesis and we give some suggestions for the future works in this chapter.



CHAPTER 2

THEORETICAL BACKGROUNDS OF METHODS USED IN THE PROPOSED ALGORITHMS

2.1 Sparse Representation Classifier

In the SRC method, a test pixel $\mathbf{p} \in \mathbb{R}^{B \times 1}$ can be sparsely and linearly represented by some of the dictionary atoms from the same class in a B dimensional space. The training dictionary is represented as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] \in \mathbb{R}^{B \times N}$, where each subdictionary $\{\mathbf{X}_m\}_{m=1,2,\dots,M}$ contains train samples from class m , M is the number of classes and N is the number of training samples. Since the aim is to find the sparse vector $\boldsymbol{\alpha}$ which satisfies $\mathbf{p} = \mathbf{X}\boldsymbol{\alpha}$, it is required to solve the following optimization problem

$$\hat{\boldsymbol{\alpha}} = \arg \min \|\boldsymbol{\alpha}\|_0 \quad s.t. \quad \mathbf{p} = \mathbf{X}\boldsymbol{\alpha} \quad (2.1)$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T \in \mathbb{R}^{N \times 1}$ is the sparse coefficient vector and $\|\boldsymbol{\alpha}\|_0$ counts the number of nonzero elements in $\boldsymbol{\alpha}$. If we consider some approximation errors $\mathbf{n}$ in the empirical data (i.e., $\mathbf{p} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{n}$) and insert it into the optimization problem in (2.1), then the equality constraint becomes an inequality constraint as follow

$$\hat{\boldsymbol{\alpha}} = \arg \min \|\boldsymbol{\alpha}\|_0 \quad s.t. \quad \|\mathbf{p} - \mathbf{X}\boldsymbol{\alpha}\|_2 \leq \varepsilon \quad (2.2)$$

ε refers to error tolerance term and $\|\cdot\|_2$ is the Euclidean norm defined as the square root of the sum of the squares of the entries in a vector. The aforementioned optimization problem in (2.2) can also be thought as minimizing the error under a specified sparsity level, so the problem can be written as

$$\hat{\boldsymbol{\alpha}} = \arg \min \|\mathbf{p} - \mathbf{X}\boldsymbol{\alpha}\|_2 \quad s.t. \quad \|\boldsymbol{\alpha}\|_0 \leq L \quad (2.3)$$

the parameter L is the upper limit of the sparsity level. The orthogonal matching pursuit (OMP) algorithm can be used to solve the problem in (2.3). The OMP is a greedy optimization algorithm whose steps are described in Algorithm 1.

After the sparse coefficient vector $\hat{\alpha}$ is calculated, the residual which is difference between the original signal and the approximated signal should be calculated for each class as;

$$E_m(\mathbf{p}) = \|\mathbf{p} - \mathbf{X}_m \hat{\alpha}_m\|_2 \quad \text{where } m = 1, 2, \dots, M \quad (2.4)$$

Finally, the class label of the test pixel $\mathbf{p}$ is determined by the class with minimum residual value as follow:

$$Class(\mathbf{p}) = \arg \min(E_m) \quad (2.5)$$

Algorithm 1: OMP

Input: Test pixel $\mathbf{p} \in \mathbb{R}^{B \times 1}$

Sparsity level L

Training dictionary $\mathbf{X} \in \mathbb{R}^{B \times N} = [x_1, x_2, \dots, x_N]$

{Normalize columns of $\mathbf{p}$ and $\mathbf{X}$ to unit norm}

Initialize: Residual $\mathbf{r}_0 = \mathbf{p}$

Index set $\Phi_0 = \emptyset$

Iteration counter $k = 1$

for $k \leq L$

1. Find the index λ_k of the atom that matches best with the residual

$$\lambda_k = \arg \max_{i=1,2,\dots,N} |\langle \mathbf{r}_{k-1}, x_i \rangle|$$

2. Update the index set $\Phi_k = \Phi_{k-1} \cup \{\lambda_k\}$

3. Compute a new estimated signal using least squares

$$\begin{aligned} \hat{\mathbf{s}}_{\Phi_k} &= \arg \min_{\mathbf{s}_{\Phi_k}} \|\mathbf{p} - \mathbf{X}_{\Phi_k} \mathbf{s}_{\Phi_k}\|_2 \\ &= (\mathbf{X}_{\Phi_k}^T \mathbf{X}_{\Phi_k})^{-1} \mathbf{X}_{\Phi_k}^T \mathbf{p} \end{aligned}$$

4. Determine the new residual $\mathbf{r}_k = \mathbf{p} - \mathbf{X}_{\Phi_k} \hat{\mathbf{s}}_{\Phi_k}$

5. $k \leftarrow k + 1$

end

Output: Sparse coefficient vector α that has nonzero rows indexed by Φ

and value of $(\mathbf{X}_{\Phi}^T \mathbf{X}_{\Phi})^{-1} \mathbf{X}_{\Phi}^T \mathbf{p}$

2.2 Joint Sparse Representation Classifier

In the SRC model, the classification process is performed on a single test pixel and thus the spatial information in HSI is neglected. In the JSRC model, a fixed size square window is formed around a test pixel. In this way, the test pixel and its neighborhoods are used in the classification stage to improve the accuracy of classification results. Let $\mathbf{p}_1$ denotes a test pixel and $\{\mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_Q\}$ denotes the neighbors of $\mathbf{p}_1$, respectively. Let $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_Q] \in \mathbb{R}^{B \times Q}$ denotes a concatenated matrix including the test pixel and its neighbors. In the JSRC method, it is assumed that the test pixel and its neighbors share a joint sparsity pattern and they can be represented by sparse linear combination of several joint atoms from a dictionary $\mathbf{X}$ as in the case of SRC model. Thus, it is required to find the sparse coefficient matrix $\mathbf{A}$ that satisfies $\mathbf{P} = \mathbf{XA}$ by solving the following optimization problem

$$\hat{\mathbf{A}} = \arg \min \|\mathbf{A}\|_0 \quad s.t. \quad \mathbf{P} = \mathbf{XA} \quad (2.6)$$

where $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_Q] \in \mathbb{R}^{N \times Q}$ is the sparse coefficient matrix and $\|\mathbf{A}\|_0$ counts the number of nonzero elements in $\mathbf{A}$. If we consider some approximation errors $\mathbf{N}$ in the empirical data (i.e., $\mathbf{P} = \mathbf{XA} + \mathbf{N}$) as in the case of SRC, the optimization problem in (2.6) becomes as

$$\hat{\mathbf{A}} = \arg \min \|\mathbf{A}\|_0 \quad s.t. \quad \|\mathbf{P} - \mathbf{XA}\|_F \leq \varepsilon \quad (2.7)$$

ε refers to error tolerance term and $\|\cdot\|_F$ is the frobenius norm defined as the square root of the sum of the squares of the each entry in a matrix. The aforementioned optimization problem in (2.7) can also be thought as minimizing the error under a specified sparsity level, so the problem can be written as

$$\hat{\mathbf{A}} = \arg \min \|\mathbf{P} - \mathbf{XA}\|_F \quad s.t. \quad \|\mathbf{A}\|_0 \leq L \quad (2.8)$$

the parameter L is the upper limit of the sparsity level. The simultaneous orthogonal matching pursuit (SOMP) algorithm can be used to solve the problem in (2.8). The SOMP is a commonly used greedy optimization algorithm whose steps are described in Algorithm 2.

After the sparse coefficient matrix $\hat{\mathbf{A}}$ is calculated, the residual which is difference between the original signal and the approximated signal should be calculated for each class as

$$E_m = \left\| \mathbf{P} - \mathbf{X}_m \hat{\mathbf{A}}_m \right\|_F \quad ; \quad m = 1, 2, \dots, M \quad (2.9)$$

Finally, the class label of the test pixel $\mathbf{p}_1$ is determined by the class with minimum residual value as follow:

$$Class(\mathbf{p}_1) = \arg \min (E_m) \quad (2.10)$$

Algorithm 2: SOMP

Input: Pixel groups $\mathbf{P} \in \mathbb{R}^{B \times Q}$

Sparsity level L

Training dictionary $\mathbf{X} \in \mathbb{R}^{B \times N} = [x_1, x_2, \dots, x_N]$

{Normalize columns of $\mathbf{P}$ and $\mathbf{X}$ to unit norm}

Initialize: Residual $\mathbf{R}_0 = \mathbf{P}$

Index set $\Phi_0 = \emptyset$

Iteration counter $k = 1$

for $k \leq L$

1. Find the index λ_k of the atom that matches best with the residual

$$\lambda_k = \arg \max_{i=1,2,\dots,N} |\langle \mathbf{R}_{k-1}^t, x_i \rangle|$$

2. Update the index set $\Phi_k = \Phi_{k-1} \cup \{\lambda_k\}$

3. Compute a new estimated signal using least squares

$$\begin{aligned} \hat{\mathbf{S}}_{\Phi_k} &= \arg \min_{\mathbf{S}_{\Phi_k}} \|\mathbf{P} - \mathbf{X}_{\Phi_k} \mathbf{S}_{\Phi_k}\|_F \\ &= (\mathbf{X}_{\Phi_k}^T \mathbf{X}_{\Phi_k})^{-1} \mathbf{X}_{\Phi_k}^T \mathbf{P} \end{aligned}$$

4. Determine the new residual $\mathbf{R}_k = \mathbf{P} - \mathbf{X}_{\Phi_k} \hat{\mathbf{S}}_{\Phi_k}$

5. $k \leftarrow k + 1$

end

Output: Sparse coefficient matrix $\mathbf{A}$ that has nonzero rows indexed by Φ

and value of $(\mathbf{X}_{\Phi}^T \mathbf{X}_{\Phi})^{-1} \mathbf{X}_{\Phi}^T \mathbf{P}$

2.3 Principal Component Analysis

PCA is a dimension reduction method which transforms the original higher dimensional space to a lower dimensional space. Since the dimension of HSI is very large, PCA can be used to reduce the number of these bands. The contiguous bands of hyperspectral images have similar spectral characteristics, so they carry

similar information about an object. The PCA utilizes some mathematical and statistical properties to analyze these correlations of the bands and then to reduce the dimension. The PCA is mainly based on the eigenvalue decomposition of the covariance matrix of the bands. The steps of the PCA can be written as;

a) Three dimensional HSI $\mathcal{H}$ is first converted to two dimensional matrix form $\mathbf{H}$.

$$\mathcal{H} \in \mathbb{R}^{M \times N \times B} \rightarrow \mathbf{H} \in \mathbb{R}^{B \times W} \quad (2.11)$$

where W stands for the number of total pixels (MN) in the HSI. After the conversion process, a pixel vector can be written in the form of

$$\mathbf{p}_i = [p_1, p_2, \dots, p_B]^T \quad \text{where } i = 1, 2, \dots, W \quad (2.12)$$

the operator $[\cdot]^T$ refers to transpose of the vector.

b) The mean, $\mathbf{m}$, of the all pixels is calculated and then the mean-centered data matrix $\mathbf{U}$ is calculated as follows

$$\mathbf{m} = \frac{1}{W} \sum_{i=1}^W \mathbf{p}_i \quad (2.13)$$

$$\mathbf{U} = [\mathbf{p}_1 - \mathbf{m} \quad \mathbf{p}_2 - \mathbf{m} \quad \dots \quad \mathbf{p}_i - \mathbf{m}] \in \mathbb{R}^{B \times W} \quad (2.14)$$

Mean centering is an essential process before the calculation of covariance matrix to produce a data matrix whose mean is equal to 0.

c) To analyze the correlations between spectral bands, the covariance matrix $\mathbf{Cov}$ is computed as

$$\mathbf{Cov} = \mathbf{U}\mathbf{U}^T \quad (2.15)$$

$\mathbf{Cov}$ is a $B \times B$ dimensional matrix whose diagonal entries show the variance of each band and off-diagonal entries show covariance between bands.

d) After the covariance matrix is found, the eigenvalue decomposition of the covariance matrix is found and takes the form of

$$\mathbf{Cov} = \mathbf{V}\mathbf{D}\mathbf{V}^T \quad (2.16)$$

$\mathbf{V}$ refers to eigenvectors of the $\mathbf{Cov}$, $\mathbf{V}^T$ refers to transpose of $\mathbf{V}$ and diagonal matrix $\mathbf{D}$ is the eigenvalues of the $\mathbf{Cov}$.

e) After all the eigenvalues and their corresponding eigenvectors are found, eigenvalues and their eigenvectors are sorted in an descending order. Then, first K ($K \leq B$) eigenvectors can be used to approximate $\mathbf{H}$. It is mentioned in [64] that

first few principal components can be used to approximate the original image because these components have highest variance and contains lots of the information about HSI. The pixels of the approximated image can be shown as

$$\mathbf{z}_i = [z_1, z_2, \dots, z_B]^T = \mathbf{V}^T \mathbf{p}_i \quad \text{where } i = 1, 2, \dots, W \quad (2.17)$$

To select only first K principal components, the following equation should be calculated.

$$\mathbf{z}_i = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_K \end{bmatrix}_i = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \dots & \mathbf{V}_{1K} & \dots & \mathbf{V}_{1B} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \dots & \mathbf{V}_{2K} & \dots & \mathbf{V}_{2B} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{V}_{K1} & \mathbf{V}_{K2} & \dots & \mathbf{V}_{KK} & \dots & \mathbf{V}_{KB} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \\ \vdots \\ p_B \end{bmatrix} \quad (2.18)$$

This step is done for the all pixels in the image, then the new image is formed by selecting only first K principal components. If the approximated image is denoted as $\mathbf{H}_{new}$, then it is needed to convert it from two dimension to three dimension as follow:

$$\mathbf{H}_{new} \in \mathbb{R}^{K \times W} \rightarrow \mathcal{H}_{new} \in \mathbb{R}^{M \times N \times K} \quad (2.19)$$

All steps of the PCA method can be seen in Figure 2.1.

2.4 Entropy Rate Superpixels Segmentation

Superpixel segmentation is a significant process to extract local homogeneous regions from an image. ERS is a superpixel segmentation method that uses entropy rate on a graph and a balancing term. Local homogeneous regions are extracted with the help of entropy rate. Thanks to the balancing term, it is ensured that these regions have similar sizes. So, ERS uses the following objective function

$$\max_A H(A) + \lambda B(A) \quad s.t. \quad A \subseteq E \quad (2.20)$$

where $H(A)$ is the entropy rate function, $B(A)$ is the balancing term, λ controls the tradeoff between the entropy rate function and the balancing term. E refers to edge set that indicates similarities between neighbor pixels and A is the subset of edge set.

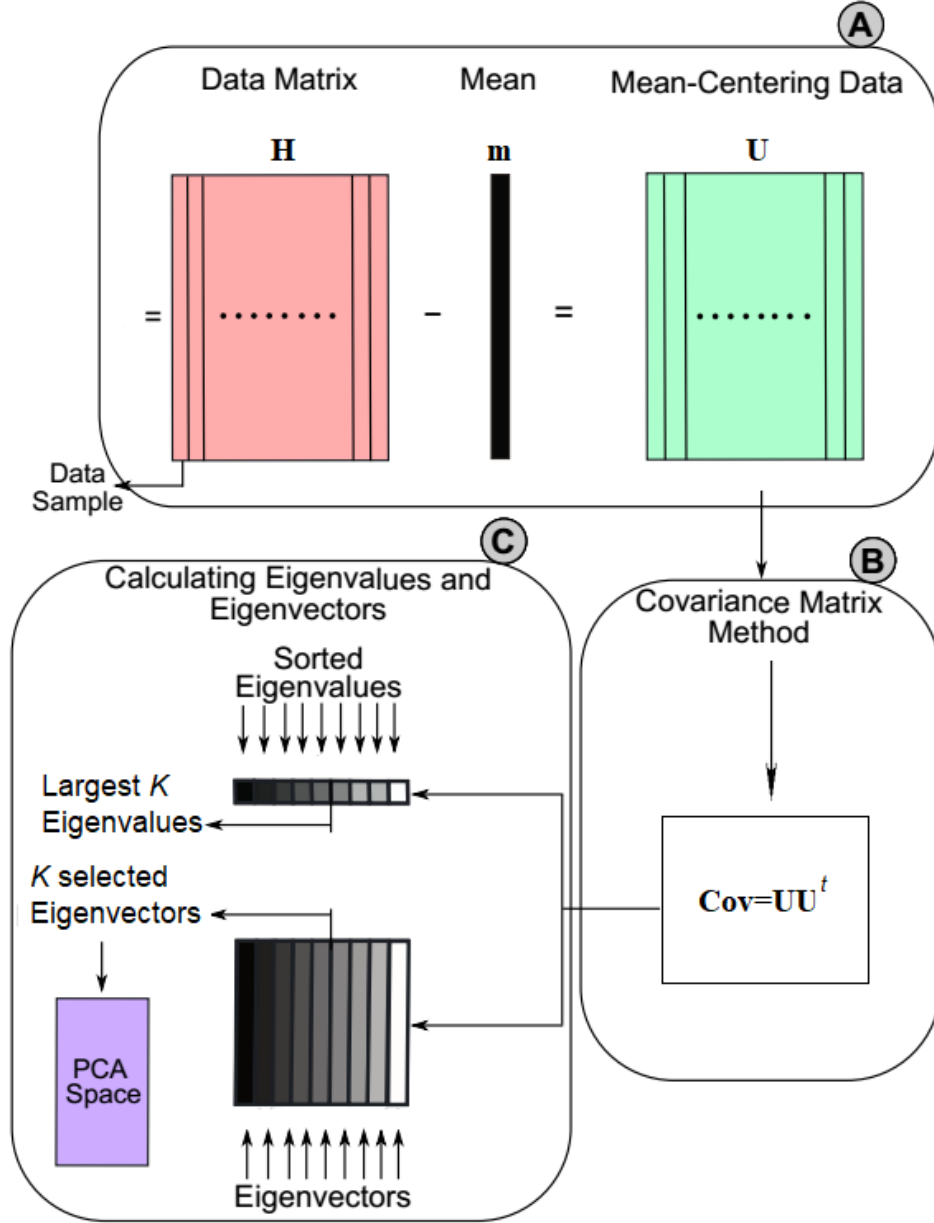


Figure 2.1 Steps of the principal component analysis [65].

ERS method needs a base image to generate superpixels. A graph $G = (V, E)$ is constructed using the base image. V refers to vertex set composed of pixels of the base image. To segment the graph into S superpixels, ERS chooses a subset of edges $A \subseteq E$. Firstly, the algorithm starts with an empty set of edges. Then, it chooses the edge from E to maximize the objective function. In this way, new edges are added to subset A in each iteration and the algorithm is terminated when the predefined number of superpixels is reached to the number of bounded subgraphs.

CHAPTER 3

DATA SETS AND PERFORMANCE EVALUATION METRICS

In this chapter, we give first the descriptions of the data sets used in the experiments. Then, formulation of the performance evaluation metrics used to measure the success of the classification methods are given.

3.1 Data Sets

In order to test the performance of the proposed methods, experiments are conducted on two hyperspectral datasets namely called Indian Pines and University of Pavia. These data sets can be downloaded online from the website in [66].

3.1.1 Indian Pines

This dataset was obtained by AVIRIS (Airborne/Visible Infrared Imaging Spectrometer) sensor during a flight over the Indian Pines test area in North-western Indiana [67]. It contains 145×145 pixels in the spatial dimension and 220 spectral bands in the spectral dimension. However, 20 spectral bands (104–108, 150–163, 220) are removed due to the water absorption [68]. So, final data size is $145 \times 145 \times 200$. The wavelength ranges of the bands are between 0.4–2.5 μm . The spatial resolution of this dataset is 20 m per pixel. It has 16 ground-truth classes, which includes mostly agricultural products, are given in Table 3.1. The term ground truth means that how the data set looks in reality. It is used to check the accuracy results of classifiers. A three band (50, 27, 17) false color image of the dataset, its corresponding ground-truth map and the names of the classes are given in Figure 3.1.

Table 3.1 The class names and the number of samples of the Indian Pines.

Class	Name	Samples
1	Alfalfa	54
2	Corn-notill	1434
3	Corn-mintill	834
4	Corn	234
5	Grass-pasture	497
6	Grass-trees	747
7	Grass-pasture-mowed	26
8	Hay-windrowed	489
9	Oats	20
10	Soybean-notill	968
11	Soybean-mintill	2468
12	Soybean-clean	614
13	Wheat	212
14	Woods	1294
15	Buildings-Grass-Trees-Drives	380
16	Stone-Steel-Towers	95
Total		10366

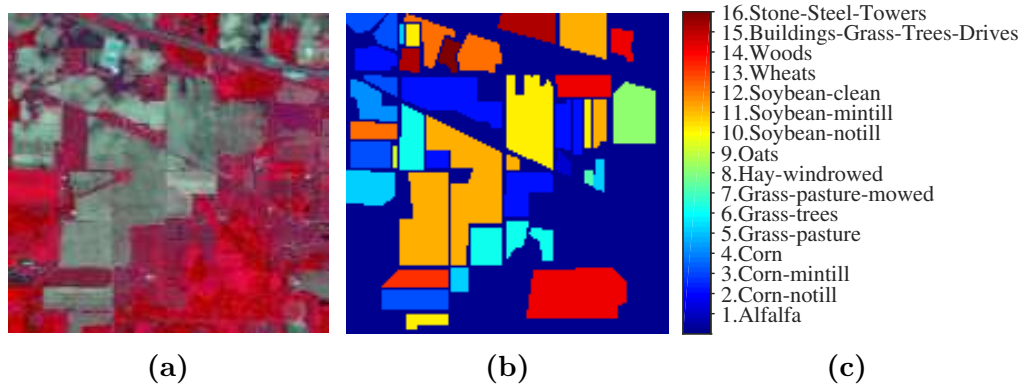


Figure 3.1 Indian Pines; (a) False color image, (b) Ground-truth map and (c) Class names.

3.1.2 University of Pavia

This dataset was obtained by ROSIS (Reflective Optics System Imaging Spectrometer) sensor during a flight over the University of Pavia test area in northern Italy [67]. It contains 610×610 pixels in the spatial dimension and 115 spectral

bands in the spectral dimension. However, 12 spectral bands are removed due to noise. In addition, some of the samples in the image have no information and have to be removed. After these preprocessing steps, final data size becomes $610 \times 340 \times 103$. The wavelength ranges of the bands are between $0.43 - 0.86 \mu\text{m}$. The spatial resolution of this dataset is 1.3 m per pixel. It has 9 ground-truth classes, which includes urban, soil and vegetation features, are given in Table 3.2. A three band (45, 27, 11) false color image of the dataset, its corresponding ground-truth map and the names of the classes are given in Figure 3.2.

Table 3.2 The class names and the number of samples of the University of Pavia.

Class	Name	Samples
1	Asphalt	6631
2	Meadows	18649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare Soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947
Total		42776

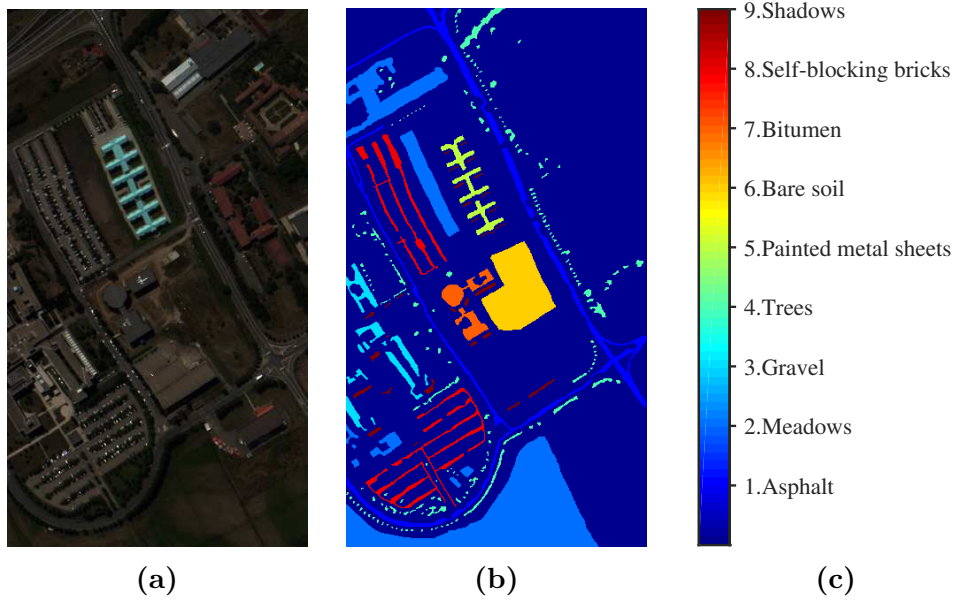


Figure 3.2 University of Pavia; (a) False color image, (b) Ground-truth map and (c) Class names.

3.2 Performance Evaluation Metrics

There are four metrics to evaluate performance of the HSI classifiers. They are overall accuracy (OA), class accuracy (CA), average accuracy (AA) and kappa statistic ($\mathcal{K}$) [69]. These metrics are calculated using a confusion matrix which is a table used to express the performance of a classifier on a set of test data.

In a confusion matrix, actual and predicted values are checked. In other words, actual and predicted class labels of test samples are compared. Actual labels are obtained from ground truth of the data set and predicted labels are determined by classifier. The size of confusion matrix is $C \times C$ where C is the total number of classes. Diagonal entries show number of correctly classified test samples and off diagonal entries show number of misclassifications in the related class.

An example of confusion matrix for a two classes data is shown in Table 3.3. In this matrix:

- a number of correctly classified test samples in Class 1
- b total number of test samples where their actual labels are Class 2 but classified as Class 1
- c total number of test samples where their actual labels are Class 1 but classified as Class 2
- d number of correctly classified test samples in Class 2
- $a + b$ total number of test samples classified as Class 1 by the classifier
- $c + d$ total number of test samples classified as Class 2 by the classifier
- $a + c$ total number of test samples in Class 1 obtained from ground truth of the data
- $b + d$ total number of test samples in Class 2 obtained from ground truth of the data
- N number of test samples should be classified

Accuracy results are evaluated using the definitions above and their explanations and formulations are given through Eqns. (3.1–3.4) for OA, CA, AA and $\mathcal{K}$, respectively.

Table 3.3 An example of two class confusion matrix table.

		Actual		Total
		Class 1	Class 2	
Predicted	Class 1	a	b	$a + b$
	Class 2	c	d	$c + d$
Total		$a + c$	$b + d$	N

$$\text{OA} = \frac{\text{Number of the test samples classified correctly}}{\text{Total number of the test samples}} \quad (3.1)$$

$$= \frac{a + d}{N}$$

$$\text{CA} = \frac{\text{Number of the test samples classified correctly in the related class}}{\text{Total number of the test samples in the related class}} \quad (3.2)$$

$$= \frac{a}{a + c} \quad \text{for the class accuracy of Class 1 (CA1)}$$

$$= \frac{d}{b + d} \quad \text{for the class accuracy of Class 2 (CA2)}$$

$$\text{AA} = \frac{\text{Sum of the classification accuracies}}{\text{Total number of the classes}} \quad (3.3)$$

$$= \frac{\text{CA1} + \text{CA2}}{2}$$

$$\mathcal{K} = \frac{\text{Observed accuracy} - \text{Expected accuracy}}{1 - \text{Expected accuracy}} \quad (3.4)$$

$$= \frac{\text{OA} - \left[\frac{(a + c)(a + b)}{N} + \frac{(b + d)(c + d)}{N} \right]}{1 - \left[\frac{(a + c)(a + b)}{N} + \frac{(b + d)(c + d)}{N} \right]}$$

CHAPTER 4

SPARSE REPRESENTATION-BASED HYPERSPECTRAL IMAGE CLASSIFICATION USING MULTISCALE SUPERPIXELS AND GUIDED FILTER

In the JSRC method, a fixed-size window is constructed around a centered test pixel and its neighbors are incorporated to use spatial information in this way. However, this method has two major drawbacks. The first one is that the fixed-size window can cause classification errors if the location of the test pixel is closer to edges. The second one is that optimal size of the window can vary for different local regions. For example, to get spatial information from a small region, size of the window should be small. Otherwise, to get spatial information from a large region, size of the window should be large.

To avoid the first problem, superpixel segmentations are used. In a superpixel, pixels that have similar spectral characteristics are included. But, the optimal number of superpixels in an image is hard to determine if we want superpixels to connect edge boundaries well.

To solve the second problem and the problem of optimal number of superpixels, multiscale superpixels are generated. However, multiscale approach may result in misclassifications near the edges. To overcome this problem, guided filter is added to the classification maps to preserve edge boundaries.

In multiscale superpixels approach, segmentation maps with different superpixel numbers are utilized to get local information effectively. However, there is not an optimal way of choosing correct number of superpixels that connect edges well in HSI. So, lots of the edge information will be discarded. In this study, multiscale superpixels and guided filter (MSS-GF) are incorporated together to fully consider the spatial and edge information. The different scale segmentation maps are obtained using multiscale superpixel approach and edge information that is not fully considered in multiscale segmentation are incorporated using guided filter.

4.1 Methodology and Formulation of the Proposed Method

In this section, backgrounds of superpixel, multiscale superpixels and guided filter are presented. In addition, problem formulation of the proposed MSS–GF method is theoretically described.

A superpixel is a group of pixels that have similar spectral characteristics. Its shape and size vary for distinct spatial structure in HSI. To obtain a superpixel segmentation map, a base image should be formed. Principal component analysis (PCA) [64] can be used to extract first or first three principal components as the base image. In this study, first three principal components are used as the base image. Then, ERS [70] is applied on the base image to generate superpixel segmentation maps. Each superpixel in segmentation maps is jointly classified by sparse representation classifier. Then, each classification map is represented as binary probability maps and these maps are filtered using guided filter. For the guided filter, first three principal components are used as the color guidance image because it provides better performance than gray-level guidance image as proposed in [71] and [72].

4.1.1 Forming Different-scaled Superpixel Segmentations

A 2-dimensional superpixel segmentation map generated on the base image contains $\mathbf{M}_{total}$ pixels in total. To determine total superpixel number $\mathbf{S}^j$ in the j^{th} scale

$$\mathbf{S}^j = \frac{\mathbf{M}_{total}}{\mathbf{RS}_n} \quad , \quad j = 1, 2, \dots, N \quad (4.1)$$

is calculated. In (4.1), $\mathbf{RS}_n = Q \times Q$ stands for region scale and N stands for total scale number. Normally, each superpixel has different shape and size. By performing equation (4.1), each superpixel nearly has $Q \times Q$ pixels inside due to the specified region scale.

4.1.2 Sparse Representation Based Multiscale Superpixels

A pixel in HSI is represented as $\mathbf{p} \in R^C$ where C is the number of spectral bands. The structured dictionary formed by all classes is denoted by $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K] \in R^{C \times D}$ where K is the total number of classes and D is the number of training samples. So, a test pixel $\mathbf{p}$ can be denoted as

$$\mathbf{p} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{n} \quad (4.2)$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_K]^t$ is the sparse coefficient vector and $\mathbf{n}$ is the possible approximation errors in the empirical data. Orthogonal Matching Pursuit (OMP) [73] can be used to solve the following optimization problem for sparse coefficient vector

$$\hat{\boldsymbol{\alpha}} = \arg \min \|\mathbf{p} - \mathbf{X}\boldsymbol{\alpha}\|_2 \quad s.t. \quad \|\boldsymbol{\alpha}\|_0 \leq L \quad (4.3)$$

the parameter L is the sparsity-level which points out the number of atoms to be selected in the dictionary at each iteration and $\|\cdot\|_0$ is the l_0 norm which counts the number of nonzero elements.

A superpixel contains a group of pixels that have similar spectral characteristics. Using (4.2) and (4.3) for a single pixel, a superpixel can be represented as

$$\mathbf{P}_s^j = \mathbf{X}\mathbf{A}_s^j + \mathbf{N}_s^j \quad (4.4)$$

$\mathbf{A}_s^j$ is the sparse coefficient matrix and s denotes the s^{th} superpixel in the j^{th} scaled segmentation map in (4.4). The sparse coefficient matrix $\mathbf{A}_s^j$ can be calculated by

$$\hat{\mathbf{A}}_s^j = \arg \min \|\mathbf{P}_s^j - \mathbf{X}\mathbf{A}_s^j\|_F \quad s.t. \quad \|\mathbf{A}_s^j\|_0 \leq L \quad (4.5)$$

We can use Simultaneous Orthogonal Matching Pursuit (SOMP) [42] to solve the optimization problem in (4.5). After $\hat{\mathbf{A}}_s^j$ is found, reconstruction residual error which is the difference between original pixel and reconstructed pixel is calculated using Frobenius norm as in (4.6)

$$E_k(\mathbf{P}_s^j) = \|\mathbf{P}_s^j - \mathbf{X}_k \hat{\mathbf{A}}_{s_k}^j\|_F, \quad k = 1, 2, \dots, K \quad (4.6)$$

where $\hat{\mathbf{A}}_{s_k}^j$ is the suitable atoms of the sub-dictionary $\mathbf{X}_k$ in $\hat{\mathbf{A}}_s^j$. Finally, the class label of $\mathbf{P}_s^j$ is obtained by the minimal reconstruction residual error as in (4.7)

$$Class(\mathbf{P}_s^j) = \arg \min (E_k(\mathbf{P}_s^j)) \quad (4.7)$$

4.1.3 Guided Filter

Guided filter aims to obtain an output image O using a guidance image G in a local window ω_v constructed around pixel u , with size $(2r+1) \times (2r+1)$ where r is the radius. By applying filtering process, edge information of G is transferred to O . The linear relationship between O and G can be written as

$$O_u = a_v G_u + b_v, \quad \forall u \in \omega_v \quad (4.8)$$

where a_v and b_v are linear coefficient and bias, respectively. To find the values of a_v and b_v , energy function $E(a_v, b_v)$ is carried out on the local window ω_v as in (4.9)

$$E(a_v, b_v) = \sum_{u \in \omega_v} ((a_v G_u + b_v - I_u)^2 + \epsilon a_v^2) \quad (4.9)$$

I is the input image and ϵ is the regularization parameter which adjusts the blurring degree. Guidance image G can be gray-level or RGB image as mentioned in [71] and [72].

4.1.4 Problem Formulation

After all the different scaled superpixel segmentations are calculated and class labels of the pixels are determined in each scale, classification maps are transformed to binary probability maps. For the j^{th} scaled classification map $\mathbf{M}^j$, the corresponding binary probability map is shown as

$$\mathbf{PM}^j = [\mathbf{PM}_1^j, \mathbf{PM}_2^j, \dots, \mathbf{PM}_K^j] \quad (4.10)$$

$\mathbf{PM}^j$ refers to probability map in (4.10). If the class label of a pixel u is k in the j^{th} scaled classification map, then

$$\mathbf{PM}_{u,k}^j = \begin{cases} 1 & \text{if } \mathbf{M}_u^j = k \\ 0 & \text{otherwise.} \end{cases} \quad (4.11)$$

After (4.11) is applied for all pixels in each scale, all probability maps must be filtered by the guided filter using a weight function. This operation for the pixel u in the j^{th} scale k^{th} class can be denoted in (4.12) as

$$\overline{\mathbf{PM}}_{u,k}^j = \sum_h W_{u,h}(G) \mathbf{PM}_{h,k}^j \quad (4.12)$$

$\overline{\mathbf{PM}}_{u,k}^j$ refers to filtered output of $\mathbf{PM}_{u,k}^j$. Index h means the neighboring pixels of pixel u inside a window. The weight function W uses and protects the edge information of the guidance image G . For the guided filter, weight function $W_{u,h}(G)$ is written as

$$W_{u,h}(G) = \frac{1}{|\omega|^2} \sum_{v:(u,h) \in \omega_v} \left(1 + \frac{(G_u - \mu_v)(G_h - \mu_v)}{\sigma_v^2 + \epsilon} \right) \quad (4.13)$$

where μ_v and σ_v are the mean and variance of G in the window ω_v , and $|\omega|$ is the total number of pixels in ω_v .

4.1.5 Determination of Class Label

After all the filtered probabilities are evaluated for each scale, the class label of a pixel u is determined by majority voting (MV) rule. First, the maximum value of filtered probabilities and its class index for the pixel u at each scale are found. Then, MV rule is applied on the class indexes and the label of the pixel u is selected as most repeated class. Maximum value and its class index can be found as

$$Class(u^j) = \mathbf{I} \left(\arg \max \left(\overline{\mathbf{PM}}_{u,k}^j \right) \right) \quad (4.14)$$

In (4.14), $\mathbf{I}$ indicates the class index of maximum value of filtered probabilities and u^j refers to a pixel u in the j^{th} scale. The label of u can be found by applying (4.15)

$$Class(u) = \mathbf{MV} (Class(u^j)) \quad (4.15)$$

4.2 Simulation Setup and Experimental Results

In this section, the proposed MSS–GF method is tested on two hyperspectral data sets called Indian Pines and University of Pavia. The methods in the literature called SVM [24], JSRC [42] and EPF [71] are used for comparison. In addition, classification is performed by directly applying joint sparse representation based classification using only multiscale superpixels by applying majority voting rule but not guided filter process. We call this method as MSS–MV. All experiments are repeated 10 times with different training samples and then the results are averaged.

The proposed MSS–GF method has parameters to be adjusted such as region scales $\mathbf{RS}_n$ in each segmentation map and sparsity-level L for each data set. In addition to these parameters, radius r and regularization parameter ϵ must be adjusted for the guided filter. Before performing guided filter, classification results of the different scaled superpixel segmentations must provide satisfactory accuracies. If the values of $\mathbf{RS}_n$ and L are chosen properly, then the classification accuracies will be high. So, firstly, different region scales and different sparsity levels are applied on two data sets to determine proper values. To determine $\mathbf{RS}_n$, different region scales which starts from 3×3 and goes to 19×19 are selected. For the L , the values 1, 3 and 5 are selected. The value of r should be set properly to capture edge information of the original image effectively. So, different r values between 1 and 9 are chosen to obtain best performance for the two data sets.

Table 4.1 Number of superpixels in each region scale for the Indian Pines.

Scale	3×3	5×5	7×7	9×9	11×11
Number of superpixel	2336	841	429	260	174

The value of ϵ is set as 0.01 for the both data sets as recommended in the work [71]. The experiments applied on the data sets are described in the following subsections.

4.2.1 Simulation Results on Indian Pines

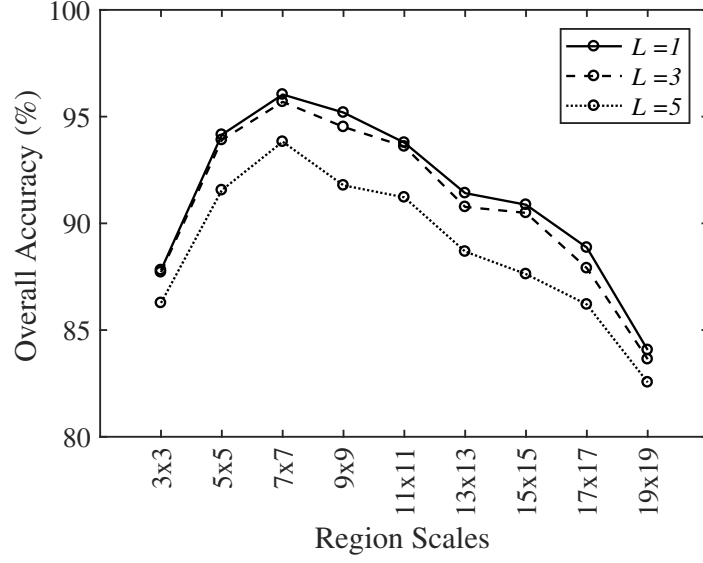
In all studies conducted on the Indian Pines data set, 10% of total samples of each class are used for training and the remaining 90% are used for testing.

Gaussian RBF kernel is used in the SVM method and 5-fold cross validation is performed to select the best parameters. For the JSRC method, L is set to 1 and window size is set as 5×5 to achieve best classification results. For the EPF method, original parameters ($r=4$ and $\epsilon=0.01$) are used in the experiment for the two data sets. First three principal components are utilized as guidance image for the guided filter process in both EPF and proposed MSS-GF methods. Figure 4.1 shows the result of the experiment to determine the parameters for the $\mathbf{RS}_n$ and L . For this data set, region scales $\mathbf{RS}_n$ are selected as 3×3 , 5×5 , 7×7 , 9×9 and 11×11 , because OA and AA values are higher. Also, $L=1$ mostly gives higher results than $L=3$ and $L=5$ in terms of both OA and AA. So, $L=1$ is chosen as the sparsity level.

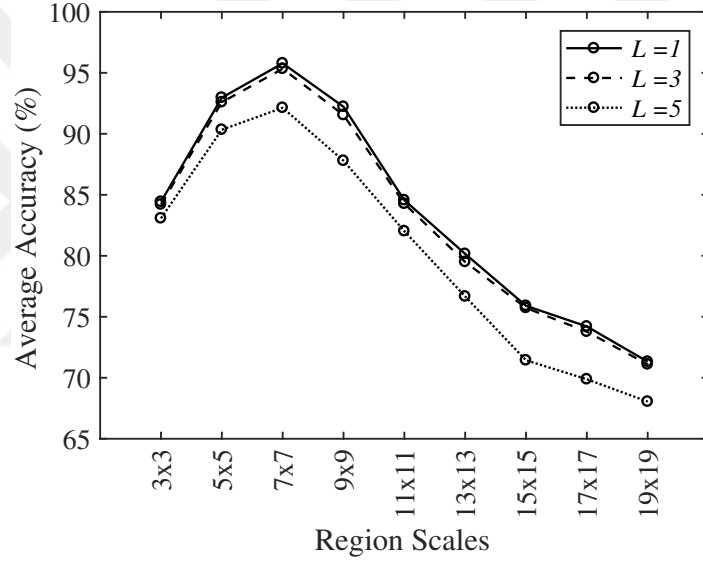
The number of superpixels in each segmentation map are presented in Table 4.1 and the superpixel segmentation maps for the selected region scales can be seen in Figure 4.2.

After the values of L and $\mathbf{RS}_n$ are determined, radius r must be set for the guided filter. Figure 4.3 shows the effect of different r values on the classification accuracy results in terms of OA. As it is seen, radius $r=4$ provides higher OA than other radius values. So, $r=4$ is set as the radius of the guided filter.

After the parameter adjustments are completed, the proposed MSS-GF is compared with SVM, JSRC, EPF and MSS-MV as mentioned before. The comparison results and the number samples used in the experiments are given in Table



(a)



(b)

Figure 4.1 Effect of region scales and sparsity-level on the Indian Pines; (a) OA and (b) AA.

4.2. As it is seen, SVM has the lowest accuracy results in terms of OA, AA and $\mathcal{K}$. This is because the spatial information is not used in SVM. Since EPF is the improved version of SVM by adding additional guided filter process to the classification map of SVM, it has higher accuracy results (approximately 15%) than SVM in terms of OA, AA and $\mathcal{K}$. MSS-MV has improved accuracy results than the results of JSRC, because MSS-MV uses multiscale superpixels instead of a fixed-size region scale and effectively capture the spatial information. The proposed MSS-GF improves the result of MSS-MV by applying guided filter and so it has the highest classification accuracy results. Thanks to the additional

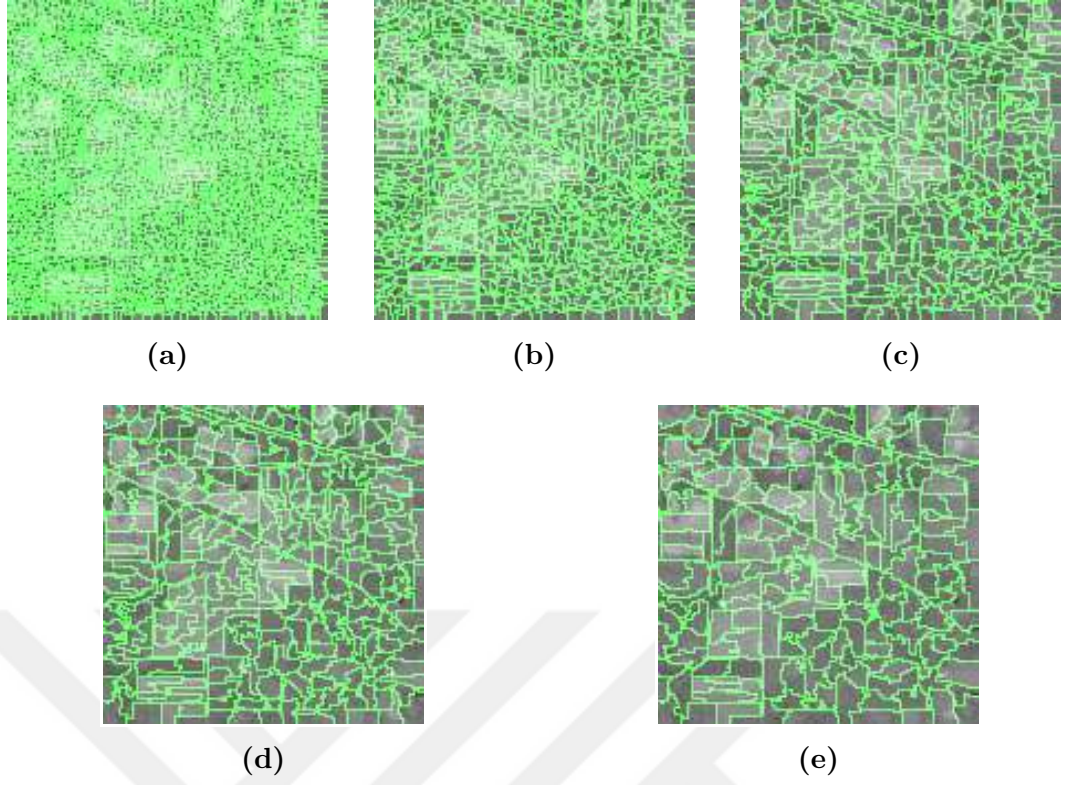


Figure 4.2 Superpixel segmentation maps of the Indian Pines for the region scales (a) 3×3 , (b) 5×5 , (c) 7×7 , (d) 9×9 and (e) 11×11 .

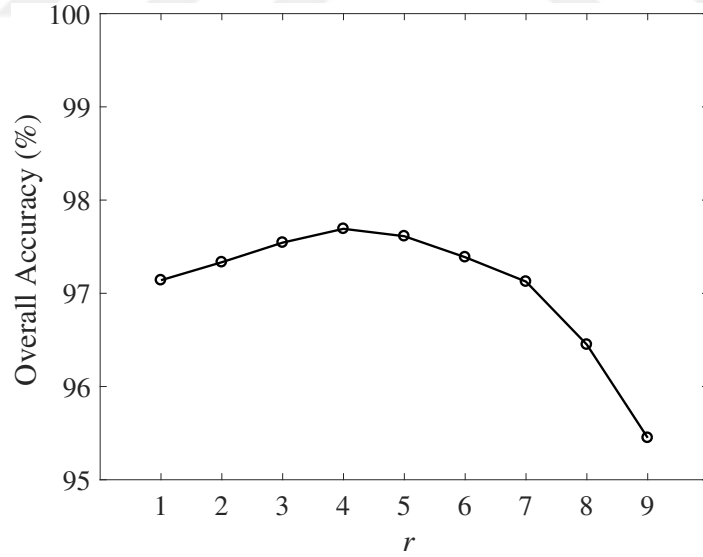


Figure 4.3 Effect of guided filter radius r for the Indian Pines.

guided filter, edge information of the original image is transferred to the classification map of the MSS-MV and eventually the accuracy results are improved. Since the aim is to improve the results of JSRC, the proposed MSS-GF fulfills this task well. It improves the OA result of JSRC nearly by 2.5%, AA result

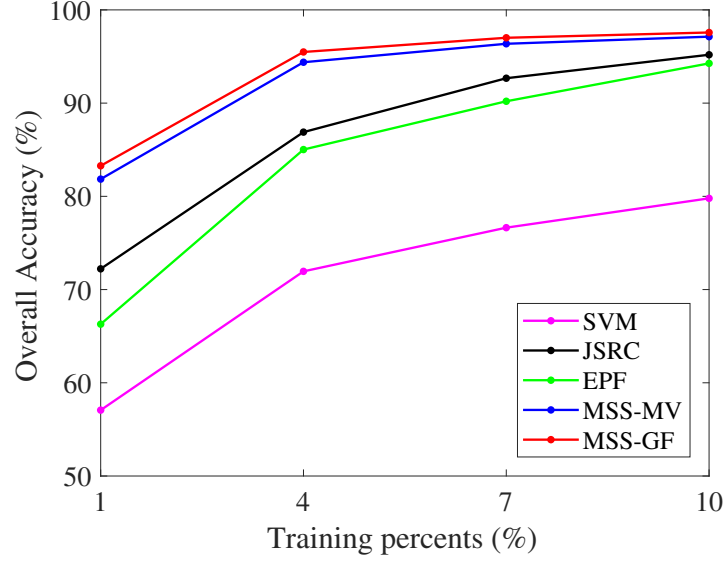
Table 4.2 Classification accuracies (%) of Indian Pines dataset obtained by SVM, EPF, JSRC, MSS–MV and the proposed MSS–GF method.

Class	Train	Test	SVM	JSRC	EPF	MSS–MV	MSS–GF
1	6	48	60.30	93.32	10.00	96.46	96.44
2	144	1290	71.98	93.22	93.86	95.37	95.24
3	84	750	74.47	92.83	95.94	96.99	98.38
4	24	210	60.33	93.03	95.73	93.90	92.86
5	50	447	88.71	96.55	97.48	96.78	99.71
6	75	672	86.63	96.59	96.10	99.30	99.04
7	3	23	80.92	84.47	100	100	100
8	49	440	90.87	99.54	91.34	99.77	100
9	2	18	77.35	69.92	0.00	78.33	71.74
10	97	871	76.00	92.96	94.21	93.48	97.31
11	247	2221	78.62	96.34	93.23	98.16	97.46
12	62	552	70.31	94.78	91.75	94.06	95.18
13	22	190	94.07	93.84	100	99.47	99.13
14	130	1164	91.49	97.25	94.76	99.66	99.52
15	38	342	69.00	96.19	96.52	97.11	98.94
16	10	85	96.48	91.29	93.77	98.00	97.89
OA	–	–	79.79	95.19	94.26	97.13	97.58
AA	–	–	79.22	92.63	84.04	96.05	96.18
$\mathcal{K}$	–	–	76.88	94.51	93.44	96.73	97.24

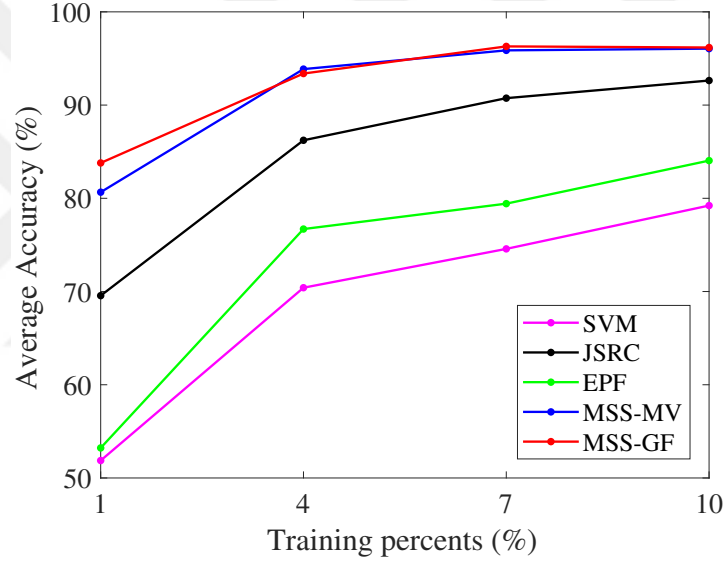
nearly by 3.5% and $\mathcal{K}$ result nearly by 3%.

In addition, a comparison is performed between MSS–GF and other methods under different training percents such as 1%, 4%, 7% and 10%. As it is seen from Figure 4.4, the proposed MSS–GF method outperforms the other methods for all training percents in terms of both OA and AA.

To observe the effect of the methods visually, classification maps of the methods are presented in Figure 4.5. Since SVM is a spectral based classification method, its classification map seems very noisy. The classification maps of the other methods seem better because the spatial information is incorporated. Effect of the guided filter can be seen from the Figures 4.5(f) and 4.5(g). Most of the misclassifications in Figure 4.5(f) near the edges in the classes ”*soybean mintill, buildings–grass–trees and grass/pasture*” are corrected by MSS–GF as it is seen from Figure 4.5(g).



(a)



(b)

Figure 4.4 Effect of different training percents on (a) OA and (b) AA using SVM, JSRC, EPF, MSS-MV and proposed MSS-GF method for the Indian Pines.

Finally, Figure 4.6 shows the superiority of using multiscale superpixels when compared to single scale superpixels in terms of OA. As it is seen, OA result of multiscale superpixels is nearly 1% higher compared to the single-scale (7×7) providing maximum OA. Eventually, the advantage of multiscale superpixels has been proven by this experiment.

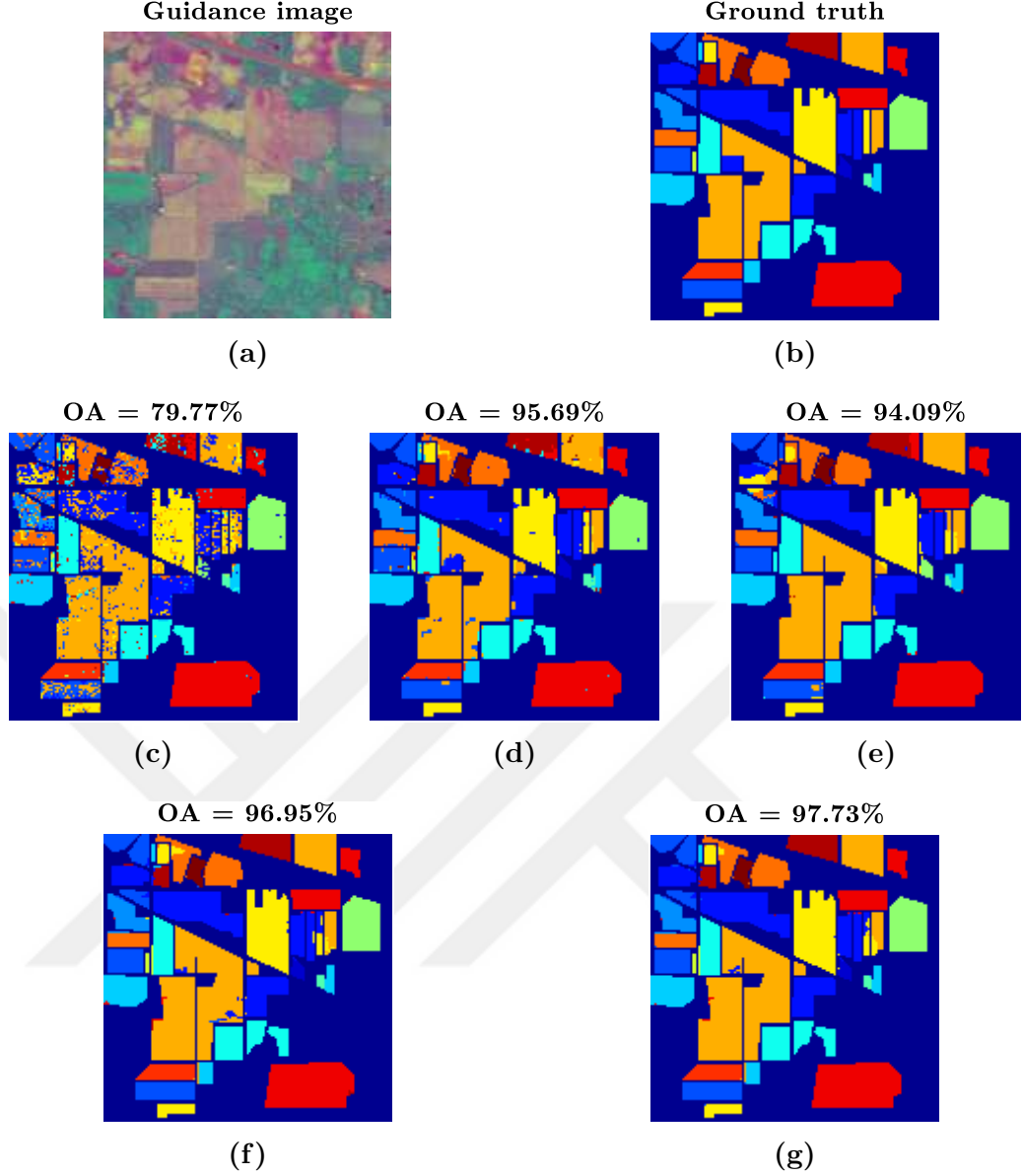


Figure 4.5 Indian Pines; (a) Guidance image, (b) Ground truth. The classification maps obtained by (c) SVM, (d) JSRC, (e) EPF, (f) MSS-MV and (g) MSS-GF.

4.2.2 Simulation Results on University of Pavia

For the University of Pavia data set, 300 train samples per class are used as training and remaining samples are utilized as test samples.

Gaussian RBF kernel is used in the SVM method and 5-fold cross validation is performed to select the best parameters. For the JSRC method, L is set to 11 and window size is set as 13×13 for to achieve best results. For the EPF method, original parameters ($r = 4$ and $\epsilon = 0.01$) are used in the experiment for the two

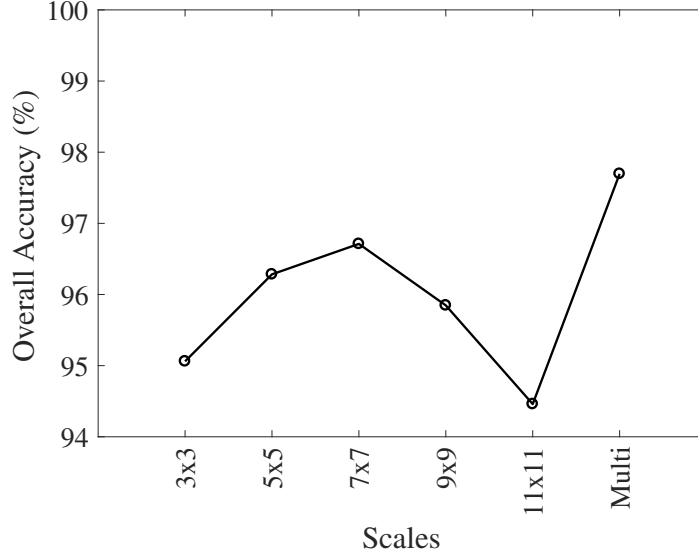


Figure 4.6 Single-scale and multiscale superpixels comparison for the Indian Pines.

data sets. First three principal components are utilized as guidance image for the guided filter process in both EPF and proposed MSS-GF methods. Figure 4.7 shows the result of the experiment to determine the parameters for the $\mathbf{RS}_n$ and L . For this data set, region scales $\mathbf{RS}_n$ are selected as 7×7 , 9×9 , 11×11 , 13×13 , 15×15 and 17×17 , because OA and AA values are higher. Also, $L=1$ mostly gives higher results than $L=3$ and $L=5$ in terms of both OA and AA. So, $L=1$ is chosen as the sparsity level.

The number of superpixels in each segmentation map are presented in Table 4.3 and the superpixel segmentation maps for the selected region scales can be seen in Figure 4.8.

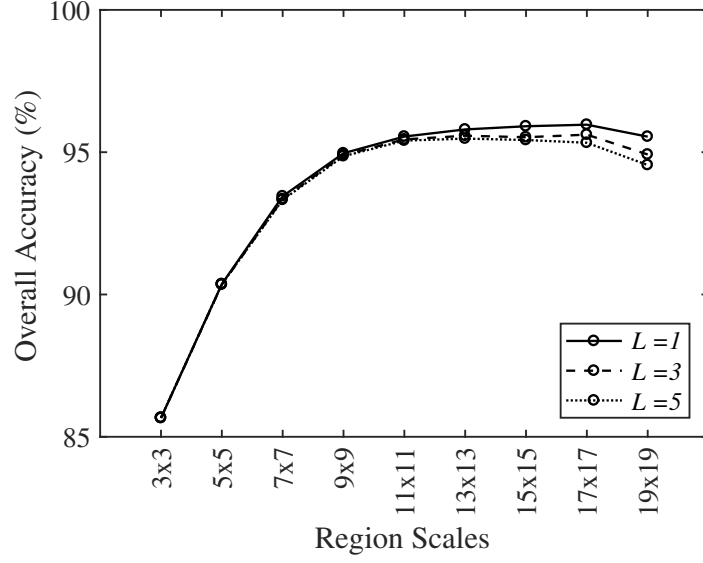
Table 4.3 Number of superpixels in each region scale for the University of Pavia.

Scale	7×7	9×9	11×11	13×13	15×15	17×17

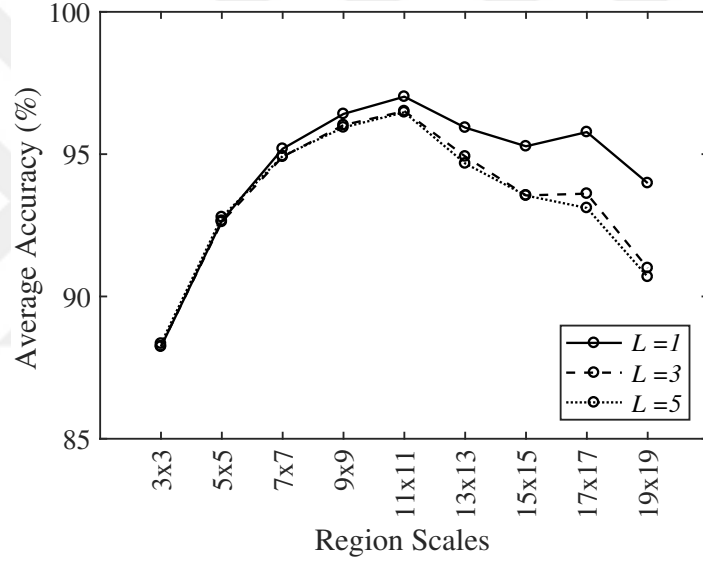
Number of superpixel	4233	2560	1714	1227	922	718

After the values of L and $\mathbf{RS}_n$ are determined, radius r must be set for the guided filter. Figure 4.9 shows the effect of different r values on the classification accuracy results in terms of OA. As it is seen, radius $r=9$ provides higher OA than other radius values. So, $r=9$ is set as the radius of the guided filter.

After the parameter adjustments are completed, the proposed MSS-GF is com-



(a)



(b)

Figure 4.7 Effect of region scales and sparsity-level on the University of Pavia; (a) OA and (b) AA.

pared with SVM, JSRC, EPF and MSS-MV as mentioned before. The comparison results and the number samples used in the experiments are given in Table 4.4. As it is seen, the accuracy results of SVM stay behind the results of other methods because SVM only uses spectral information. Since EPF is the improved version of SVM by adding additional guided filter process to the classification map of SVM, it has higher accuracy results than SVM in terms of OA, AA and $\mathcal{K}$. MSS-MV has improved accuracy results of the JSRC, because MSS-MV uses multiscale superpixels instead of a fixed-size region scale and effectively capture the spatial information. The proposed MSS-GF improves the result of MSS-MV

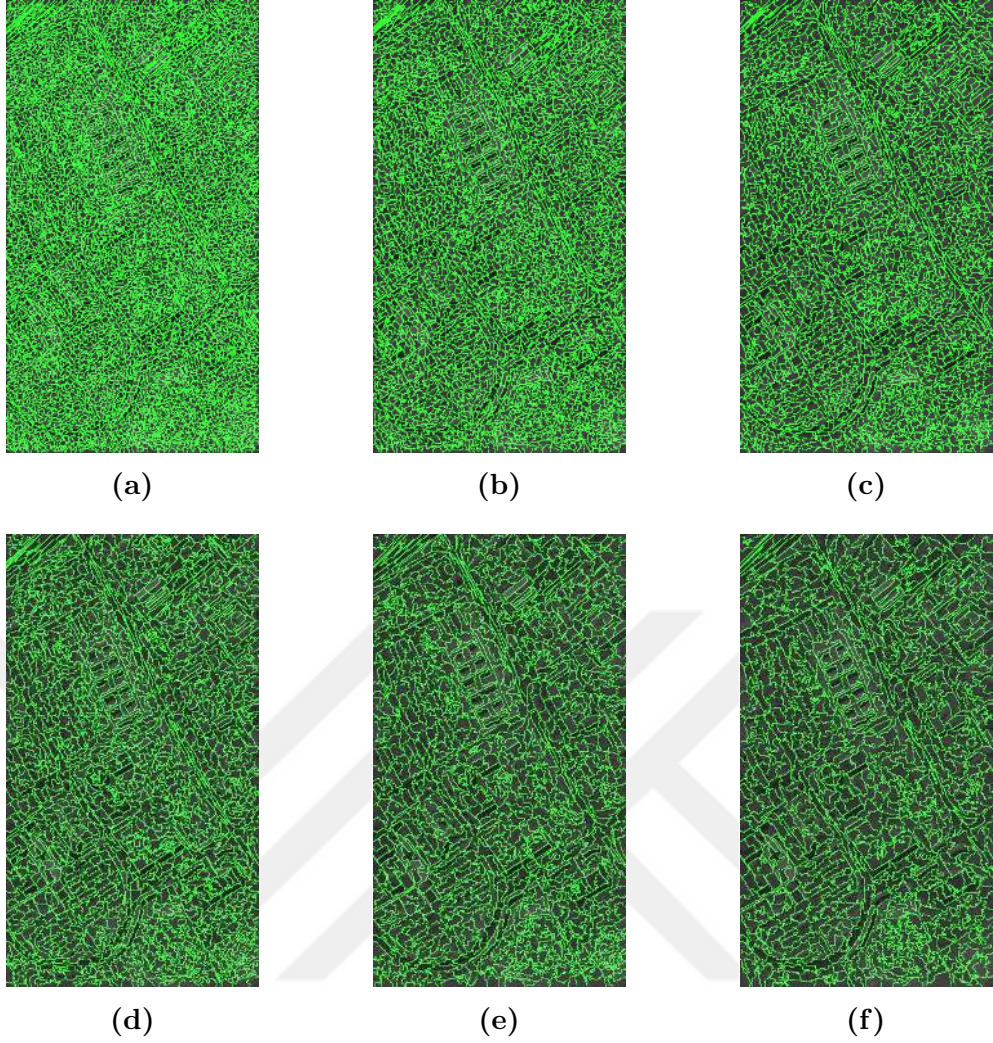


Figure 4.8 Superpixel segmentation maps of the University of Pavia for the region scales (a) 7×7 , (b) 9×9 , (c) 11×11 , (d) 13×13 , (e) 15×15 and (f) 17×17 .

by applying guided filter and so it has the highest classification accuracy results. Thanks to the additional guided filter, edge information of the original image is transferred to the classification map of the MSS–MV and eventually the accuracy results are improved. Since the aim is to improve the results of JSRC, the proposed MSS–GF fulfills this task well. It improves the OA result of JSRC nearly by 3%, AA result nearly by 7% and $\mathcal{K}$ result nearly by 4%.

In addition, a comparison is performed between MSS–GF and other methods under different training percents such as 1%, 4%, 7% and 10%. For this dataset, SVM performs better results compared to Indian Pines due to the less number of classes. It may be because SVM is basically a binary-classifier and not convenient for multiclass datasets as in the Indian Pines. So, SVM-based EPF method outperforms the other methods up to nearly 5% training percent for the Pavia

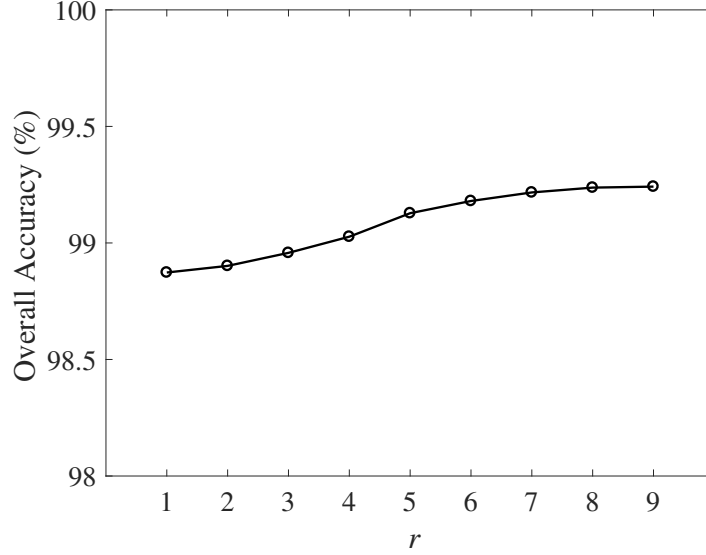


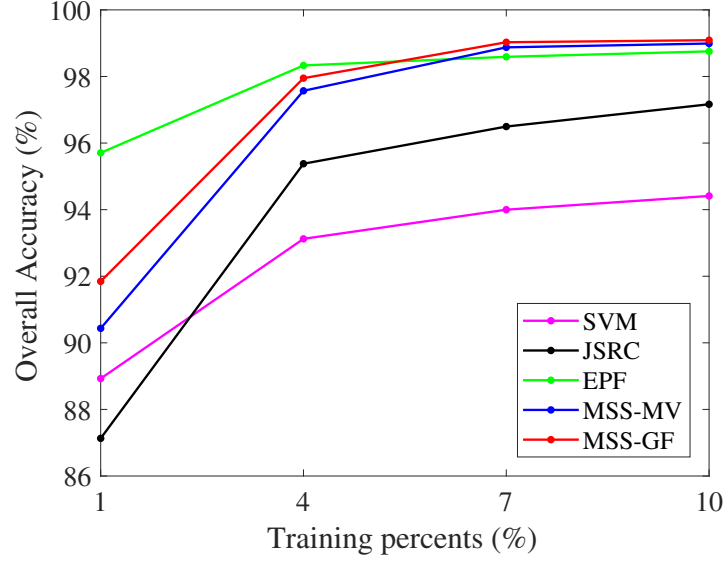
Figure 4.9 Effect of guided filter radius r for the University of Pavia.

Table 4.4 Classification accuracies (%) of University of Pavia dataset obtained by SVM, EPF, JSRC, MSS-MV and the proposed MSS-GF method.

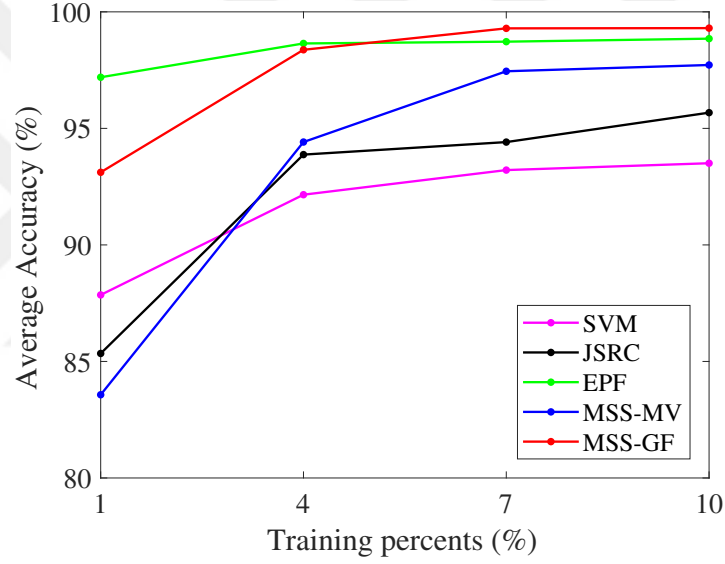
Class	Train	Test	SVM	JSRC	EPF	MSS-MV	MSS-GF
1	300	6331	97.01	95.68	99.01	97.58	99.85
2	300	18349	98.01	99.56	99.77	99.18	99.63
3	300	1799	78.44	91.68	99.36	99.96	99.47
4	300	2764	87.73	91.67	98.88	95.30	96.60
5	300	1045	97.75	88.71	99.39	99.45	99.64
6	300	4729	80.59	96.90	96.27	99.98	99.79
7	300	1030	67.70	89.95	99.49	100	99.10
8	300	3382	87.29	92.79	95.98	99.28	96.54
9	300	647	99.85	82.04	99.78	96.51	99.84
OA	–	–	91.75	96.24	98.79	98.78	99.17
AA	–	–	88.26	92.11	98.66	98.58	98.94
$\mathcal{K}$	–	–	88.96	94.91	98.36	98.35	98.88

dataset. And after this percent, the proposed MSS-GF method outperforms the other methods in terms of both OA and AA as can be seen in Fig. 4.10.

To observe the effect of the methods visually, classification maps of the methods are presented in Figure 4.11. Since SVM is a spectral based classification method, its classification map seems very noisy. The classification maps of the other methods seem better because the spatial information is incorporated. Effect of the guided filter can be seen from the Figures 4.11(f) and 4.11(g). Most of the



(a)



(b)

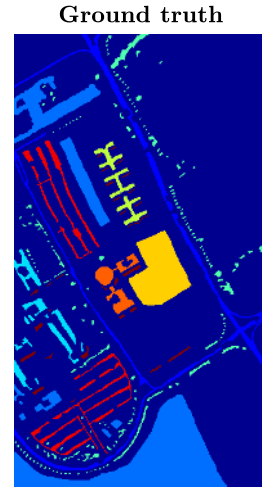
Figure 4.10 Effect of different training percents on (a) OA and (b) AA using SVM, JSRC, EPF, MSS-MV and proposed MSS-GF method for the University of Pavia.

misclassifications in Figure 4.11(f) near the edges in the classes ”*asphalt, meadows and brick*” are classified correctly in the proposed MSS-GF as it is seen from Figure 4.11(g).

Finally, Figure 4.12 shows the superiority of using multiscale superpixels when compared to single scale superpixels in terms of OA. As it is seen, OA result of multiscale superpixels is nearly 0.5% higher compared to the single-scale (11×11) providing maximum OA. Eventually, the advantage of multiscale superpixels has

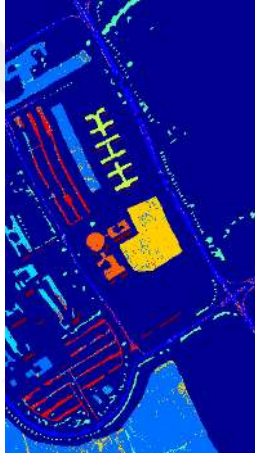


(a)



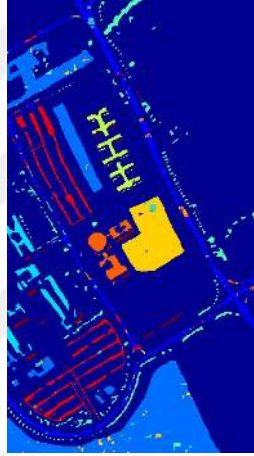
(b)

OA = 92.01%



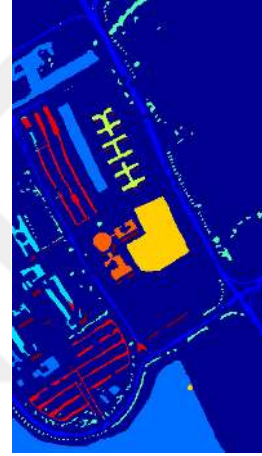
(c)

OA = 95.22%



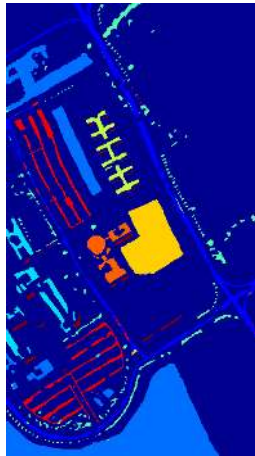
(d)

OA = 99.02%



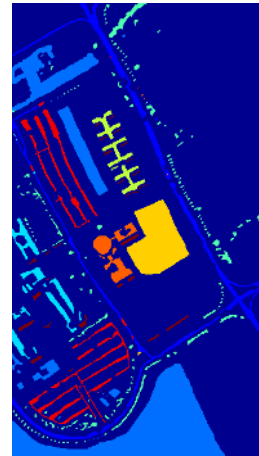
(e)

OA = 98.71%



(f)

OA = 99.21%



(g)

Figure 4.11 University of Pavia; (a) Guidance image, (b) Ground truth. The classification maps obtained by (c) SVM, (d) JSRC, (e) EPF, (f) MSS-MV and (g) MSS-GF.

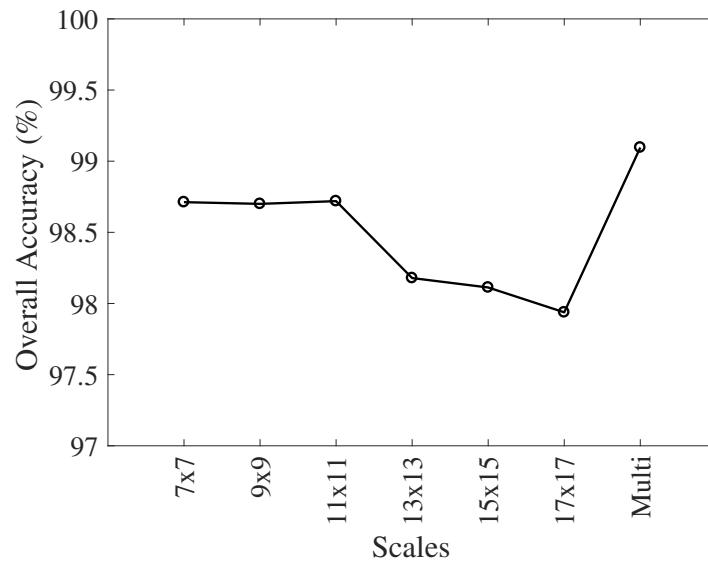


Figure 4.12 Single-scale and multiscale superpixels comparison for the University of Pavia.

been proven by this experiment.

CHAPTER 5

SPARSE REPRESENTATION-BASED CLASSIFICATION OF HYPERSPPECTRAL IMAGES USING SPECTRAL MATCHING METHODS

In the JSRC method, it is assumed that the weights of all neighboring pixels in a fixed size window around the test pixel are equal. Particularly, as the window size increases, the error rate will increase if it is considered that the pixels belonging to different classes will be included in the classification process. In order to solve this problem, 3SM-JSRC (3 Spectral Matching Joint Sparse Representation Classifier) method which applies 3 spectral matching methods to central test pixel and each neighbor pixel in the window and combines with JSRC is proposed. Three spectral matching methods called euclidean distance (ED), spectral angle mapper (SAM) and pearson correlation coefficient (PCC) are used in common to decide which neighboring pixels to be selected or not to be selected in the classification process. These matching methods are used in the problems such as band selection and classification of HSI [74–76]. The matching methods are applied to all adjacent pixels respectively so that the corresponding neighboring pixel will be selected or not be selected according to the specified threshold value.

5.1 Spectral Matching Methods

In this section, ED, SAM and PCC methods are explained and their related formulas are given.

5.1.1 Euclidean Distance

The ED is used to determine the distance between the X and Y vectors in a B dimensional space. If the distance between two vectors is small, then their

spectral characteristics can be thought as similar. The formula of ED is given in Eqn. (5.1) as follow:

$$ED(X, Y) = \sqrt{\sum_{i=1}^B (X_i - Y_i)^2} \quad (5.1)$$

5.1.2 Spectral Angle Mapper

Thanks to the SAM method, the angle between the X and Y vectors in a B dimensional space can be calculated. The small angle value shows that their spectral characteristics are similar. The value of SAM can be calculated with the help of the following Eqn. (5.2) as:

$$SAM(X, Y) = \cos^{-1} \left(\frac{\sum_{i=1}^B X_i Y_i}{\sqrt{\sum_{i=1}^B X_i^2} \sqrt{\sum_{i=1}^B Y_i^2}} \right) \quad (5.2)$$

5.1.3 Pearson Correlation Coefficient

The PCC also shows the similarity of two vectors, like other methods, and is a value between -1 and 1. A value close to -1 indicates that the two vectors are similar but they lie in the opposite direction. Conversely, a value close to 1 indicates that the two vectors are similar and lie in the same direction. Therefore, if the X and Y vectors are similar in B dimensional space, the absolute value of PCC should be close to 1. On the contrary, it should be close to 0. The PCC can be calculated as:

$$PCC(X, Y) = \frac{\sum_{i=1}^B (X_i - \bar{X}) (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^B (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^B (Y_i - \bar{Y})^2}} \quad (5.3)$$

where $\bar{X} = \left(\frac{1}{B} \sum_{i=1}^B X_i \right)$ and $\bar{Y} = \left(\frac{1}{B} \sum_{i=1}^B Y_i \right)$ in Eqn. (5.3) stand for the mean of the vectors X and Y , respectively.

5.2 Formulation of the Proposed 3SM-JRC Method

The pixels in the window are indicated by $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{Q \times Q}]$, provided that $\mathbf{p}_1$ is the test pixel and the others are the neighbors. ED, SAM and PCC matching methods are applied on $\mathbf{p}_1$ and its neighbors, respectively. This process can be modeled as the following Eqns. (5.4), (5.5) and (5.6):

$$ED_j = \text{ED}(\mathbf{p}_1, \mathbf{p}_j) \quad \text{where } j = 2, 3, \dots, Q \times Q \quad (5.4)$$

$$SAM_j = \text{SAM}(\mathbf{p}_1, \mathbf{p}_j) \quad \text{where } j = 2, 3, \dots, Q \times Q \quad (5.5)$$

$$PCC_j = \text{PCC}(\mathbf{p}_1, \mathbf{p}_j) \quad \text{where } j = 2, 3, \dots, Q \times Q \quad (5.6)$$

After these values are found for the all neighbor pixels, the mean values of the vectors ED_j , SAM_j and PCC_j should be calculated using the Eqns. (5.7), (5.8) and (5.9), respectively, as:

$$\overline{ED} = \frac{1}{(Q \times Q) - 1} \sum_{j=2}^{Q \times Q} ED_j \quad (5.7)$$

$$\overline{SAM} = \frac{1}{(Q \times Q) - 1} \sum_{j=2}^{Q \times Q} SAM_j \quad (5.8)$$

$$\overline{PCC} = \frac{1}{(Q \times Q) - 1} \sum_{j=2}^{Q \times Q} PCC_j \quad (5.9)$$

The calculated mean values will be used as a threshold for each matching method. The next step is to check if the calculated ED_j , SAM_j and PCC_j values are greater or smaller than the threshold values. Then, according to the results, $\mathbf{ED}_{1,0}$, $\mathbf{SAM}_{1,0}$ and $\mathbf{PCC}_{1,0}$ vectors which consist of only 0 and 1 are constructed. The value 0 means that related neighbor pixel has different spectral characteristics than the test pixel and so it should be discarded. The value 1 means that related neighbor pixel has similar spectral characteristics with the test pixel and so it should be selected. All these process can be formulated as the following Eqns. (5.10), (5.11) and (5.12):

$$\mathbf{ED}_{1,0} = \begin{cases} 1 & ; \quad ED_j \leq \overline{ED} \\ 0 & ; \quad ED_j > \overline{ED} \end{cases} \quad (5.10)$$

$$\mathbf{SAM}_{1,0} = \begin{cases} 1 & ; \quad SAM_j \leq \overline{SAM} \\ 0 & ; \quad SAM_j > \overline{SAM} \end{cases} \quad (5.11)$$

$$\mathbf{PCC}_{1,0} = \begin{cases} 1 & ; \quad PCC_j \geq \overline{PCC} \\ 0 & ; \quad PCC_j < \overline{PCC} \end{cases} \quad (5.12)$$

After $\mathbf{ED}_{1,0}$, $\mathbf{SAM}_{1,0}$ and $\mathbf{PCC}_{1,0}$ vectors are constructed, majority voting (MV) is applied to these three vectors and the $\mathbf{V}$ vector is formed as follow:

$$\begin{aligned}\mathbf{V} &= \text{MV} [\mathbf{ED}_{1,0} ; \mathbf{SAM}_{1,0} ; \mathbf{PCC}_{1,0}] \\ &= [V_2 \ V_3 \ \cdots \ V_{Q \times Q}] \quad \text{where} \quad V_{j=2,3,\dots,Q \times Q} \in [0, 1]\end{aligned}\quad (5.13)$$

In Eqn. (5.13), the selection is done for the neighbor pixels around the center test pixel $\mathbf{p}_1$. Since the aim of the proposed algorithm is to classify the pixel $\mathbf{p}_1$, a weight vector $\mathbf{W}$ which includes the weights for $\mathbf{p}_1$ and $\mathbf{V}$ is formed as follow:

$$\mathbf{W} = [1 \ V_2 \ V_3 \ \cdots \ V_{Q \times Q}] \quad (5.14)$$

where the weight of $\mathbf{p}_1$ is determined as 1. Finally, $\tilde{\mathbf{P}}$ matrix containing the selected test pixel $\mathbf{p}_1$ and its neighbor pixels is formed by the following Eqn. (5.15) as:

$$\tilde{\mathbf{P}} = \mathbf{W}\mathbf{P} \quad (5.15)$$

A visual representation of the steps described so far can be seen in Figure 5.1.

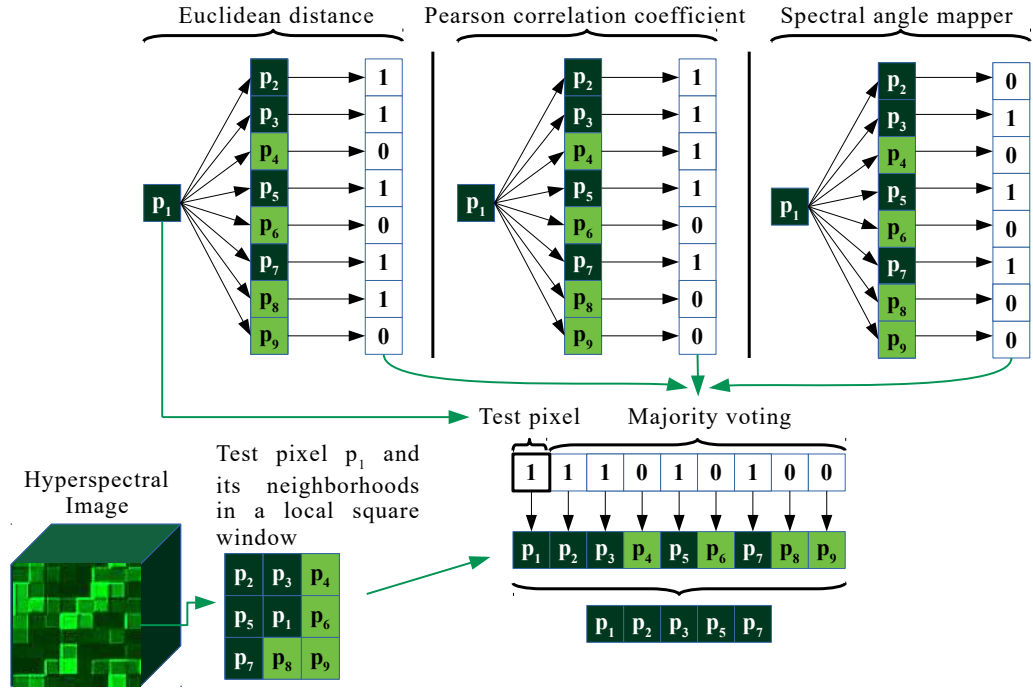


Figure 5.1 Selection of the neighbor pixels within the window created around the test pixel by the methods ED, SAM and PCC.

After the neighbor selection process is completed, the steps of JSRC method should be applied. The pixel matrix $\tilde{\mathbf{P}}$ is linearly and sparsely represented using

a few training samples in a structured dictionary $\mathbf{X}$. Also, some representation error $\mathbf{N}$ is added to problem in Eqn. (5.16) as:

$$\tilde{\mathbf{P}} = \mathbf{X}\mathbf{A} + \mathbf{N} \quad s.t. \quad \|\mathbf{A}\|_0 \leq L \quad (5.16)$$

$\mathbf{A}$ denotes the sparse coefficient matrix and L represents the sparsity level. In sparse representation based approaches, the goal is to find the reconstructed test pixel using sparse coefficient matrix. So, the problem in Eqn. (5.17) must be solved for $\mathbf{A}$ using SOMP [42] algorithm as:

$$\hat{\mathbf{A}} = \arg \min \left\| \tilde{\mathbf{P}} - \mathbf{X}\mathbf{A} \right\|_F \quad s.t. \quad \|\mathbf{A}\|_0 \leq L \quad (5.17)$$

After $\hat{\mathbf{A}}$ is calculated, the representation residual error E_m should be computed for each class using the Eqn. (5.18) and then class label of the test pixel $\mathbf{p}_1$ is determined as the class with minimum E_m using the Eqn. (5.19) :

$$E_m = \left\| \tilde{\mathbf{P}} - \mathbf{X}_m \hat{\mathbf{A}}_m \right\|_F \quad where \quad m = 1, 2, \dots, M \quad (5.18)$$

$$Class(\mathbf{p}_1) = \arg \min(E_m) \quad (5.19)$$

5.3 Simulation Setup and Experimental Results

In this section, the proposed 3SM-JSRC method is tested on two hyperspectral data sets called Indian Pines and University of Pavia. The methods in the literature called SVM [24], SRC [42], JSRC [42] and NLW-JSRC [57] are used for comparison. All experiments are repeated 10 times with different training samples and then the results are averaged.

5.3.1 Simulation Results on Indian Pines

In all studies conducted on the Indian Pines data set, 10% of total samples of each class are used for training and the remaining 90% are used for testing. Detailed information about classes and samples is given in Table 5.1.

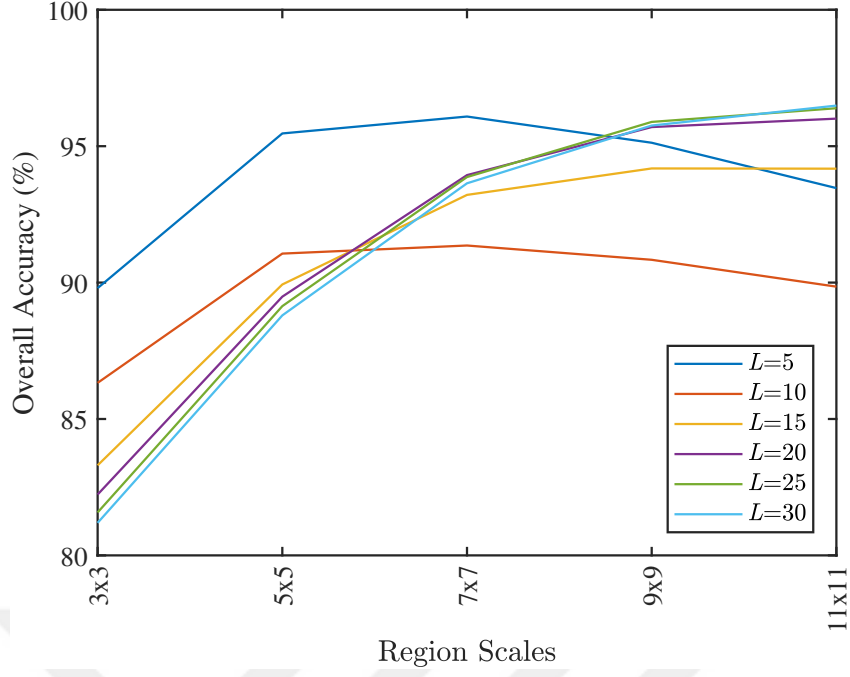
Before comparison, the parameter setting should be made to ensure that the classification results of the other methods and the proposed method are the highest. Gaussian radial basis function (RBF) kernel are used for SVM and 5-fold cross validation is performed to select the best parameters. In the SRC method, the sparsity level of $L=5$ is used. For the JSRC, region scale is selected as 9×9 and sparsity level is chosen as 30. In the NLW-JSRC, the region scale is set as 9×9 ,

Table 5.1 The number of train and test samples for each class used in the experiments for the Indian Pines data set.

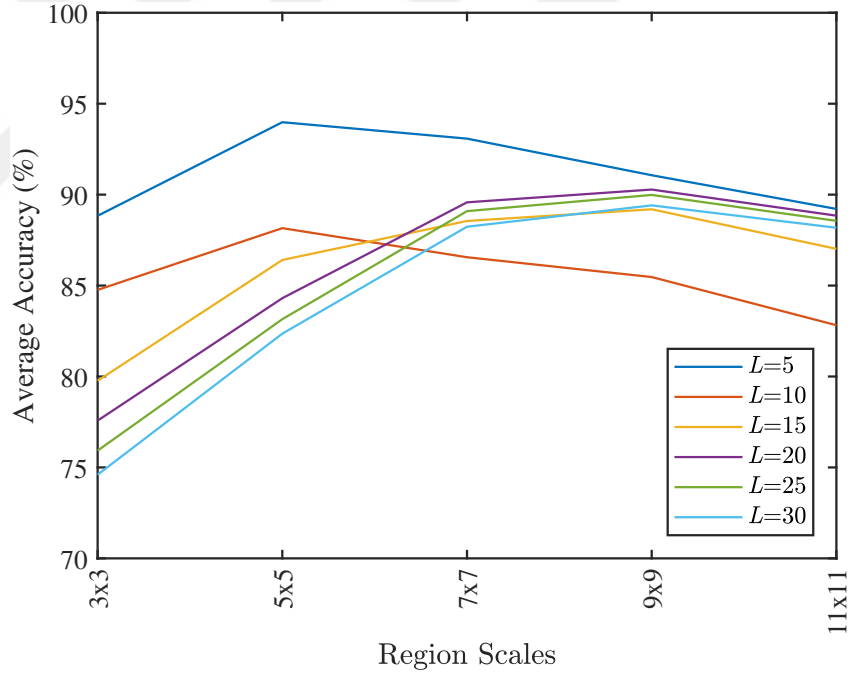
Class	Name	Train	Test
1	Alfalfa	6	48
2	Corn-notill	144	1290
3	Corn-mintill	84	750
4	Corn	24	210
5	Grass-pasture	50	447
6	Grass-trees	75	672
7	Grass-pasture-mowed	3	23
8	Hay-windrowed	49	440
9	Oats	2	18
10	Soybean-notill	97	871
11	Soybean-mintill	247	2221
12	Soybean-clean	62	552
13	Wheat	22	190
14	Woods	130	1164
15	Buildings-Grass-Trees-Drives	38	342
16	Stone-Steel-Towers	10	85
Total		1043	9323

sparsity level is chosen as 30 and the region scale is set as 7×7 for the non-local weight calculation. The proposed 3SM-JSRC method are tested with different region scales and sparsity levels in order to determine the optimum values. The region scales are selected as 3×3 , 5×5 , 7×7 , 9×9 and 11×11 . The sparsity levels are chosen as 5, 10, 15, 20, 25 and 30.

Figure 5.2 shows the OA and AA results for the different region scales and sparsity levels specified. As can be seen, there is a decrease in AA results for all sparsity levels between 9×9 and 11×11 region scales. An increase in both OA and AA results is observed between 3×3 and 5×5 region scales. Therefore, region scales of 5×5 , 7×7 or 9×9 must be selected. Although the OA result at 9×9 scale and sparsity level of 25 is close to the results of 5×5 and 7×7 scales, the AA result at 9×9 scale is lower than the results of 5×5 and 7×7 scales. Hence, only 5×5 and 7×7 region scales remain for selection. For both region scales, the results of OA and AA are higher at the sparsity level of $L=5$. Since 7×7 region scale gives higher OA than 5×5 region scale, the optimal parameters are selected as 7×7 region scale and sparsity level $L=5$. The higher OA result indicates that the number of correctly classified test samples is higher.



(a)



(b)

Figure 5.2 The results obtained using different region scales and sparsity levels for the Indian Pines data set; (a) OA, (b) AA.

After the parameter adjustments are done, a fair comparison is made between the proposed 3SM-JSRC and SVM, SRC, JSRC and NLW-JSRC methods. Classification accuracy results are presented in Table 5.2. As can be seen, the proposed

Table 5.2 Classification accuracies (in %) obtained by the SVM, SRC, JSRC, NLW-JSRC and the proposed 3SM-JSRC methods for the Indian Pines.

Class	SVM	SRC	JSRC	NLW-JSRC	3SM-JSRC
1	76.67	68.33	86.67	85.83	96.88
2	81.01	65.15	93.74	94.53	93.90
3	76.65	61.55	89.91	89.29	93.68
4	71.00	43.33	93.81	91.52	89.67
5	94.97	89.08	93.42	97.23	96.00
6	95.67	94.94	99.14	99.88	98.44
7	79.57	82.17	51.74	33.04	86.96
8	97.48	97.25	99.80	100	100
9	71.11	41.67	4.44	23.89	51.67
10	70.54	72.45	89.07	86.99	92.92
11	85.41	76.47	97.25	98.04	97.35
12	86.68	59.22	89.00	94.49	91.47
13	99.16	97.74	99.21	99.74	98.79
14	95.40	93.77	99.24	99.24	99.66
15	61.87	44.71	98.45	93.71	94.74
16	89.06	87.53	92.24	98.24	90.94
OA	84.83	75.91	94.86	95.31	95.83
AA	83.26	73.46	86.07	86.60	92.06
$\mathcal{K}$	82.69	72.49	94.13	94.64	95.24

3SM-JSRC overcomes other methods in terms of OA, AA and $\mathcal{K}$ values. The classification accuracy results in SVM and SRC, which do not use spatial information, are lower than JSRC, NLW-JSRC and 3SM-JSRC methods, which use spatial information. Therefore, it can be easily observed that use of spatial information increases the results of classification accuracy. When it comes to the JSRC, which does not have any selection process, 3SM-JSRC has better accuracy results than JSRC in terms of OA, AA and $\mathcal{K}$. The OA, AA and $\mathcal{K}$ results of the method 3SM-JSRC are nearly 1% , 6% and 1% higher than the results of the JSRC, respectively. According to NLW-JSRC, where weight is assigned to neighboring pixels, there is no significant differences in OA and $\mathcal{K}$ values between NLW-JSRC and 3SM-JSRC methods. However, AA result of the 3SM-JSRC is nearly 6% higher than the result of the NLW-JSRC.

For a better visualization of the classification results, the ground truth map of the Indian Pines data, the classification maps obtained by SVM, SRC, JSRC, NLW-

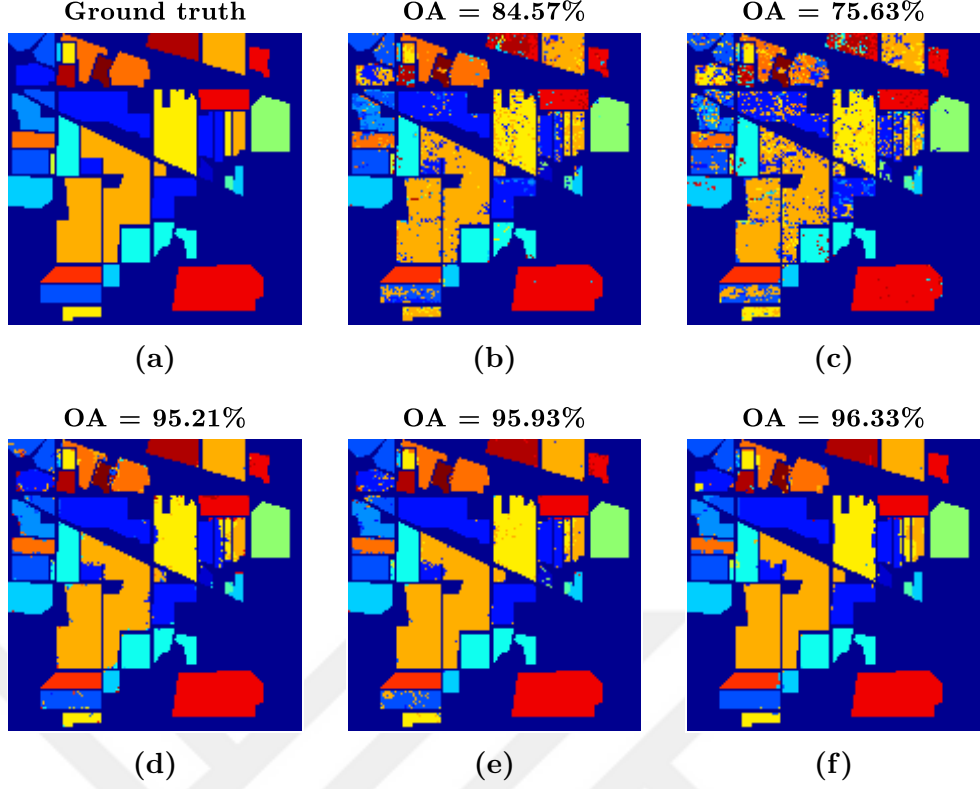


Figure 5.3 Indian Pines data set: (a) Ground truth; Classification maps and OA results obtained by the (b) SVM, (c) SRC, (d) JSRC, (e) NLW-JSRC, (f) 3SM-JSRC methods.

JSRC and 3SM-JSRC are given in Figure 5.3. As can be seen, the classification map of the 3SM-JSRC better coincides with the ground truth map. The maps of the SVM and SRC look so noisy because the spatial information in HSI is not incorporated. However, most of the noise is removed by the use of spatial information as can be seen in the maps of the JSRC, NLW-JSRC and 3SM-JSRC, respectively. In addition, most of the misclassifications seen in JSRC and NLW-JSRC are improved by the proposed 3SM-JSRC. This is because the neighbor pixels belonging to different classes in a window are eliminated in 3SM-JSRC by the use of spectral matching methods.

5.3.2 Simulation Results on University of Pavia

Like the Indian Pines data set, 10% of total samples of each class are used for training and the remaining 90% are used for testing for the experiments conducted on the University of Pavia data set. Detailed information about classes and samples is given in Table 5.3.

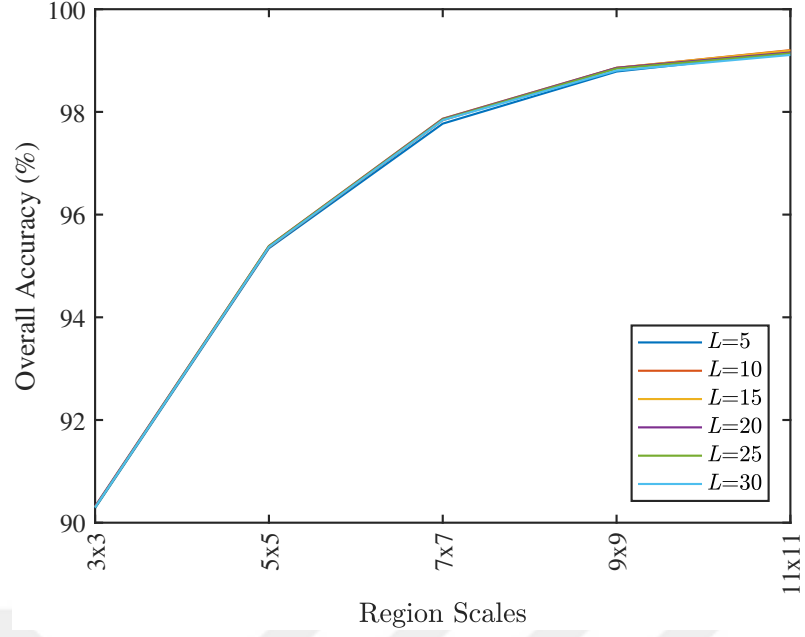
Table 5.3 The number of train and test samples for each class used in the experiments for the University of Pavia data set.

Class	Name	Train	Test
1	Asphalt	664	5967
2	Meadows	1865	16784
3	Gravel	210	1889
4	Trees	307	2757
5	Painted metal sheets	135	1210
6	Bare Soil	503	4526
7	Bitumen	133	1197
8	Self-Blocking Bricks	369	3313
9	Shadows	95	852
Total		4281	38495

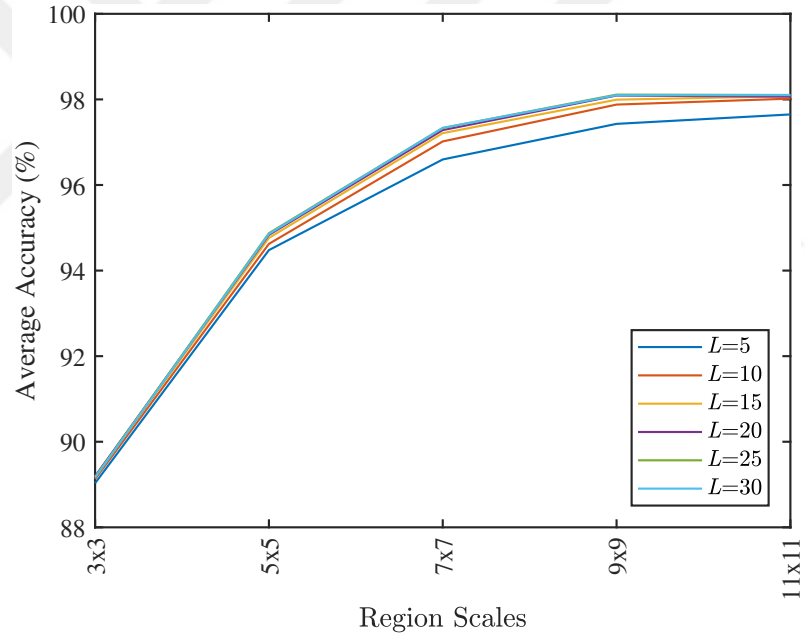
Before comparison, the parameter adjustments of the methods should be made to ensure that the classification results of the other methods and the proposed method are the highest. Gaussian radial basis function (RBF) kernel are used for SVM and 5-fold cross validation is performed to select the best parameters. In the SRC method, the sparsity level of $L=5$ is used. For the JSRC method, region scale is selected as 3×3 and sparsity level is chosen as 10. In the NLW-JSRC method, the region scale is set as 5×5 , sparsity level is chosen as 15 and the region scale is set as 7×7 for the non-local weight calculation. The proposed 3SM-JSRC method are tested with different region scales and sparsity levels in order to determine the optimum values like done for the Indian Pines data set. The region scales are selected as 3×3 , 5×5 , 7×7 , 9×9 and 11×11 . The sparsity levels are chosen as 5, 10, 15, 20, 25 and 30.

Figure 5.4 shows the OA and AA results obtained with these parameters. As can be seen, both OA and AA results are increasing for all sparsity levels as region scales and sparsity levels increases. However, when the region scale goes from 9×9 to 11×11 , accrual in AA results decreases compared to other region scales and reaches to maximum values at 11×11 . In Figure 5.4(a), the OA result reaches the highest value for the $L=15$ sparsity level at 11×11 scale. On the other hand, in the AA results, the values are almost same at 11×11 region scale for the all sparsity levels except $L=5$. For this reason, 11×11 scale and $L=15$ are selected as the optimum values for this data set to obtain highest classification accuracies.

After the parameter tunings are made, a fair comparison is performed between



(a)



(b)

Figure 5.4 The results obtained using different region scales and sparsity levels for the University of Pavia data set; (a) OA, (b) AA.

the proposed 3SM-JSRC and the other methods. Classification accuracy results are presented in Table 5.4. As can be seen, OA, AA and $\mathcal{K}$ values are significantly higher in 3SM-JSRC compared to other methods. The OA, AA and $\mathcal{K}$ results of 3SM-JSRC are approximately 7%, 7% and 9.50% better than the results of JSRC, respectively. In the same way, the OA, AA and $\mathcal{K}$ results of 3SM-JSRC

Table 5.4 Classification accuracies (in %) obtained by the SVM, SRC, JSRC, NLW-JSRC and the proposed 3SM-JSRC methods for the University of Pavia.

Class	SVM	SRC	JSRC	NLW-JSRC	3SM-JSRC
1	91.53	77.18	82.14	93.30	98.73
2	97.19	94.97	98.79	99.60	99.94
3	70.70	61.90	83.34	84.66	99.33
4	93.80	84.54	91.58	94.60	95.50
5	99.51	99.49	100	100	99.98
6	75.92	56.36	79.54	73.29	99.51
7	83.13	78.59	94.06	92.44	99.56
8	87.60	76.88	91.34	83.39	99.62
9	99.07	94.92	94.99	97.51	82.58
OA	91.12	83.38	91.84	92.79	98.93
AA	88.72	80.54	90.64	90.98	97.19
$\mathcal{K}$	88.13	77.64	89.09	90.30	98.58

are almost 6%, 6% and 8% better than the results of NLW-JSRC, respectively. Since spatial information is not used in SVM and SRC methods, classification accuracy results are low when compared to the other 3 methods. According to JSRC and NLW-JSRC methods, 3SM-JSRC has very high classification results because of the large region scale for this data set. In JSRC and NLW-JSRC, spatial information could not be used sufficiently in small region scales and so high classification accuracy values could not be obtained. In the 3SM-JSRC, spatial information is efficiently used thanks to the large region scale. In addition, the pixels of different classes in the window are eliminated by spectral matching methods. In this way, the classification accuracy results have reached very high values.

For a better visualization of the classification results, the ground truth map of the University of Pavia data, the classification maps obtained by SVM, SRC, JSRC, NLW-JSRC and 3SM-JSRC are given in Figure 5.5. As can be seen, the classification map of the 3SM-JSRC method better coincides with the ground truth map and most of the classes almost match with the ground truth map completely. Although the classification result of the class "*Bare soil*" is quite noisy for SVM, SRC, JSRC and NLW-JSRC, the result of this class in the 3SM-JSRC method overlap well with the ground truth map. Similarly, if the classification results of the "*Gravel*" and "*Self-blocking bricks*" classes are examined on classification maps, there are noises in the comparison methods other than 3SM-JSRC.

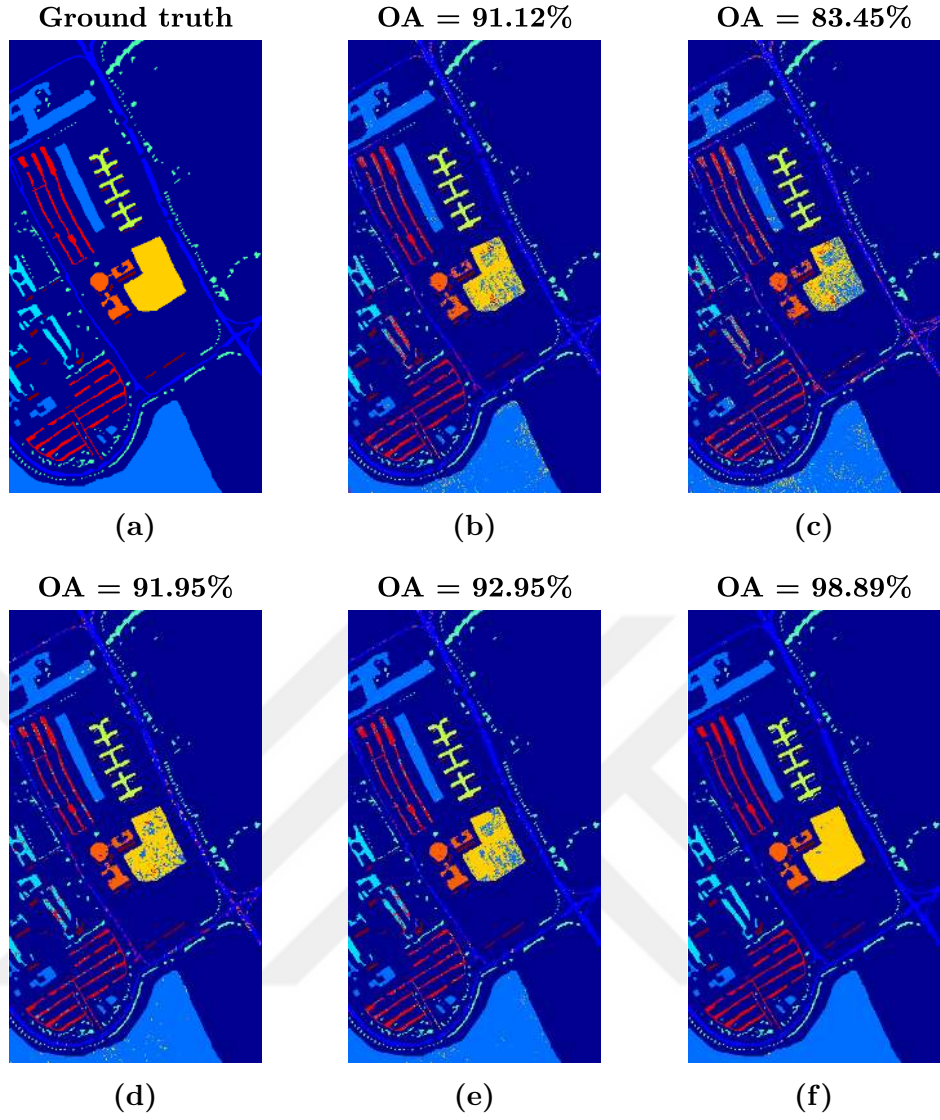


Figure 5.5 University of Pavia data set: (a) Ground truth; Classification maps and OA results obtained by the (b) SVM, (c) SRC, (d) JSRC, (e) NLW-JSRC, (f) 3SM-JSRC methods.

In 3SM-JSRC, the results of these classes are almost identical to those in the ground truth map. This is because the neighbor pixels belonging to different classes in a window are eliminated in 3SM-JSRC by the use of spectral matching methods. In this way, classification accuracy results of the classes become so high.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

In the scope of this thesis, two new sparse representation based HSI classification methods have been developed. They perform better classification performance than the state-of-the-art HSI classification methods proposed in the literature.

The first method called MSS-GF utilizes spatial and edge information effectively in HSI. Instead of using multiscale superpixels as certain numbers, different region scales have been used to construct multiscale superpixels. In this way, small and large local regions are formed to acquire spatial information well. Each superpixel in a segmentation map has been classified using SOMP algorithm and this process has been repeated for all segmentation maps. After all the classification maps were constructed, guided filter has been applied on the classification maps to improve the misclassifications near the edges. Then, the label of a pixel was determined by applying majority voting rule on the final classification maps. Experimental results show that the MSS-GF method has left behind some methods in the literature even in low train sample percentages. In addition, the OA classification result obtained by multiscale segmentation were higher for all two hyperspectral data sets compared to the OAs obtained by single-scale superpixel segmentations.

The second method called 3SM-JSRC uses three spectral matching methods (Euclidean distance, Pearson correlation coefficient and Spectral angle mapper) simultaneously to find the similarity between test pixel and its each neighbor pixel in a window centered around the test pixel. Using these similarity measurements and specified thresholds, the neighbor pixels belonging to different classes were discarded. In this way, only the test pixel and its neighbors with the same class were solved by the SOMP algorithm. Experimental results show that the 3SM-JSRC algorithm has increased the classification accuracy when compared

to classical JSRC method.

6.2 Future Work

The training dictionary has been randomly picked from HSI and so its size may be large and it adversely affect the results for the both data sets. Therefore, an efficient training algorithm for structured dictionary matrix used in the training stage may improve the results. In addition, when determining the class label of a test pixel, the similarities between the pixel groups and the training samples may be calculated with the help of a weight function. Then, these weights may be inserted into the residual value calculation to improve the classification accuracy results.

For the MSS–GF method, a feature extraction method that extracts local features well compared to PCA may increase the classification accuracy. In addition, if an efficient superpixel segmentation method compared to ERS is applied, then classification results will increase more. So, efficient local feature extraction and superpixel segmentation methods will be adopted to MSS–GF method for the future work.

For the 3SM–JSRC method, classification accuracy results can be increased by using matching methods that can better find out the similarity between pixels in addition to the different weight assignment methods.

REFERENCES

- [1] Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., and Abbas, A. (2018). Modern trends in hyperspectral image analysis: A review, *IEEE Access*. **6**, 14118–14129.
- [2] Shaw, G. and Manolakis, D. (2002). Signal processing for hyperspectral image exploitation, *IEEE Signal Processing Magazine*. **19(1)**, 12–16.
- [3] Kale, K. V., Solankar, M. M., Nalawade, D. B., Dhumal, R. K., and Gite, H. R. (2017). A research review on hyperspectral data processing and analysis algorithms, *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*. **87(4)**, 541–555.
- [4] Ghamisi, P., et al. (2018). New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, markov random fields, segmentation, sparse representation, and deep learning, *IEEE Geoscience and Remote Sensing Magazine*. **6(3)**, 10–43.
- [5] Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., and Sousa, J. J. (2017). Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry, *Remote Sensing*. **9(11)**.
- [6] Dundar, T. and Ince, T. (2019). Sparse representation-based hyperspectral image classification using multiscale superpixels and guided filter, *IEEE Geoscience and Remote Sensing Letters*. **16(2)**, 246–250.
- [7] Li, W., Wu, G., Zhang, F., and Du, Q. (2017). Hyperspectral image classification using deep pixel-pair features, *IEEE Transactions on Geoscience and Remote Sensing*. **55(2)**, 844–853.
- [8] Fang, L., He, N., Li, S., Ghamisi, P., and Benediktsson, J. A. (2018). Extinction profiles fusion for hyperspectral images classification, *IEEE Transactions on Geoscience and Remote Sensing*. **56(3)**, 1803–1815.
- [9] Tu, B., Zhang, X., Kang, X., Zhang, G., and Li, S. (2019). Density peak-based noisy label detection for hyperspectral image classification, *IEEE Transactions on Geoscience and Remote Sensing*. **57(3)**, 1573–1584.
- [10] Zhang, Y., Du, B., Zhang, L., and Liu, T. (2017). Joint sparse representation and multitask learning for hyperspectral target detection, *IEEE Transactions on Geoscience and Remote Sensing*. **55(2)**, 894–906.

- [11] Matteoli, S., Diani, M., and Corsini, G. (2018). Automatic target recognition within anomalous regions of interest in hyperspectral images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **11(4)**, 1056–1069.
- [12] Zhu, D., Du, B., and Zhang, L. (2019). Binary-class collaborative representation for target detection in hyperspectral images, *IEEE Geoscience and Remote Sensing Letters*. 1–5.
- [13] Li, W., Wu, G., and Du, Q. (2017). Transferred deep learning for anomaly detection in hyperspectral imagery, *IEEE Geoscience and Remote Sensing Letters*. **14(5)**, 597–601.
- [14] Díaz, M., Guerra, R., López, S., and Sarmiento, R. (2018). An algorithm for an accurate detection of anomalies in hyperspectral images with a low computational complexity, *IEEE Transactions on Geoscience and Remote Sensing*. **56(2)**, 1159–1176.
- [15] Yuan, Y., Ma, D., and Wang, Q. (2019). Hyperspectral anomaly detection via sparse dictionary learning method of capped norm, *IEEE Access*. **7**, 16132–16144.
- [16] Gong, M., Li, H., Luo, E., Liu, J., and Liu, J. (2017). A multiobjective cooperative coevolutionary algorithm for hyperspectral sparse unmixing, *IEEE Transactions on Evolutionary Computation*. **21(2)**, 234–248.
- [17] Zhou, Y., Rangarajan, A., and Gader, P. D. (2018). A gaussian mixture model representation of endmember variability in hyperspectral unmixing, *IEEE Transactions on Image Processing*. **27(5)**, 2242–2256.
- [18] Hong, D., Yokoya, N., Chanussot, J., and Zhu, X. X. (2019). An augmented linear mixing model to address spectral variability for hyperspectral unmixing, *IEEE Transactions on Image Processing*. **28(4)**, 1923–1938.
- [19] Ghamisi, P., Plaza, J., Chen, Y., Li, J., and Plaza, A. J. (2017). Advanced spectral classifiers for hyperspectral images: A review, *IEEE Geoscience and Remote Sensing Magazine*. **5(1)**, 8–32.
- [20] Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges, *IEEE Geoscience and Remote Sensing Magazine*. **1(2)**, 6–36.
- [21] Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., and Plaza, A. (2017). Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art, *IEEE Geoscience and Remote Sensing Magazine*. **5(4)**, 37–78.
- [22] Plaza, A., Martinez, P., Plaza, J., and Perez, R. (2003). Spatial/spectral analysis of hyperspectral image data, *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*. 298–307.

- [23] Soille, P. (2003). *Morphological Image Analysis: Principles and Applications*. Springer-Verlag: Berlin.
- [24] Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*. **42(8)**, 1778–1790.
- [25] Chi, M. and Bruzzone, L. (2007). Semisupervised classification of hyperspectral images by svms optimized in the primal, *IEEE Transactions on Geoscience and Remote Sensing*. **45(6)**, 1870–1880.
- [26] Bruzzone, L., Chi, M., and Marconcini, M. (2006). A novel transductive svm for semisupervised classification of remote-sensing images, *IEEE Transactions on Geoscience and Remote Sensing*. **44(11)**, 3363–3373.
- [27] Scholkopf, B. and Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press: Cambridge.
- [28] Ma, L., Crawford, M. M., and Tian, J. (2010). Local manifold learning-based k -nearest-neighbor for hyperspectral image classification, *IEEE Transactions on Geoscience and Remote Sensing*. **48(11)**, 4099–4109.
- [29] Goel, P. K., Prasher, S. O., Patel, R. M., Landry, J. A., Bonnell, R. B., and Viau, A. A. (2003). Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn, *Computer and Electronics in Agriculture*. **39(2)**, 67 – 93.
- [30] Chen, J. and Wang, R. (2007). A pairwise decision tree framework for hyperspectral classification, *International Journal of Remote Sensing*. **28(12)**, 2821–2830.
- [31] Wang, M., Gao, K., Wang, L., and Miu, X. (2012). A novel hyperspectral classification method based on c5.0 decision tree of multiple combined classifiers, *2012 Fourth International Conference on Computational and Information Sciences*. 373–376.
- [32] Ham, J., Chen, Y., Crawford, M. M., and Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing*. **43(3)**, 492–501.
- [33] Dalponte, M., Ørka, H. O., Gobakken, T., Gianelle, D., and Næsset, E. (2013). Tree Species Classification in Boreal Forests With Hyperspectral Data, *IEEE Transactions on Geoscience and Remote Sensing*. **51(5)**, 2632–2645.
- [34] Wei, W., Zhang, L., Tian, C., Plaza, A., and Zhang, Y. (2017). Structured sparse coding-based hyperspectral imagery denoising with intracluster filtering, *IEEE Transactions on Geoscience and Remote Sensing*. **55(12)**, 6860–6876.

- [35] Zhuang, L. and Bioucas-Dias, J. M. (2018). Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **11(3)**, 730–742.
- [36] He, W., Zhang, H., and Zhang, L. (2017). Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing*. **55(7)**, 3909–3921.
- [37] Gong, M., Li, H., Luo, E., Liu, J., and Liu, J. (2017). A multiobjective cooperative coevolutionary algorithm for hyperspectral sparse unmixing, *IEEE Transactions on Evolutionary Computation*. **21(2)**, 234–248.
- [38] Niu, Y. and Wang, B. (2017). Extracting target spectrum for hyperspectral target detection: An adaptive weighted learning method using a self-completed background dictionary, *IEEE Transactions on Geoscience and Remote Sensing*. **55(3)**, 1604–1617.
- [39] Lu, X., Zhang, W., and Li, X. (2018). A hybrid sparsity and distance-based discrimination detector for hyperspectral images, *IEEE Transactions on Geoscience and Remote Sensing*. **56(3)**, 1704–1717.
- [40] Lai, Z., Wong, W. K., Xu, Y., Zhao, C., and Sun, M. (2014). Sparse alignment for robust tensor learning, *IEEE Transactions on Neural Networks and Learning Systems*. **25(10)**, 1779–1792.
- [41] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. (2009). Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **31(2)**, 210–227.
- [42] Chen, Y., Nasrabadi, N., and Tran, T. (2011). Hyperspectral Image Classification Using Dictionary-Based Sparse Representation, *IEEE Transactions on Geoscience and Remote Sensing*. **49(10)**, 3973–3985.
- [43] Chen, Y., Nasrabadi, N. M., and Tran, T. D. (2013). Hyperspectral image classification via kernel sparse representation, *IEEE Transactions on Geoscience and Remote Sensing*. **51(1)**, 217–231.
- [44] Qian, Y., Ye, M., and Zhou, J. (2013). Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features, *IEEE Transactions on Geoscience and Remote Sensing*. **51(4)**, 2276–2291.
- [45] Srinivas, U., Chen, Y., Monga, V., Nasrabadi, N. M., and Tran, T. D. (2013). Exploiting sparsity in hyperspectral image classification via graphical models, *IEEE Geoscience and Remote Sensing Letters*. **10(3)**, 505–509.
- [46] Ghamisi, P., Benediktsson, J. A., and Ulfarsson, M. O. (2014). Spectral-spatial classification of hyperspectral images based on hidden markov random fields, *IEEE Transactions on Geoscience and Remote Sensing*. **52(5)**, 2565–2574.

- [47] Chen, Y., Jiang, H., Li, C., Jia, X., and Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks, *IEEE Transactions on Geoscience and Remote Sensing*. **54**(10), 6232–6251.
- [48] Ghamisi, P., Dalla Mura, M., and Benediktsson, J. A. (2015). A survey on spectral–spatial classification techniques based on attribute profiles, *IEEE Transactions on Geoscience and Remote Sensing*. **53**(5), 2335–2353.
- [49] Ghamisi, P., Souza, R., Benediktsson, J. A., Rittner, L., Lotufo, R., and Zhu, X. X. (2016). Hyperspectral data classification using extended extinction profiles, *IEEE Geoscience and Remote Sensing Letters*. **13**(11), 1641–1645.
- [50] Fu, W., Li, S., Fang, L., Kang, X., and Benediktsson, J. A. (2016). Hyperspectral image classification via shape-adaptive joint sparse representation, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **9**(2), 556–567.
- [51] Fang, L., Li, S., Kang, X., and Benediktsson, J. A. (2015). Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model, *IEEE Transactions on Geoscience and Remote Sensing*. **53**(8), 4186–4201.
- [52] Li, J., Zhang, H., and Zhang, L. (2015). Efficient superpixel-level multi-task joint sparse representation for hyperspectral image classification, *IEEE Transactions on Geoscience and Remote Sensing*. **53**(10), 5338–5351.
- [53] Fang, L., Li, S., Kang, X., and Benediktsson, J. A. (2014). Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation, *IEEE Transactions on Geoscience and Remote Sensing*. **52**(12), 7738–7749.
- [54] Fu, W., Li, S., Fang, L., Kang, X., and Benediktsson, J. A. (2016). Hyperspectral image classification via shape-adaptive joint sparse representation, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **9**(2), 556–567.
- [55] Fang, L., Li, S., Kang, X., and Benediktsson, J. A. (2015). Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model, *IEEE Transactions on Geoscience and Remote Sensing*. **53**(8), 4186–4201.
- [56] Li, J., Zhang, H., and Zhang, L. (2015). Efficient superpixel-level multi-task joint sparse representation for hyperspectral image classification, *IEEE Transactions on Geoscience and Remote Sensing*. **53**(10), 5338–5351.
- [57] Zhang, H., Li, J., Huang, Y., and Zhang, L. (2014). A nonlocal weighted joint sparse representation classification method for hyperspectral imagery, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. **7**(6), 2056–2065.

- [58] Tu, B., Zhang, X., Kang, X., Zhang, G., Wang, J., and Wu, J. (2018). Hyperspectral image classification via fusing correlation coefficient and joint sparse representation, *IEEE Geoscience and Remote Sensing Letters*. **15**(3), 340–344.
- [59] Bai, J., Zhang, W., Gou, Z., and Jiao, L. (2017). Nonlocal-similarity-based sparse coding for hyperspectral imagery classification, *IEEE Geoscience and Remote Sensing Letters*. **14**(9), 1474–1478.
- [60] Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images, *IEEE Transactions on Image Processing*. **16**(5), 1395–1411.
- [61] Buades, A., Coll, B., and Morel, J. (2005). A non-local algorithm for image denoising, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. 60–65.
- [62] Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2009). Non-local sparse models for image restoration, *2009 IEEE 12th International Conference on Computer Vision*. 2272–2279.
- [63] Dong, W., Zhang, L., Shi, G., and Wu, X. (2011). Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, *IEEE Transactions on Image Processing*. **20**(7), 1838–1857.
- [64] Rodarmel, C. and Shan, J. (2002). Principal Component Analysis for Hyperspectral Image Classification, *Surveying and Land Information Science*. **62**(2), 115–122.
- [65] Tharwat, A. (2016). Principal component analysis (pca) : An overview.
- [66] Hyperspectral remote sensing data sets. http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, 01.04.2019.
- [67] Tarabalka, Y., Benediktsson, J. A., and Chanussot, J. (2009). Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques, *IEEE Transactions on Geoscience and Remote Sensing*. **47**(8), 2973–2987.
- [68] Gualtieri, J. A. and Crompton, R. F. (1999). Support vector machines for hyperspectral remote sensing classification, *Proceedings of SPIE 27th AIPR Workshop: Advances in Computer Assisted Recognition*. 221–232.
- [69] Foody, G. (2004). Thematic map comparison: evaluating the statistical significance of differences in classification accuracy, *Photogrammetric Engineering and Remote Sensing*. **70**(5), 627–633.
- [70] Liu, M.-Y., Tuzel, O., Ramalingam, S., and Chellappa, R. (2011). Entropy rate superpixel segmentation, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2097–2104.
- [71] Kang, X., Li, S., and Benediktsson, J. A. (2014). Spectral-Spatial Hyperspectral Image Classification With Edge-Preserving Filtering, *IEEE Transactions on Geoscience and Remote Sensing*. **52**(5), 2666–2677.

- [72] He, K., Sun, J., and Tang, X. (2013). Guided Image Filtering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **35(6)**, 1397–1409.
- [73] Tropp, J. A. and Gilbert, A. C. (2007). Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit, *IEEE Transactions on Information Theory*. **53(12)**, 4655–4666.
- [74] Keshava, N. (2004). Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries, *IEEE Transactions on Geoscience and Remote Sensing*. **42(7)**, 1552–1565.
- [75] Yan, L., Cui, M., and Prasad, S. (2018). Joint euclidean and angular distance-based embeddings for multisource image analysis, *IEEE Geoscience and Remote Sensing Letters*. **15(7)**, 1110–1114.
- [76] Petropoulos, G. P., Vadrevu, K. P., and Kalaitzidis, C. (2013). Spectral angle mapper and object-based classification combined with hyperspectral remote sensing imagery for obtaining land use/cover mapping in a mediterranean region, *Geocarto International*. **28(2)**, 114–129.

APPENDIX A

SELECTION OF THE REGION SCALES FOR THE PROPOSED MSS-GF ALGORITHM

A.1 University of Pavia

Table A.1 Classification accuracies (%) of different region scales for the University of Pavia.

C	3×3	5×5	7×7	9×9	11×11	13×13	15×15	17×17	19×19
1	66.02	72.30	80.62	85.71	90.34	92.96	94.85	95.59	96.80
2	88.57	92.60	93.96	94.80	95.60	95.58	95.39	95.29	95.33
3	89.76	95.92	98.66	99.71	99.75	99.94	100	99.02	98.87
4	95.26	96.49	96.49	94.67	92.98	90.97	89.97	88.41	85.23
5	99.61	99.48	99.38	99.34	99.40	98.43	99.27	99.65	97.03
6	81.85	91.94	98.38	99.35	99.87	100	99.83	99.98	99.54
7	97.37	99.65	98.93	99.49	99.84	98.81	96.15	100	94.40
8	83.79	90.85	94.08	95.90	97.39	96.36	96.99	98.00	97.37
9	91.00	91.99	95.45	97.03	98.85	89.15	86.58	87.57	79.04
OA	84.88	89.93	93.06	94.48	95.69	95.69	95.78	95.92	95.45
AA	88.14	92.35	95.10	96.22	97.11	95.80	95.45	95.94	93.74
$\mathcal{K}$	79.89	86.55	90.72	92.60	94.21	94.20	94.33	94.52	93.88

In Table A.1, classification accuracies of different region scales are presented for the University of Pavia. Both 3×3 and 5×5 region scales have low OA, AA and $\mathcal{K}$ values (mostly lower than 90%) compared to other scales. So, they are discarded. OA and $\mathcal{K}$ values are starting to increase from 3×3 region scale to 17×17 region scale. But, after this scale, these values and AA are starting to decrease. So, we discarded 19×19 region scale too. As a result, only 7×7, 9×9, 11×11, 13×13, 15×15 and 17×17 region scales were selected for the University of Pavia dataset.

A.2 Indian Pines

Table A.2 Classification accuracies (%) of different region scales for the Indian Pines.

C	3×3	5×5	7×7	9×9	11×11	13×13	15×15	17×17	19×19
1	62.50	82.29	96.25	90.00	96.25	90.00	38.54	67.29	39.17
2	82.57	91.80	92.43	91.59	91.54	86.10	88.15	90.49	67.21
3	83.61	93.63	96.36	99.33	94.43	92.29	91.60	87.25	73.92
4	67.71	86.14	93.52	92.33	81.62	77.00	81.48	69.10	79.38
5	90.94	94.50	97.14	95.77	93.65	93.87	93.60	93.40	92.95
6	97.10	99.02	98.72	97.89	98.32	97.75	97.75	97.63	97.75
7	90.87	98.26	100	100	10.00	10.00	10.00	10.00	10.00
8	97.55	98.30	99.77	99.77	99.34	99.73	99.77	99.77	99.77
9	56.67	93.33	90.00	70.00	70.00	20.00	20.00	20.00	20.00
10	89.78	92.56	94.98	89.80	88.23	84.64	88.58	82.33	67.70
11	89.45	95.38	95.66	95.25	94.73	91.01	88.03	89.08	87.20
12	75.76	92.14	92.93	94.09	91.76	89.96	87.92	66.25	69.75
13	99.58	99.58	99.58	100	100	100	100	100	100
14	96.25	98.26	99.09	99.22	98.64	100	98.84	98.73	100
15	65.44	86.26	97.43	97.19	89.91	87.19	78.57	78.57	78.63
16	83.76	95.06	98.82	97.88	98.00	98.00	97.88	97.88	97.88
OA	87.69	94.48	96.09	95.48	93.81	91.38	90.51	88.71	83.04
AA	83.10	93.53	96.42	94.38	87.28	82.35	78.79	77.99	73.83
$\mathcal{K}$	85.95	93.71	95.55	94.86	92.95	90.19	89.20	87.15	80.70

If the results in the Table A.2 are examined, the accuracy results of the region scales 5×5, 7×7 and 9×9 are considerably higher than the results of the other scales. Therefore, these scales should be selected to achieve high classification accuracy results. The results at 15×15, 17×17 and 19×19 scales are not very satisfactory because in some classes (7, 9) accuracy results are quite low. Therefore, these scales should be eliminated. There are only two remaining scales, which are 3×3 and 11×11. In the Indian Pines data set, the number of samples

in the classes 7 and 9 is low, so scales with satisfactory accuracy results should be selected in these classes. A value of 90.87 is obtained for class 7 in the 3×3 region scale. Also, satisfactory results are mostly observed in the other classes. Therefore, this scale should be included. For the 11×11 scale, very good accuracy results are observed in the classes other than class 7. So, this scale should be chosen. As a result, 3×3 , 5×5 , 7×7 , 9×9 and 11×11 region scales are selected for the Indian Pines data set.



CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name : DÜNDAR, Tuğcan
Nationality : Turkish (TC)
Date and Place of Birth : 23 March 1993, Gaziantep/Turkey
Office Phone : +90 342 317 2146
E-mail : dundar@gantep.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
M.Sc.	University of Gaziantep	2019
B.Sc.	University of Gaziantep	2016

WORK EXPERIENCE

Year	Place	Enrollment
2017-present	University of Gaziantep	Research Assistant

PUBLICATIONS

- [1] Dundar, T., Ince, T. (2019). Sparse representation-based hyperspectral image classification using multiscale superpixels and guided filter, *IEEE Geoscience and Remote Sensing Letters*. **16(2)**, 246–250.
- [2] Dundar, T., Ince, T. (2019). Classification of hyperspectral images with adaptive neighborhood selection and weighting methods, *Journal Geomatik*. (In press).
- [3] Dundar, T., Ince, T. (2019). Sparse representation-based classification of hyperspectral images using spectral matching methods, *Journal Geomatik*. (In press).