Old Dominion University
ODU Digital Commons

Electrical & Computer Engineering Theses & Disssertations

Electrical & Computer Engineering

Spring 2016

Sparse Coding Based Feature Representation Method for Remote Sensing Images

Ender Oguslu
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/ece_etds

Part of the Electrical and Computer Engineering Commons, and the Remote Sensing Commons

Recommended Citation
Oguslu, Ender. "Sparse Coding Based Feature Representation Method for Remote Sensing Images" (2016). Doctor of Philosophy (PhD), dissertation, Electrical/Computer Engineering, Old Dominion University, DOI: 10.25777/ayt5-ks59
https://digitalcommons.odu.edu/ece_etds/4

This Dissertation is brought to you for free and open access by the Electrical & Computer Engineering at ODU Digital Commons. It has been accepted for inclusion in Electrical & Computer Engineering Theses & Disssertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

SPARSE CODING BASED FEATURE REPRESENTATION METHOD FOR REMOTE SENSING IMAGES

by

Ender Oguslu
B.S. August 1995, Turkish Air Force Academy, Turkey
M.S. May 2006, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
May 2016

Approved by:

Jiang Li (Director)

Dimitrie C. Popescu (Member)

Khan Iftekharuddin (Member)

Richard C. Zimmerman (Member)

ABSTRACT

SPARSE CODING BASED FEATURE REPRESENTATION METHOD
FOR REMOTE SENSING IMAGES

Ender Oguslu
Old Dominion University, 2016
Director: Dr. Jiang Li

In this dissertation, we study sparse coding based feature representation method for the classification of multispectral and hyperspectral images (HSI). The existing feature representation systems based on the sparse signal model are computationally expensive, requiring to solve a convex optimization problem to learn a dictionary. A sparse coding feature representation framework for the classification of HSI is presented that alleviates the complexity of sparse coding through sub-band construction, dictionary learning, and encoding steps. In the framework, we construct the dictionary based upon the extracted sub-bands from the spectral representation of a pixel. In the encoding step, we utilize a soft threshold function to obtain sparse feature representations for HSI. Experimental results showed that a randomly selected dictionary could be as effective as a dictionary learned from optimization.

The new representation usually has a very high dimensionality requiring a lot of computational resources. In addition, the spatial information of the HSI data has not been included in the representation. Thus, we modify the framework by incorporating the spatial information of the HSI pixels and reducing the dimension of the new sparse representations. The enhanced model, called sparse coding based dense feature representation (SC-DFR), is integrated with a linear support vector machine (SVM) and a composite kernels SVM (CKSVM) classifiers to discriminate different types of land cover. We evaluated the proposed algorithm on three well known HSI datasets and compared our method to four recently developed classification methods:

SVM, CKSVM, simultaneous orthogonal matching pursuit (SOMP) and image fusion and recursive filtering (IFRF). The results from the experiments showed that the proposed method can achieve better overall and average classification accuracies with a much more compact representation leading to more efficient sparse models for HSI classification.

To further verify the power of the new feature representation method, we applied it to a pan-sharpened image to detect seafloor scars in shallow waters. Propeller scars are formed when boat propellers strike and break apart seagrass beds, resulting in habitat loss. We developed a robust identification system by incorporating morphological filters to detect and map the scars. Our results showed that the proposed method can be implemented on a regular basis to monitor changes in habitat characteristics of coastal waters.

Copyright, 2016, by Ender Oguslu, All Rights Reserved.

To my wife, son, and daughter,

I thank you for your unconditional love, trust and support.

ACKNOWLEDGMENTS

I would like to offer my sincerest appreciation to my advisor, Dr. Jiang Li for his continuous support, assistance, guidance, and patience to complete this dissertation. I would like to thank the rest of my dissertation committee: Dr. Richard C. Zimmerman, Dr. Khan Iftekharuddin, and Dr. Dimitrie C. Popescu for their insightful comments and encouragement. I would like to also thank to Dr. Oktay Baysal and Dr. Sefer Kurnaz for initiating the collaborative doctoral program between Old Dominion University and Turkish Air Force Aerospace and Space Technologies Institute. Finally, I appreciate the financial support from the Turkish Air Force that funded my doctoral program.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES x

Chapter

1. INTRODUCTION 1
   1.1 PROBLEM STATEMENT 2
   1.2 CONTRIBUTIONS 4
   1.3 ORGANIZATION OF THE DISSERTATION 5

2. RELATED WORK 6
   2.1 INTRODUCTION 6
   2.2 SPARSE CODING BASED FEATURE REPRESENTATION METHODS 7
   2.3 CLASSIFICATION TECHNIQUES 16
   2.4 FEATURE SELECTION 21
   2.5 MORPHOLOGICAL OPERATIONS 22

3. PROPOSED METHOD 27
   3.1 HYPERSPECTRAL IMAGERY INTRODUCTION 28
   3.2 SPECTRAL-BASED SPARSE CODING MODEL 32
   3.3 SPARSE CODING BASED DENSE FEATURE REPRESENTATION MODEL 37
   3.4 SPECTRAL-BASED EXPERIMENTS 39
   3.5 EXPERIMENTAL RESULTS OF SC-DFR METHOD 44
   3.6 DISCUSSIONS 60
   3.7 CONCLUSION 62

4. DETECTION OF SEAGRASS SCARS USING SPARSE CODING AND MORPHOLOGICAL FILTER 63
   4.1 INTRODUCTION OF SCAR DETECTION PROBLEM 63
   4.2 REMOTE SENSING DATA 66
   4.3 METHOD 67
   4.4 RESULTS 75

5. CONCLUSION AND FUTURE WORK 86

REFERENCES 88

Page

VITA 102

LIST OF TABLES

Table Page

1. Class names and number of labeled KSC data. 41
2. Comparison of dictionary learning algorithms. 43
3. Comparison of proposed algorithm and SLLE-W*k*NN classifier. 43
4. Class names and number of labeled Indian Pines data. 47
5. Classification accuracies (%) for Indiana Pines HSI. 48
6. OA of the proposed algorithm with various training data samples. 49
7. Effect of receptive field length on performance. 50
8. Effect of soft thresholding value, *t*. 51
9. Effect of tradeoff value μ on OA. 52
10. Effect of feature selection. 53
11. Class names and number of labeled University of Pavia data. 54
12. Classification accuracies (%) with fixed samples for University of Pavia HSI. 56
13. Classification accuracies (%) with random samples for University of Pavia HSI. 57
14. Class names and number of labeled Center of Pavia data. 59
15. Classification accuracies (%) for Center of Pavia HSI. 60
16. Classification results of the first data set using different amounts of data for training (%) ...77
17. Classification results of the second data set using different amounts of data for training (%)77
18. Detected scars using different amounts of data for training (%) 83

LIST OF FIGURES

Figure Page

1. Feature representation system. 7

2. Representation of soft thresholding. 15

3. Representation of hyperplanes. 18

4. Examples of SE. (a) linear, (b) square, and (c) discrete disk. 23

5. Examples of erosion operations on (a) grey-level, (b) binary image. 23

6. Examples of dilation operations on (a) grey-level, (b) binary image. 24

7. Implementation of opening operation. (a) Original image, (b) image opened by a horizantial linear SE, and (c) image opened by a vertical linear SE. 25

8. Implementation of closing operation. (a) Original image, (b) image closed with a disk SE. 26

9. A distinctive structure of HSI. 29

10. A real HSI cube obtained over the center of Pavia. 29

11. Flowchart of the spectral-based feature representation method. 32

12. Extraction of sub-bands. 33

13. Dictionary learning from random sub-bands. 34

14. Encoding step. 35

15. Pooling step. 36

16. Flowchart of the SC-DFR method. 38

17. A sample band of Kennedy Space Center HSI data. 40

18. A sample band of Indian Pines HSI data. 46

Figure Page

19. Pines image. (a) ground truth, (b) training data, and (c) test data. Classification maps obtained by (d) SVM, (e) CKSVM, (f) SOMP, (g) IFRF, (h) SC-DFR with SVM and (i) SC-DFR with CKSVM. 47

20. A sample band of the University of Pavia HSI data.....54

21. University of Pavia image. (a) ground truth, (b) training data, and (c) test data. Classification maps obtained by (d) SVM, (e) CKSVM, (f) SOMP, (g) IFRF, (h) SC-DFR with SVM and (i) SC-DFR with CKSVM. 55

22. A sample band of the Center of Pavia HSI data. 58

23. Center of Pavia image. (a) ground truth, (b) training data, and (c) test data. Classification maps obtained by (d) SVM, (e) CKSVM, (f) SOMP, (g) IFRF, (h) SC-DFR with SVM and (i) SC-DFR with CKSVM. 58

24. Examples of prop scar. 64

25. Panchromatic image obtained over shallow vegetated area near Hagen's Cove Park.....67

26. Flowchart of the proposed algorithm. 68

27. (a) Green, (b) yellow, (c) red, (d) red edge bands of the multispectral image, and (e)-(h) pan-sharpened multispectral bands after fusion. 69

28. Construction of patches. 71

29. Samples of (a) scar and (b) non-scar blocks.....71

30. Linear type of structural elements in size 5 and various orientations (a) 0°, (b) 45°, (c) 90°, (d) 135°. 73

31. Implementation of "opening by reconstruction" filter on (a) initial map with orientation (b) 30°, (c) 60° and (d) 110°. 74

32. MDP of a scar pixel. 75

33. The ground truth data of the (a) first and (b) second sets.....76

34. Classification maps of test image after the classification step. The maps were obtained with training (a) 10%, (b) 25%, (c) 50% and (d) 75% of the first ground truth set. 79

Figure Page

35. Classification maps of test image after the classification step. The maps were obtained with training (a) 10%, (b) 25%, (c) 50% and (d) 75% of the second ground truth set. 80

36. Final scar maps of test image after the local orientation step. The maps were obtained with training (a) 10%, (b) 25%, (c) 50% and (d) 75% of the first ground truth set. 81

37. Final scar maps of test image after the local orientation step. The maps were obtained with training (a) 10%, (b) 25%, (c) 50% and (d) 75% of the second ground truth data set. 82

Chapter 1

INTRODUCTION

In remote sensing, the reflectance of an object can be characterized by spectral responses over a mixture of wavelengths. The comparability of these responses over distinct wavelengths can be applied to identify different materials. Some materials, such as water, can be identified using broad wavelength ranges in visible and near-infrared portions of the electromagnetic spectrum. On the other hand, more specific classes, such as different vegetation types in a forest, can be broken up using a much finer wavelength range. Therefore, different types of remote sensing images are required to categorize the objects on the ground.

Remote sensing images can be categorized into three different types depending on the number of spectral bands used in the imaging process. The first one is the panchromatic image that consists of a single band. It is captured by a sensor, which has a single channel detector sensitive to radiation within a broad wavelength range. The second type is referred to as the multispectral image which records energy over several separate wavelengths ranges of several spectral resolutions. In multichannel detector, each channel is sensitive to radiation within a narrow wavelength band. The resulting multilayer image contains both the brightness (spatial) and color (spectral) information from the objects. The third type of remote sensing images is the hyperspectral image (HSI) that acquires images in about a hundred or more contiguous spectral bands throughout the visible, near-infrared, and mid-infrared portions of the electromagnetic spectrum.

During the last two decades, significant progress has been made in remote sensing sensors with finer resolutions in both the spatial and spectral domains. The development in panchromatic systems, such as IKONOS, QuickBird, and WorldView-2, can produce an elaborate and accurate mapping of the urban environment with a sub-meter resolution. However, panchromatic images have a low spectral resolution, leading to poor visualization and classification performances [1]. Contrarily, multispectral images are designed to take advantage of the different spectral properties of materials on Earth surface. They have low spatial resolution leading less detail information of targets. Therefore, a panchromatic image has been often fused with a multispectral image using pan-sharpening methods [2, 3] to obtain high spectral and spatial resolution image called pan-sharpened image.

1.1 PROBLEM STATEMENT

Different materials can be identified by classifying remote sensing pixels into different categories based on statistical characteristics of the pixel values. The spectral bands on multispectral and HSI images are essential source of information for the classification of land cover materials. A variety of techniques have been employed, including support vector machines (SVM) [4-6], neural networks [7], and graph-based methods [8] for the classification of multispectral and HSI images. All of these methods have proved to be powerful tools for this purpose. However, the performance of the classification methods is heavily dependent on the feature representation on which they are applied. For that reason, machine learning algorithms are employed to transform a pixel into a new representation that can be effectively exploited in classification tasks. In this dissertation, we focus on the feature representation methods for the classification of HSI and multispectral images.

Sparse coding is a recently developed machine learning technique. In sparse coding, it is assumed that high dimensional natural signals belonging to the same class usually lie in a low-dimensional subspace. Sparse coding reconstructs each unknown input data as a weighted linear combination of a few basis functions of a dictionary learned from the data. The dictionary is intentionally made overcomplete so that many of the reconstruction weights (sparse codes) are zeros leading to a sparse representation for the input data. Extraction of the reconstruction weights and dictionary requires a two-step optimization where alternating minimization is solved over the sparse codes and the dictionary while holding the other fixed. To overcome this difficult computational problem, approximate solutions are considered and various numbers of minimization techniques, such as greedy algorithms or l_1 norm convex relaxation, have been developed [9]. Sparse coding has been successfully employed in many applications, including classification of HSI images. In HSI model, it is usually assumed that high dimensional spectral bands of the pixels in the same class lie in a subspace of reduced dimensionality. Therefore, the HSI model is consistent with the sparse signal model. Initial studies in this direction usually propose different dictionary learning models using sparse coding for the new representation of HSI data [10-14]. However, the feature representation systems based on the sparse signal model are computationally expensive, requiring to solve a convex optimization problem. We need more practical and efficient method to solve this problem.

Besides the computational complexity problem, another problem of sparse coding algorithm is the high dimensionality of the new representation requiring plenty of computational resources. In addition, for the specific classification task, some of the new features may be correlated or redundant, leading to the overfitting problem. This problem occurs when the number of features is larger compared to the number of training samples. Thus, given a classification technique, it is

conceivable to select the best set of features satisfying a given “model selection”. One way to overcome this problem is to use a feature selection step that may reduce the dimensionality of the new pixel representation without performance degradation.

This dissertation also discusses the detection of seafloor propeller seagrass scars in shallow water. The existing techniques usually tend to detect scars based on an object recognition method which is a combination of segmentation and classification steps. However, the success of these methods depends on a segmentation scale that determines the maximum detectable scar size in images. In addition, these methods usually need some manual editing such as field observations and injection of the classification rules that make them semi-automatic.

1.2 CONTRIBUTIONS

In this dissertation, we address the challenge of feature representation problem for the classification of hyperspectral images. We present a sparse coding based dense feature representation (SC-DFR) model which consists of sub-band construction, dictionary learning, encoding and feature selection steps. We systematically integrated multiple sparse representation techniques for accurate HSI image classification. In addition, we experimentally showed that we can reduce the size of the final representation of the pixel, which is a natural challenge of sparse coding approaches, without performance degradation for HSI classification. We also experimentally prove that a randomly selected dictionary could be efficient as a dictionary learned from optimization solutions.

We also introduce a pixel based algorithm to detect and map the seagrass scars. The SC-DFR model extracts the new representation of the pan-sharpened multispectral bands obtained from multispectral and panchromatic images. The classification of the new features successfully detects both bright and dark-toned scars which have very similar pixel values with their

neighborhood structures and are difficult to discriminate. The morphological filtering process eliminates the false positive pixels to improve the detection results achieved by the classification step. The proposed system is fully automatic once the model is trained, requiring no human intervention for scar detection.

1.3 ORGANIZATION OF THE DISSERTATION

This dissertation is organized as follows. Chapter 2 reviews the related works including sparse coding based feature learning techniques, recent sparse coding based HSI feature representation and classification approaches. Chapter 3 presents the proposed SC-DFR method for the classification of HSI. The experiment results on multiple HSI data are also discussed in this part. Chapter 4 introduces the application of the SC-DFR method to the multi-spectral pan-sharpened bands for the detection of seafloor propeller seagrass scars in shallow water. Chapter 5 concludes this dissertation.

Chapter 2

RELATED WORK

2.1 INTRODUCTION

The focus of this dissertation is feature representation of pixels for the classification of HSI and multispectral images. Before getting involved in details, we outline a few explanatory remarks on the classification task and the importance of the feature representation. In machine learning aspect, classification is the process of identifying the class of an input sample containing specific features, on the basis of a training set of data whose class is known. A classification algorithm requires the features which are convenient to process in terms of mathematically and computationally. However, in real world, the image pixels may have complex, noisy, redundant and high dimensional features which can not be categorized easily. Two different vegetation types may have very similar spectral responses, on the other hand, a certain vegetation type may have different spectral responses that make them difficult to categorize into appropriate classes. Thus, in order to model real-world data, it is significant to find new representations from raw data.

In this chapter, we explain some fundamental concepts and related works studied. In this content, sparse coding based feature representation, classification, feature selection and basic morphological operation methods are reviewed, respectively.

2.2 SPARSE CODING BASED FEATURE REPRESENTATION METHODS

Recent researches on sparse coding have attracted increased attention in various computer vision and pattern recognition applications, including image classification [15-17] and HSI classification [10-14, 18, 19]. In sparse coding, a set of basis functions, called as dictionary, is first learned from the data. Then it reconstructs each original data point as a linear combination of the basis functions. With sparse coding, local patterns embedded in data are extracted and composed of sparse vectors in a high-dimensional space, where a linear classifier is sufficient for classification. Sparse coding based feature representation methods can be set apart into two parts (Figure 1): (1) learning part that creates a set of basis functions, dictionary **D,** and (2) encoding part that extracts the new representation of an input data with the learned dictionary [16]. In the following part of this chapter, we first review the dictionary research, then discuss the two steps of sparse coding.

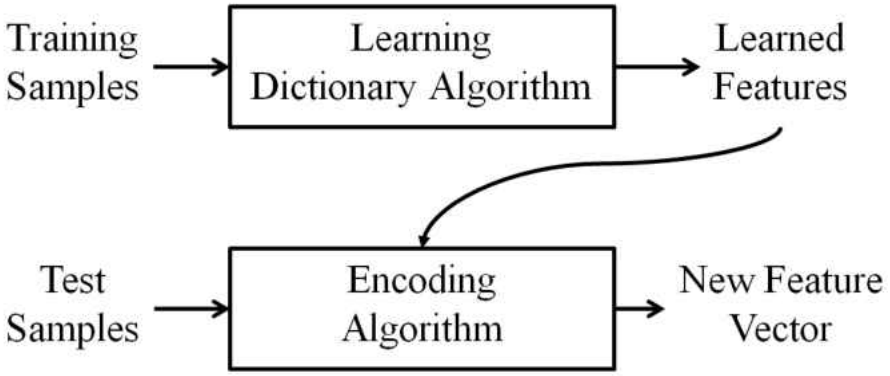


Figure 1: Feature representation system.

2.2.1 REVIEW OF DICTIONARY RESEARCH

Signal transforms have been studied by many researchers since the linear-time invariant operators being introduced. In this regard, the Fast Fourier Transform (FFT) came up as a remarkable tool by Cooley and Turkey in 1965 for signal processing. The Fourier basis represents a signal in global frequency content and describes a signal in the form of a combination of orthogonal waveforms [20]. The K lowest frequency basis functions are used for projection of the signal onto the basis functions to analyze or smooth the signal. Fourier transform assumes a periodic extension of finite signals, which causes a discontinuity at boundaries. Subsequently, the adaptation of anti symmetric extension property to signal produced the Discrete Cosine Transform (DCT) [21], which results in continues boundaries. This made it more desirable than the FFT in practical applications.

Signal approximation in turn paved the way for linear approximation. In particular, given a set of basis functions, a signal is linearly approximated by projecting it into a fixed subset of basis functions where the number of basis functions are smaller than the dimension of the signal. The goal of this undercomplete linear transformation is to capture a significant part of the signal with only a few coefficients and this capability is called as compaction [22]. In fact, the concept of this approximiation is closely linked to sparsity which we will discuss later.

For the revolution of more efficient representations, the data itself was brought to light during the 1970's and 1980's. The statistical tools, Karhunen-Loeve Transform (KLT) [23] and Principal Component Analysis (PCA) [24], were two main research areas in this field. Basis functions in KLT are the first K eigenvectors of the data covariance matrix. KLT is considered to be superior to Fourier transform due to its representational efficiency. However, its complexity is the major drawback.

The robust statistics, which was the powerful approach in statistic research in 1980's, revealed that the sparsity property of data is a promosing tool for data recovery and analysis tasks. This discovery led to the search for more sparse representations and subsequently more efficient transforms. It was conducted by focusing on the non-linear formulations allowing the use of a different set of basis functions to approximate each signal. Due to the achievement of more efficient transforms through robust statistics, modern dictionary designs emerged during the 1980's and 1990's as described as follows [25].

It was found that the locally concentrated basis functions can give more flexible representations and limit the effects of irregularities. The basis functions based on the local signal characteristics become the base for Short Time Fourier Transform (STFT) [26], which is a component of the Fourier transform. The STFT describes the signal as a space-frequency applying the Fourier Transforms locally. The JPEG image compression algorithm [27] is an example of the STFT method named as Gabor transform later. The name was given after Dennis Gabor whose work was based on time-frequency decomposition in 1946 [28]. The discrete version of this transform had successful applications in 1990s [29-30]. The contributions in [31, 32] improved the performance of complex Gabor structures in higher dimensions. This improvement motivated researchers to apply the Gabor dictionaries to image processing tasks such as in [33] and [34].

It was discovered in the 1980's that natural signals and images can be analyzed and described through the multi-scale constructions. Laplacian pyramid [35] is one of the most known structures introduced by multi-scale instructions. In the Laplacian pyramid, each image can be approximated by a series of images in different scales and frequency bands.

Wavelet analysis [24, 36, 37] came into the light in the second half of the 1980's, which is based on a signal combination over a series of translated and dilated varieties of an elementary function proposed in [38]. The wavelet theory was put forward consequently by Meyer, Daubechies, Mallat and others for a complete mathematical framework of continuous and discrete domains. Meyer's work in 1985 [39] revealed that the wavelet framework could produce orthogonal design. The concept of wavelet decomposition was established by Mallat and his colleagues [40-42], which is a multi-resolution expansion on the basis of a pair of localized functions called the scaling function and the mother wavelet. The scaling function is a low frequency signal that presents the approximation of the signal with its translations; on the other hand the mother wavelet is high frequency signal that reveals the signal details with its various scales and translations. Non-linear approximation of the wavelet basis produced optimal piecewise-smooth 1-D signals without prior information of the discontiniuty regions; however, it loses its optimality in higher dimensions resulting in smooth artifacts near edges.

In 1990's, the development of sparsity paved the way in orthogonal bases. The inadequate of sparse representation was due to the small number of basis functions in the dictionary. For providing further sparsity, the idea of adapting basis functions to the signal was proposed in [43]. In this regard, the wavelet packet transform, which is an extension of wavelet transform, was introduced by Coifman, Meyer and Wickerhauser in 1992 [43]. A large number of dyadic time-frequency basis functions obtained from a pair of scaling function and mother wavelet are unified in the wavelet packet dictionary. This process produces an optimized orthogonal sub-dictionary out of the large dictionary for any given signal.

In 1992, Simoncelli *et al.* [44] introduced a dictionary based on the variation of basis functions produced by translation, rotation and scaling deformations. In this work, they used

overcomplete dictionaries in order to acquire the best result. Simoncelli *et al.* also introduced the steerable wavelet transform focusing on steerable filters and localized 2-D wavelet basis functions in various orientations, scales and translations.

Mallat and Zhang [45] proposed a novel sparse signal expansion scheme in 1993. This scheme was based on the selection of a few basis function from an overcomplete dictionary. Then, Chen, Donoho and Saunders [46] proposed their sparsity based algorithm, called as basis pursuit. These two studies showed that a signal could have more than one description in the representation domain and the goal of the task is the determination of the best one. In the study, they also split the sparse representation process into two parts: designing the dictionary and coding the signal. This categorization made it available to use different algorithms in the content of sparse representation. In the first study of sparse representation, the existing transforms such as Fourier, STFT, Gabor, and wavelet transforms, were employed as a dictionary [45, 46].

The dictionaries described above model a signal by a simple mathematical function. In particular, the Fourier dictionary is created by a low pass filter and the wavelet dictionary is modeled by piecewise-smooth functions. The major advantage of these dictionaries is their simplicity, which does not require a multiplication by the dictionary matrix. However, the models presented are over-simplistic compared to the complexity of natural signals [25]. In 1996, Olshausen and Field [34] introduced a training dictionary algorithm detailed in subsection 2.2.2 for sparse representation model of small image patches selected from the natural images. This study produced superior results for the benefit of sparse representation and widely appealed in machine learning community. The proposed algorithm in [21] proved that the approach, which directly learns from the image itself, can reveal the elementary structures from complex signal data.

2.2.2 LEARNING A DICTIONARY IN SPARSE CODING

In learning a dictionary, the goal is to compose a set of basis function that represents each member of training samples. The dictionary is a matrix $\mathbf{D} = \{d^{(i)}\}_{i=1}^{K}$ that contains K basis functions on columns $\{d^{(i)}\}_{i=1}^{K}$ and a given training dataset, $\mathbf{X} = \{x^{(i)}\}_{i=1}^{M}$, can be represented as a sparse linear combination of these basis functions. The representation of x can be either exact $x = \mathbf{D}a$ or $x \approx \mathbf{D}a$ satisfying $\left\|\mathbf{D}a^{(i)} - x^{(i)}\right\|_p \leq \varepsilon$, where the vector $a^{(i)}$ is the reconstruction weight for $x^{(i)}$ using the basis function in $\mathbf{D}$ and $\|\cdot\|_p$ denotes the l_p norms for p = 1, 2, and ∞ [9]. In sparse coding, the dictionary is intentionally made overcomplete. Thus, the dictionary may capture a large number of patterns from the input data and also many of the reconstruction weights can be zeros leading to a sparse representation of the original data. In this section, we review three types of dictionary learning techniques.

Spherical K-means

Spherical K-means is a variety of K-means clustering algorithm and also named as "gain shape vector quantization" [47]. To learn the dictionary **D**, this algorithm minimizes the following equation:

$$\min_{\mathbf{D},a^{(i)}} \sum_i \left\|\mathbf{D}a^{(i)} - x^{(i)}\right\|_2^2 \tag{1}$$

$$\text{subject to } \left\|d^{(j)}\right\|_2^2 = 1, \forall j \quad \text{and} \quad \left\|a^{(i)}\right\|_0 \leq 1, \forall i$$

Here $\left\|a^{(i)}\right\|_0 \leq 1$ constrains each reconstruction weight to have at most a single non-zero component. The solution of the Equation (1) can be obtained using alternating iteration over the coefficients a and dictionary **D** while holding the other fixed. In short, this iteration determines

which element of $a^{(i)}$ should be non-zero to minimize the Equation (1) and assigns each $a^{(i)}$ to its optimal value. The optimal choice of each $d^{(j)}$ is computed holding the $a^{(i)}$ fixed.

Orthogonal Matching Pursuit

Equation (2) is the generalization of the spherical K-means algorithm that allows multiple non-zero components in $a^{(i)}$:

$$\min_{\mathbf{D},a^{(i)}} \sum_i \left\| \mathbf{D}a^{(i)} - x^{(i)} \right\|_2^2 \tag{2}$$

$$\text{subject to } \left\| d^{(j)} \right\|_2^2 = 1, \forall j \text{ and} \left\| a^{(i)} \right\|_0 \leq \mathrm{T}, \forall i$$

where $\left\| a^{(i)} \right\|_0$ is the number of non-zero components in $a^{(i)}$ and T determines the largest number of non-zero elements allowed to represent each $x^{(i)}$. This optimization is difficult to solve since the constraint is non-convex. However, the reconstruction weight $a^{(i)}$, with at most T non-zero elements, can be computed approximately using a greedy algorithm called Orthogonal Matching Pursuit (OMP) [48]. Given a single training sample $x^{(i)}$, OMP starts with $a^{(i)} = 0$ and at each iteration greedily selects a component of $a^{(i)}$ to be made non-zero to minimize the remaining reconstruction error. After each $a^{(i)}$ is updated to minimize Equation (2) over $a^{(i)}$ allowing only the selected components to be non-zero.

Sparse Coding

In the sparse coding framework [49], for a given training dataset, a dictionary can be learned by minimizing the following cost function:

$$\min_{\mathbf{D},a^{(i)}} \sum_i \left\| \mathbf{D}a^{(i)} - x^{(i)} \right\|_2^2 + \lambda \left\| a^{(i)} \right\|_1 \tag{3}$$

$$\text{subject to } \left\| d^{(j)} \right\|_2^2 = 1, \forall i$$

where $\|\cdot\|_1$ denotes the l_1 norm and λ is a trade-off parameter. The l_1 norm penalty produces the sparse resulting weights $a^{(i)}$ while λ controls the number of zeros. The solution of the Equation (3) can be obtained using alternating minimization over the sparse codes $a^{(i)}$ and the dictionary **D** while holding the other fixed [16]. A major drawback of this efficient learning system is its complexity and expense.

Dictionary learning plays an important role in the sparse coding framework, but it is difficult and time-consuming. Recent research showed that randomly selected dictionaries can also perform well. For example, Jarrett et al. found that features from a one-layer convolution pooling architecture with a random dictionary could achieve a sufficient recognition rate for image classification [50]. Coates et al. have experimentally proved that the choice of basis functions does not affect much on classification performances as long as it is overcomplete [16]. In this dissertation, we randomly select the basis functions from the constructed HSI sub-bands as described in Chapter 3.

2.2.3 ENCODING

Once the dictionary is obtained by solving the optimization problems described above, an encoding step is performed to transform the input test data $x^{(i)}$ into the new representation vector. For a data sample $x^{(i)}$, the feature representation $a^{(i)}$ is usually obtained by solving the matching algorithm employed in dictionary learning with **D** fixed. However, it is not essential to use the matching training algorithm; instead, another encoding algorithm can be employed to encode the data samples. In Ref. 16, the learning and encoding algorithms were compared in a crosswise style to analyze the contributions of each of them.

Another existing possibility to compute the new representation is the soft thresholding method (Figure 2) [16, 51, 52] such as

$$a^{(i)} = sign(z^{(i)}) \max(0, \left|z^{(i)}\right| - t) \tag{4}$$

where t is an adjustable threshold and $z^{(i)} = D^T \cdot x^{(i)}$. The soft thresholding function, also called as shrinkage function, reduces the absolute value of $z^{(i)}$ by a certain amount depends on the parameter t. Thus, the soft thresholding function produces new sparse representations by transferring relatively small values of $z^{(i)}$ into zero.

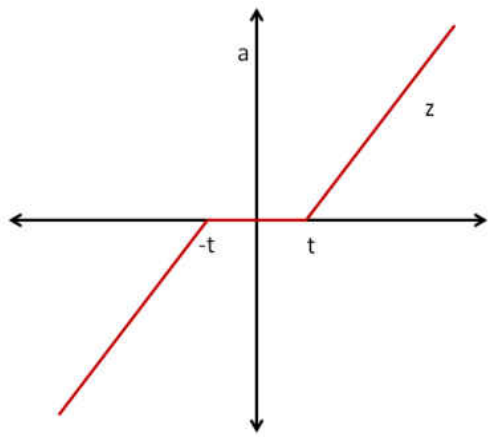


Figure 2: Representation of soft thresholding.

Coates et al. showed that an efficient and effective image classification system can be implemented by combining the random basis function strategy and the soft thresholding method [16].

2.2.4 WHY DOES SPARSE CODING WORK

Sparse coding is an algorithm to find a sparse representation of input data. For a given input data set, it learns a dictionary that contains a set of representative low level features (building blocks also named as dictionary) in the data. Using the basis functions in the dictionary, a new data point can be reconstructed as a linear combination of a few basis functions. In this part of

the dissertation, we list the reasons why sparse coding is appropriate for representation of images. First, in sparse coding, it is assumed that natural images usually consists of low level features such as edges and lines [53]. Thus, in the application of sparse coding to natural images, learned basis functions are formed similarly to the receptive fields of neurons in the visual cortex system [53]. Second, the techniques prior to sparse coding such as PCA and wavelet transform utilize a set of orthogonal and complete basis functions. Contrarily, sparse coding uses an overcomplete dictionary, in which the number of basis functions is larger than the number of features of the data. The use of overcomplete dictionary allows more flexibility in adaptation of learned features to data. Third, the learned dictionary has similar properties as the human visual cortex such that the basis functions are localized, oriented, and band-passed. This property makes sparse coding a conceivable model for the human visual cortex system [54].

2.3 CLASSIFICATION TECHNIQUES

Classification is one of the most basic and important research contents of the HSI and multispectral image processing. It can be described as an analytical technique of identifying the land object into a class. The main task in classification is to assign each pixel of an image into a class to generate the thematic map using the features of the interested pixel. There are several typical HSI and multispectral image classification methods, including *k*-nearest neighbor (*k*NN), spectral angle match (SAM) [55], maximal likelihood (ML) [56], SVM [4-6], and composite-kernel SVM (CKSVM) [57]. In the following subsections, we review *k*NN, SVM, and CKSVM, classification methods in the content of this dissertation.

2.3.1 *k*NN CLASSIFIER

The *k*NN classifier [58], one of the simplest and most widely used classification method, is based on the distance/similarity of training (known) and test (unknown) samples. A nearest

neighbor classifier locates the nearest neighbor in the training set and labels the test sample with the same class label as that of the training neighbor. More robust model of the nearest neighbor classifier is achieved by assigning *k* neighbors, where *k* >1, and letting the majority vote decide the outcome of the class labeling. The *k*NN classifier finds the *k* nearest neighbors of a test sample in the training samples and classifies the test sample into the most common class of the *k* nearest neighbors. In distance-weighted *k*nn (W*k*NN) [59], the closest neighbors in training samples have more weighted than the further ones, thus, the classification result of the test sample is made by the majority weighted voting. A number of algorithms are proposed to determine the weights of the neighbors, including locally linear embedding (LLE) [60], which is proposed for dimension reduction. LLE-based *k*NN first computes the distance from a test sample to every other training sample to find a set of the nearest neighbors and then computes a set of weights for each test sample that best describes the sample as a linear combination of its neighbors. Lastly, an eigenvector-based optimization technique is applied to determine the low-dimensional embedding of the sample best reconstructed by the weights.

2.3.2 SVM CLASSIFIER

SVM [61] is a supervised classification method that finds a function that assigns each test sample to its corresponding class for a given set of training samples and their corresponding classes. Linear SVM constructs hyperplanes in a higher dimension space for the purpose of classification. Obviously, a successful classification is achieved by the hyperplane that maximizes margin, which is the distance between the nearest training samples and the separating hyperplanes (Figure 3). The ideal SVM should construct a hyperplane that classifies the samples into two non-overlapping classes. However, this ideal case may not be possible in most cases. Thus, SVM searches for a hyperplane that maximizes the margin and simultaneously minimizes

the error of classifier. The strength of SVM is that if the data is linearly separable, there is only one global minimum value.

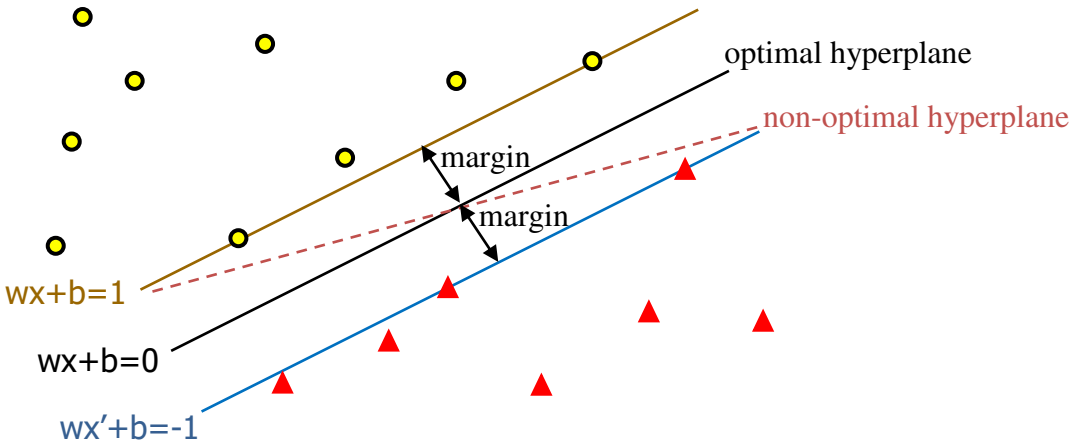


Figure 3: Representation of hyperplanes.

Mathematically, given a training set of sample-class pairs $(x_i, y_i), i = 1,...,l$, where $x_i \in R^n$ and $y \in \{1,-1\}^l$, SVM solves the following optimization problem:

$$\min_{\theta,b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i^l \xi_i \tag{5}$$

$$\text{subject to } y_i\left(\mathbf{w}^T\phi(x_i)+b\right) \geq 1-\xi_i \text{ and } \xi_i \geq 0$$

where $\mathbf{w}$ is a the normal to the optimal decision hyperplane, b represents the closest distance to the origin of the coordinate system, ξ is a slack variable for each x_i, C is an adjustable parameter controlling the trade-off between the slack variable and the margin and the function and ϕ maps the training samples x_i into a higher dimensional space. Here it should be noted that maximizing the distance of samples to the optimal decision hyperplane (Figure 3) is equivalent to minimizing the norm of $\mathbf{w}$. The dimensionality of the vector variable w is usually very high,

thus, Equation (5) is usually solved through its Lagrangian dual problem, which consists of maximizing

$$L_d \equiv \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \left(\phi(x_i).\phi(x_j)\right) \tag{6}$$

subject to $0 \le \alpha_i \le C$ and $\sum_i \alpha_i y_i = 0$, $\forall i = 1,...,l$

where α_i are Lagrange multipliers. As noticed in Equation (6), the mappings occur in the form of inner products, thus, a kernel function can be defined where $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$. In linear SVM, $K(x_i, x_j) = x_i^T x_j$ is used as a kernel function.

However, the linear separation may not be a good solution for some cases, thus, more flexible decision boundary, i.e. non-linear decision boundary, is needed. In non-linear SVM, first, the data in the original finite-dimensional space are mapped into a much higher dimensional space using a non-linear kernel function, such as polynomial, gaussian, and exponential radial basis function (RBF) kernels in Equations (7), (8), and (9), respectively. Then, the linear separation is performed in that space efficiently.

$$K(x_i, x_j) = \left(x_i^T x_j + 1\right)^d \tag{7}$$

$$K(x_i, x_j) = \exp\left(\frac{-\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{8}$$

$$K(x_i, x_j) = \exp\left(\frac{-\left\|x_i - x_j\right\|}{2\sigma^2}\right) \tag{9}$$

where d is the polynomial kernel degree, σ is the kernel parameter. In recent years, SVM has been successfully employed in HSI classification due to its excellent performance in high-dimensional data [12, 13].

2.3.3 COMPOSITE KERNEL SVM (CKSVM) CLASSIFIER

SVM takes into account only the spectral information of data. Motivated by the fact that pixels in an HSI are often surrounded by pixels of the same class, CKSVM method [57] is proposed to exploit the spatial information in HSI through composite kernels. CKSVM considers spectral, spatial and local cross information in HSI constructing kernels composed from the spectral and contextual kernels. In CKSVM, first the mean and standard deviation of HSI pixels are extracted to define the spatial information of HSI. Then, kernel matrices are computed from the spectral and spatial blocks of HSI (K_{ws} and K_{sw}) using any suitable Kernel function, such as in Equations (7) - (9). SVM classification algorithm is applied to the combination of spectral dedicated kernel matrix (K_w), contextual dedicated kernel matrix (K_s) and cross information of spectral and spatial features (K_{ws} and K_{sw}). CKSVM proposes different kernel combination approaches to extract the final representation of a HSI pixel as listed below:

(1) Direct summation kernel:

$$K(x_i, x_j) = K_s(x_i^s, x_j^s) + K_w(x_i^w, x_j^w)$$

(2) Weighted summation kernel:

$$K(x_i, x_j) = \mu K_s(x_i^s, x_j^s) + (1-\mu) K_w(x_i^w, x_j^w)$$

(3) Cross-information kernel:

$$K(x_i, x_j) = K_s(x_i^s, x_j^s) + K_w(x_i^w, x_j^w) + K_{sw}(x_i^s, x_j^s) + K_{ws}(x_i^w, x_j^w)$$

where x_i^w and x_i^s are the spectral and spatial features of HSI, respectively, and μ is the parameter $(0 < \mu < 1)$ to tune the spectral and spatial information.

In this dissertation, the new representations of HSI pixels are classified with SVM and CKSVM classifiers as described in Chapter 3.

2.4 FEATURE SELECTION

The dimensionality of the new features obtained by sparse coding is usually very high that may lead to the overfitting problem and require many computational resources in terms of time and memory [62]. A feature extraction or feature selection algorithm can be used to reduce the dimensionality of the features. Feature extraction algorithms reduce the dimension by creating new combinations as features, whereas feature selection algorithms select the most relevant features without incurring much loss of information. We propose to use a feature selection algorithm due to its efficiency in large dimension. The benefits of the feature selection are: reducing the measurement and storage requirements, reducing training, testing and classification time, resisting overfitting to improve the classification performance and interpreting data easily. In this dissertation, to select effective features for HSI classification, we propose to utilize a multi-class logistic regression [63-66] as shown in Equation (10)

$$\min_{\theta} \sum_{l=1}^{p} \sum_{i=1}^{m} w_{il} \log(1+\exp(-y_{il}(\theta_l^T \alpha_{il} + c_l))) + r(\theta) \tag{10}$$

where α_{il} is the *i*'th feature for the *l*'th class, w_{il} is the weight for α_{il}, y_{il} is the response of α_{il}, c_l is the intercept (scalar) for the *l*'th class, θ is the obtained weight, *m* is the total number of the training samples and *p* is the total number of class. In Equation (10), the first term is the logistic loss function, also known as logistic regression, and the second term is a penalty term, also known as the regularization term, which is used to address the overfitting problem in the logistic regression. The following l_1 and l_2 regularization terms are commonly used.

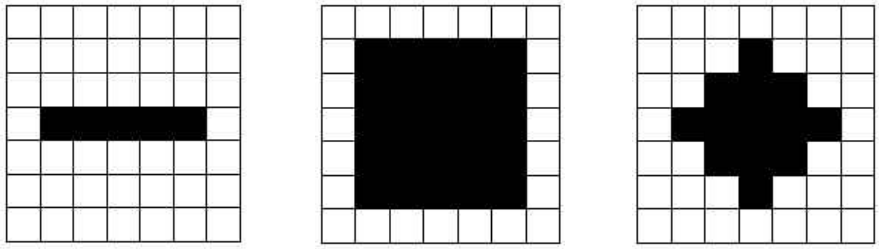
$$r_{l2}(\theta) = \beta \frac{1}{2} \|\theta\|_2^2 \tag{11}$$

$$r_{l1}(\theta) = \beta \|\theta\|_1 \tag{12}$$

where β is a regularization parameter, which controls the trade-off between the logistic loss function and the size of the weights, $\|\cdot\|_1$ denotes the l_1 and $\|\cdot\|_2$ denotes the l_2 norms, which enforce the resulting weights to be sparse. Both Equations (11) and (12) are convex and separable functions. Directly solving the logistic model (11) is ill-posed and may lead to overfitting when a high number of training samples ($m >> p$) are used. Therefore, in the recent years, the l_1 norm regularization approach, also known as LASSO [67], has been applied to address the overfitting problem. However, solving the l_1-regularized logistic regression is more challenging than solving the l_2-regularized logistic regression, since the regularization term (12) is non-differentiable [66]. In addition, the l_1-regularized logistic regression may produce too sparse weights resulting in a highly unstable feature selection. Thus, many algorithms have been interested in extending the l_1 regularization to l_1/l_q regularization due to its success to overcome the problems described above [68-71]. l_1/l_q regularization also facilitates the group effect of features with high correlation. Assuming HSI pixels are often surrounded by pixels of the same class, in this dissertation, we propose to use l_1/l_q regularization in Equation (10).

2.5 MORPHOLOGICAL OPERATIONS

A morphological operation is a tool for extracting or modifying the structure of objects in an image based on the characteristics of a structuring element (SE). An SE is a binary matrix with arbitrary shape, size, and orientation. Figure 4 shows linear, square and discrete disk types of structural elements. The common usages of morphological filters are edge detection, feature detection, counting objects, segmentation and noise reduction. The morphological filters can be applied to both gray-level and binary images. In this section, four fundamental morphological filters; erosion, dilation, opening, and closing are explained.



(a) (b) (c)

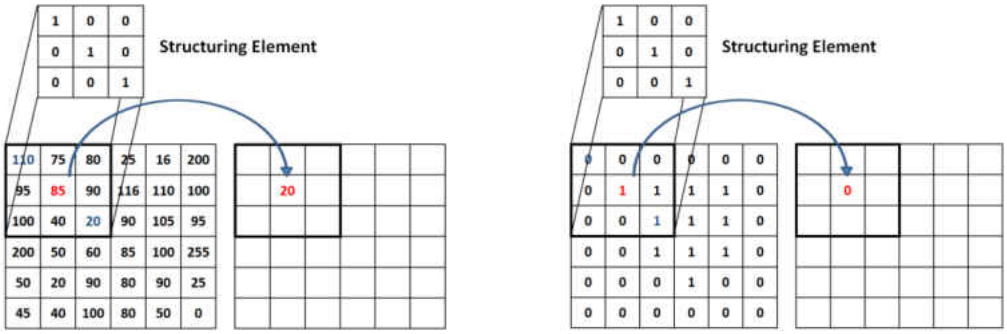
Figure 4: Examples of SE. (a) linear, (b) square, and (c) discrete disk.

2.6.1 EROSION

An erosion filter decreases the pixel values on object boundaries in an image. This process can be formulated as Equation (13):

$$(x \ominus SE)(i, j) = \min_{(m,n,)\in SE} \{x(i+m, j+n)\} \tag{13}$$

where $\ominus$ represents erosion process. In erosion, the new value of a pixel is the minimum pixel value within the neighborhood of the pixel of interest where the neighborhood is chosen based on SE. In the binary case, therefore, if there is a zero within the neighborhood then the value of the pixel will be zero after erosion operation. Figures 5(a) and 5(b) show erosion operations on gray-level and binary images, respectively, where a linear SE is used.



(a) (b)

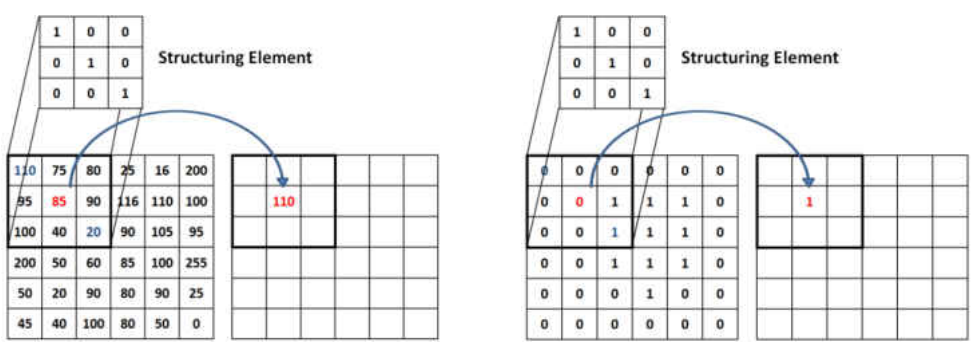
Figure 5: Examples of erosion operations on (a) grey-level, (b) binary image.

2.6.2 DILATION

A dilation filter increases the pixel values on the boundaries of objects in an image. This process can be formulated as Equation (14):

$$(x \oplus SE)(i, j) = \max_{(m,n,)\in SE} \{x(i-m, j-n)\} \tag{14}$$

where ⊕ represents dilation process. In dilation, the new value of a pixel is the maximum pixel value within the neighborhood of the pixel of interest in which the neighborhood is chosen according to the shape and size of the SE. In the binary case, therefore, if there is one within the neighborhood then the value of the pixel will be one after dilation operation. Figures 6(a) and 6(b) show dilation operations on gray-level and binary images, respectively, where a linear SE is used.



(a) (b)

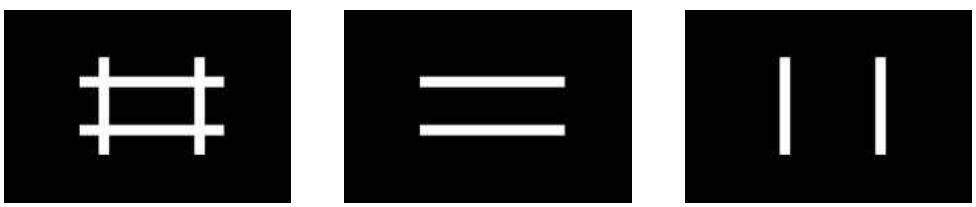
Figure 6: Examples of dilation operations on (a) grey-level, (b) binary image.

2.6.3 OPENING

An opening operation is the concatenating the two fundamental morphological filters, an erosion filter followed by a dilation filter with the same SE. Mathematically, it is formulated as:

$$(x \circ SE)\ (i, j) = ((x \ominus SE) \oplus SE)\ (i, j) \tag{15}$$

where "o" represents the opening operation. In the binary case, erosion typically removes objects smaller than the SE and the subsequent dilation tends to restore the shape of the objects that remain. Figure 7 shows a binary input image and corresponding output images by the opening operation with a horizontal (Figure 7(b)) and a vertical (Figure 7(c)) linear SEs. From Figure 7(b), it can be seen that the horizontal SE successfully restores the horizontal objects; however, the vertical objects are removed from the image since the erosion step completely eliminates them. Conversely, the vertical SE restores only the vertical objects as seen in Figure 7(c).



(a) (b) (c)

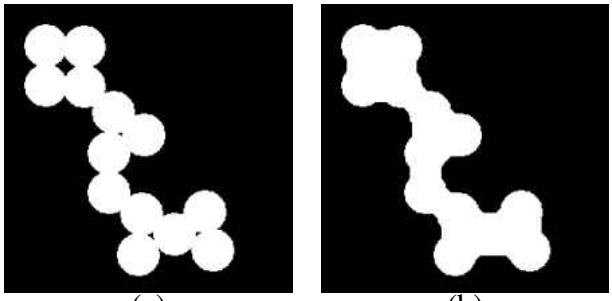
Figure 7: Implementation of opening operation. (a) Original image, (b) image opened by a horizontal linear SE, and (c) image opened by a vertical linear SE.

2.6.4 CLOSING

As opening operation, the closing operation is also the concatenating the two fundamental morphological filters in which a dilation filter followed by an erosion filter with the same SE. Mathematically, it is formulated as:

$$(x \bullet SE)\,(i, j) = ((x \oplus SE) \ominus SE)\,(i, j) \tag{16}$$

where "•" represents the closing operation. In the binary case, closing smoothes the boundaries of foreground objects but, in contradistinction to the opening, it merges narrow gaps eliminating small black holes as seen from Figure 8.



(a) (b)

Figure 8: Implementation of closing operation. (a) Original image, (b) image closed with a disk SE.

As seen from the explanations above, the opening and closing filters tend to restore the shape of the objects that remain. However, the accuracy of this restoration depends on the similarity between the shapes of the objects and the SE.

In this dissertation, we used another morphological filter, named "opening by reconstruction" to eliminate the false positives obtained in the classification step of the scar detection algorithm. This process is explained in Chapter 4.

Chapter 3

PROPOSED METHOD

In this chapter, a new feature representation method based on sparse coding for the classification of HSI is presented. First, we introduce the model of hyperspectral imagery and existing classification techniques for HSI. Next, a new representation model is proposed using spectral information of the HSI. In spectral-based sparse coding model, we extract overlapping sub-bands from the spectral representation of a pixel then construct a dictionary based upon these sub-bands. Subsequently, we encode each HSI pixel to a new representation for classification. In the experiment part, we construct another dictionary by solving Equation (3) to compare the effectiveness of random dictionary. We experimentally prove that a randomly selected dictionary could be as effective as a dictionary learned from optimization. However, the dimensionality of the new representation for HSI is usually very high requiring a lot of computational resources, which significantly limit the applicability of sparse coding in HSI classification. In addition, spatial information is not considered. Thus, the proposed algorithm is enhanced by including the spatial information of the HSI data to improve classification and utilizing the multinomial logistic regression technique [63-66] to select the most effective features in the new representation. We call the enhanced model as sparse coding based dense feature representation (SC-DFR) method. In this chapter, we also evaluate the SC-DFR method on a variety of benchmark data and compare it with multiple state-of-the-art HSI classifiers.

3.1 HYPERSPECTRAL IMAGERY INTRODUCTION

In remote sensing, the reflectance of a target can be characterized by spectral responses over a variety of wavelengths. The comparison of these responses over distinct wavelengths can be used to distinguish different materials. Hyperspectral imagery is an imaging modality that represents ground reflectance across hundreds of very narrow spectral bands throughout the visible, near-infrared, and mid-infrared portions of the electromagnetic spectrum. This yields fine discrimination between different ground targets based on their spectral response in each of the narrow bands, making it a useful modality for fine differentiation between ground objects. It also has the potential for accurate determination of terrain properties such as, material classification, geological feature identification and environmental monitoring [72]. A distinctive structure of HSI and a real HSI cube from reflective optics system imaging spectrometer (ROSIS) obtained over the center of Pavia are shown in Figure 9 and 10, respectively. Each pixel in hyperspectral image (HSI) data consists of hundreds of spectral bands captured in different spectral channels, where the pixel $x_i \in R^B$ is represented by a B-dimensional feature vector and B is the number of spectral bands. This feature vector is called the spectrum of the material in the pixel. Different materials can be identified by classifying HSI pixels into different categories based on the statistical characteristics of the pixel values. A variety of techniques have been used for the classification including SVM [4-6], neural networks [7], and graph-based methods [8].

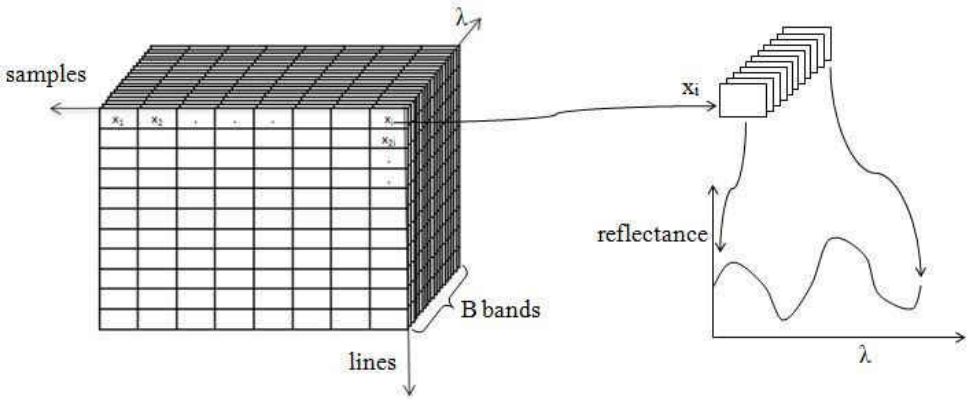


Figure 9: A distinctive structure of HSI.

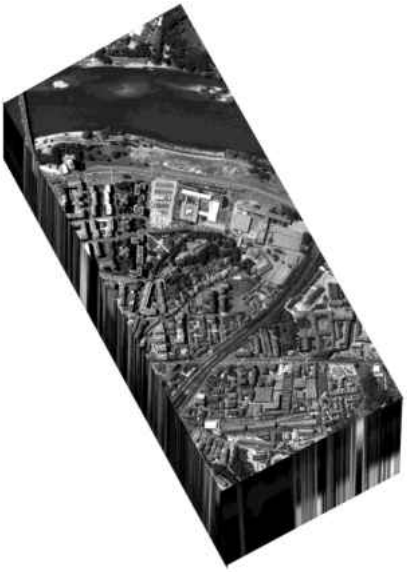


Figure 10: A real HSI cube obtained over the center of Pavia.

Though the abundant information in each pixel increases the capability of distinguishing different materials, it is often difficult to exploit HSI data due to the particular challenges in the remote sensing environment such as noise of measurement devices, energy interaction between the targeted area and the spectrometer, and spectral mixing, where each pixel is composed of a

combination of different materials. All those challenges will jeopardize a precise material identification in HSI. In addition, the non-negligible changes of spectral responses within the same land-cover class [73] and the limited availability of the ground truth data (labeled pixel) make the spatial structures of HSI as critical as spectral structures in data analysis. Therefore, many recent techniques have integrated the spatial and spectral information for HSI classification such as, the stacking vector approach [57, 74], segmentation-based techniques [75, 76], the Markov random field (MRF) method [77-79], context-based techniques [80], and mathematical morphology [81-82]. In the stacking vector approaches [57, 74], the spatial features of a pixel are built by calculating the average or standard deviation of its neighbors and then concatenated with spectral features for classification. In segmentation techniques [75, 76], pixel classification results are refined by a postoperation step such as segmentation or majority voting. MRF methods combine the spectral and spatial information using the Bayesian framework [77]. Context-based techniques use the random set framework [80] or the hierarchical segmentation technique [81] to characterize the complex and irregular patterns in the pixel feature space. Mathematical morphology based methods [82-86] usually utilize the morphological approaches to integrate spatial information in the feature extraction process of the HSI classification task.

It is often assumed that the measured energy in HSI data is proportional to the covered ground and the reflectivity of its environment as described by the following linear mixture model [11],

$$\mathbf{X} = \mathbf{DA} + \mathbf{N} \tag{17}$$

where $\mathbf{X}$ is the HSI image consists of M pixels with B bands, $\mathbf{X} \in R^{B\times M}$, $\mathbf{D}$ is the dictionary consists of K basis functions (atoms of materials) with a length of B, $\mathbf{D} \in R^{B\times K}$, $\mathbf{A}$ is the coefficient matrix representing the mixture of dictionary functions, $\mathrm{A} \in R^{K\times M}$ and $\mathbf{N}$ is additive

noise. Due to the high spatial resolution of latest imagery technology, i.e., only a few land cover materials may be present in one pixel and thus, each pixel can be approximated by a few basis functions, leading to a sparse coefficient matrix **A**. Therefore, the HSI model in Equation (17) is consistent with the sparse signal model in Equation (3).

Sparse coding has been successfully applied to HSI classification with encouraging results [10-13,18,19,87,88] due to its success in discriminative tasks with small training data. For example, [10] uses sparse coding minimization formula in Equation (3) to learn spectral signatures of material in the dictionary training phase. The extracted features are then classified through a linear SVM classifier. A spatial-spectral based HSI classification study [11] learns class-specific spectral dictionaries for each class of data using a sparse coding based projected gradient iteration. Then, each pixel is assigned to the class whose dictionary have the lowest representation error. Another sparse coding based algorithm [12] employs two different methods to integrate the spectral and spatial contents of HSI data: laplacian constraint and joint sparsity model. The dictionary is learned by a generalized form of OMP algorithm, called simultaneous OMP [89], as described in Section 2.2.1. The reconstructed pixel is then labeled as the class whose training samples have the largest similarity to that pixel. In [13], a structured dictionary based feature representation is proposed incorporating the both spectral and spatial information. The pixels of HSI are grouped into a number of spatial neighborhoods, called as contextual groups, and reconstructed with a linear combination of a few dictionary elements learned from the data. The convex sparse coding minimization technique in Equation (3) is also used in [13]. The estimated sparse coefficients are considered new representations to be classified using a linear SVM.

These algorithms are useful in obtaining the features of HSI for the classification, however, HSI classification systems based on the sparse signal model in Equation (3) are computationally expensive, requiring the solution of a convex optimization problem. In this dissertation's proposed method, sparse representation is achieved by a simple, but efficient approach without using convex optimization solution as described in the following subsections.

3.2 SPECTRAL-BASED SPARSE CODING MODEL

The proposed spectral-based feature representation method for HSI data classification consists of three steps as illustrated in Figure 11 [90]: (1) sub-band construction from HSI data, (2) dictionary learning, and (3) encoding. Each step is detailed in the following subsections.

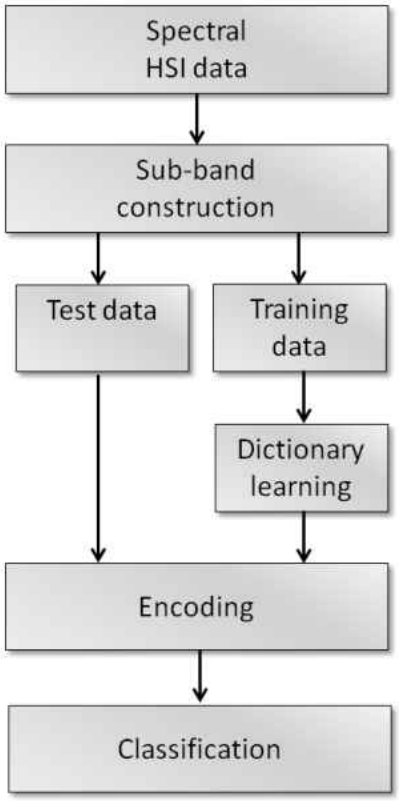


Figure 11: Flowchart of the spectral-based feature representation method.

3.2.1 SUB-BAND CONSTRUCTION FROM HSI DATA

To divide the pixel into sub-bands, a window with a size of b bands was used. The window was moved along the spectral direction as shown in Figure 12 with a step size of 1, resulting in B-b+1 sub-bands for the total B bands of a pixel, where b is called the receptive field length. Note that these sub-bands were overlapped and $x^{(i)} \epsilon R^b$ is the vector representing the i'th sub-band of a HSI pixel. Then the sub-bands were preprocessed as follows. First, each sub-band $x^{(i)}$ was normalized as in Equation (18).

$$\tilde{x}^{(i)} = \frac{x^{(i)} - mean\left(x^{(i)}\right)}{\sqrt{\mathrm{var}\left(x^{(i)}\right) + 10}} \tag{18}$$

where $\tilde{x}^{(i)}$ is the normalized sub-band.

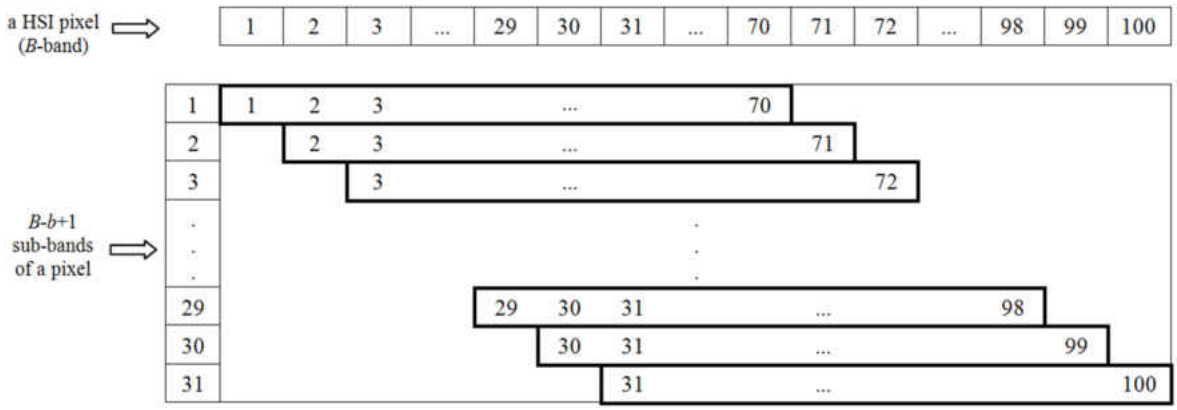


Figure 12: Extraction of sub-bands.
(From a 100-band (B) HSI pixel, 31 sub-bands are extracted with 70 receptive field length (b).)

To provide uncorrelated bands and to distribute more basis functions in the orthogonal directions, the zero-phase components analysis (ZCA) whitening process [91] was applied to sub-bands. In Ref. 92, this process was identified as a critical step for the quality of the learned feature representations. Moreover, according to Ref. 93, the whitened image patches can result in

sharper edge features similar to those created by sparse coding and independent component analysis (ICA).

3.2.2 DICTIONARY LEARNING

The main purpose of dictionary learning is to find a set of basis functions $\{d^{(j)}\}_{i=1}^{K}$ that forms the building blocks of **X**. Compared to the dictionary learning methods mentioned in Section 2.2.1, we employed a very simple method as in Ref. 16. The method for dictionary learning is relatively fast and simple, where a portion of preprocessed sub-bands were randomly selected from the sub-band set to compose the dictionary **D** (Figure 13). If k is the total number of sampled sub-bands for dictionary learning, we denote it as $\mathbf{D}=\{d^{(1)}, d^{(2)}, \ldots, d^{(k)}\}$, where $d^{(j)} \in R^{b}$.

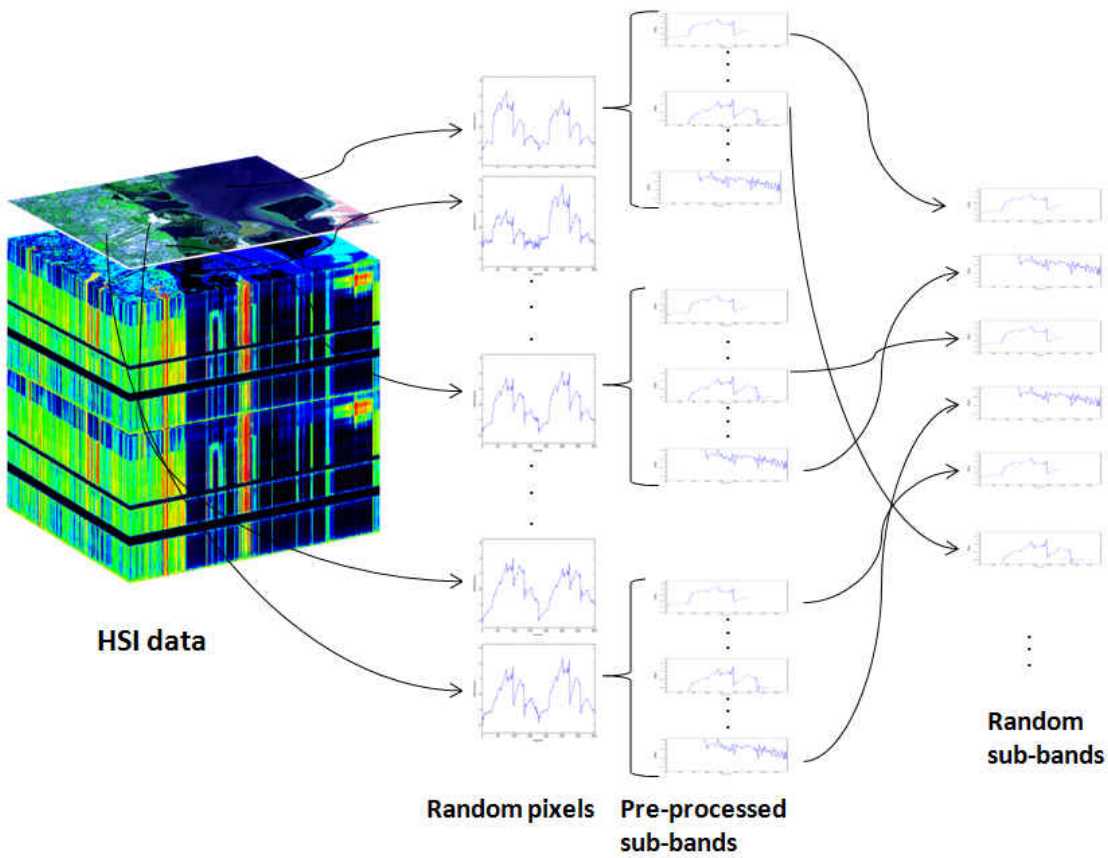


Figure 13: Dictionary learning from random sub-bands.

3.2.3 ENCODING

With the dictionary **D**, we obtained a new feature representation for an HSI pixel in a convolutional manner as follows (Figure 14).

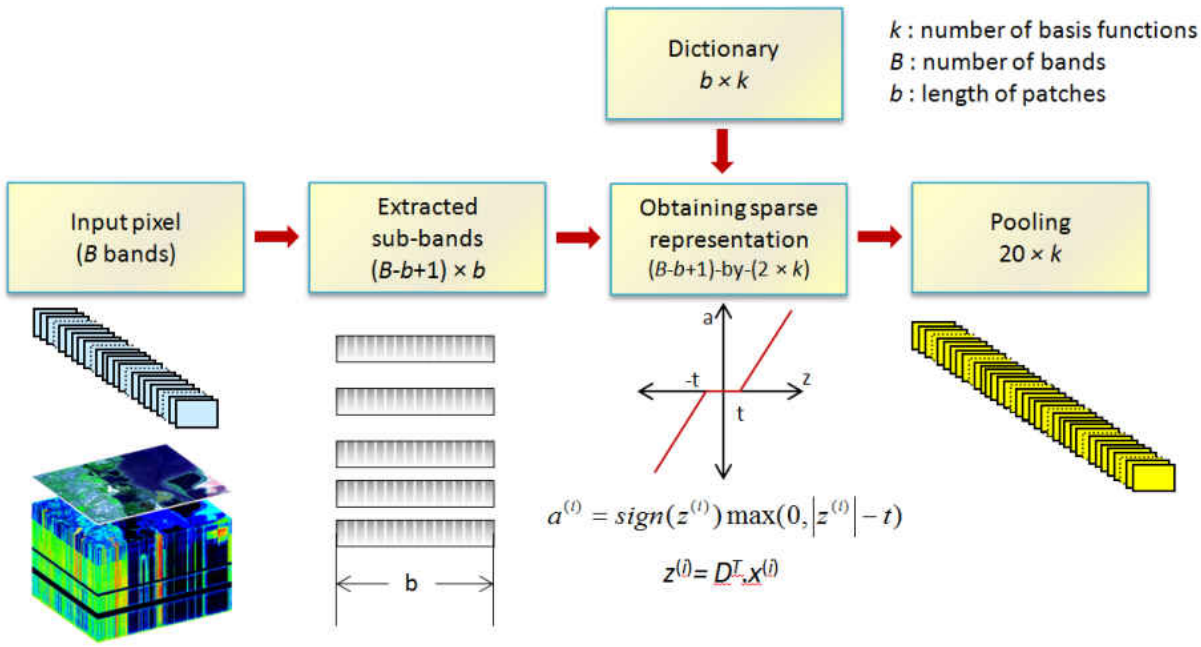


Figure 14: Encoding step.

Obtaining sparse representation

For each normalized sub-band, we obtained a new representation by applying the soft threshold function in Equation (4). Then we split the positive and negative components of the new representation as two separate features. Note that the dimension of the representation for the sub-band is 2 × *k*. We applied the encoding step to each overlapping sub-band of the pixel; thus, we had (*B*-*b*+1)-by-(2 × *k*) matrix for each pixel to be converted into a single feature representation in the pooling step.

$$a^{(i)} = sign(z^{(i)})\max(0, |z^{(i)}| - t) \tag{4}$$

Pooling

In this step, we split all representations from the *B*-*b*+1 sub-bands of the pixel into 10 equal-sized groups and summed them in each group to obtain 10 representations (Figure 15). The final representation, α_w with a dimension of 20 × *k* for the HSI pixel was achieved by concatenating the 10 representations. This step also aimed to reduce the dimensionality of the new feature representation and has been widely studied in different applications [94]. It was also noticed that the feature representation may not be sparse after this step.

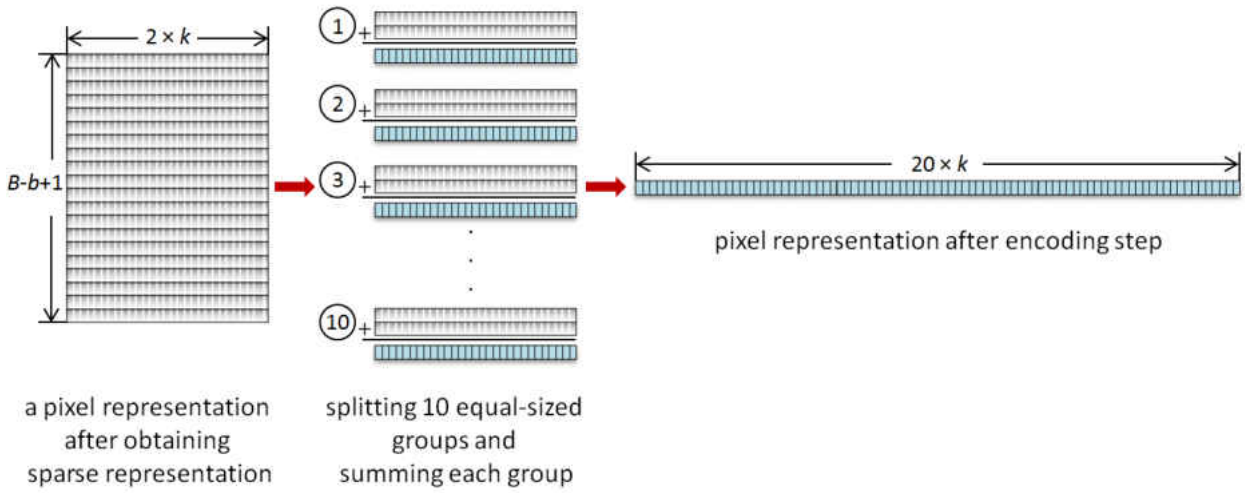


Figure 15: Pooling step.

Once we obtained the final representation vector for each pixel, we used the linear SVM classifier with the kernel function $K(x_i, x_j) = x_i^T x_j$ as described in Section 2.3.2 to evaluate the efficiency of the spectral-based sparse coding feature representation model. In the experiments, we constructed another dictionary using sparse coding optimization technique in Equation (3) as described in Section 2.2.1 for a comparison. The experimental results of the spectral-based proposed feature representation method will be reported in Section 3.4.

3.3 SPARSE CODING BASED DENSE FEATURE REPRESENTATION MODEL

The proposed feature representation method described in Section 3.2 was succesfully extracted new features of HSI data and increased the classification accuracy using spectral information of the pixels [95]. However, the spatial information of the pixels was not considered in the method. In addition, the dimensionality of the feature representation was very high. In this section, the general method discussed in Section 3.2 is modified to incorporate the spatial information and to reduce the dimension of the new representation.

Recall from Section 3.2 the spectral based sparse coding model consists of three steps. The modified method, SC-DFR, has 4 steps as illustrated in Figure 16 [95, 96]: (1) sub-band construction from HSI data, (2) dictionary learning, (3) encoding and (4) feature selection. The modification parts of the proposed method are detailed in the following subsections.

3.3.1 SPATIAL INFORMATION

Material information at one pixel may spread over its neighbors due to the effect of the point spread function (PSF) of imaging devices. Therefore, it is beneficial for characterizing a pixel if its spectral and spatial information are simultaneously considered. In the proposed method, we computed the average value of the neighbors of each pixel in each spectral band, resulting in another B value for the pixel. It should be noted that the steps described in Section 3.2 were applied to the spectral and spatial contents to form the spectral information (α_w) and spatial information (α_s) separately. Then we balanced the spatial and spectral contents by using Equation (19) as

$$\alpha = \mu\alpha_w + (1-\mu)\alpha_s \tag{19}$$

where μ is a trade-off $(0 \le \mu \le 1)$ between α_w and α_s, α is the balanced representation of a pixel.

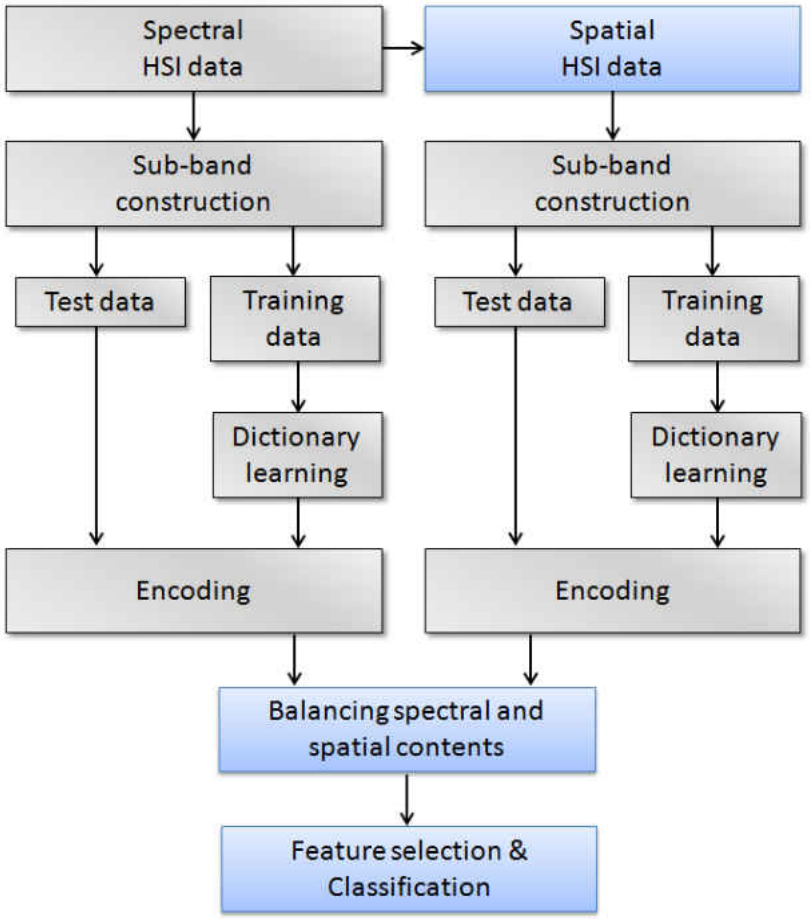


Figure 16: Flowchart of the SC-DFR method.

3.3.2 FEATURE SELECTION

One challenge of the proposed method described in Section 3.2 was that the dimensionality of the feature representation was very high. In Section 3.2, we implemented the pooling method to reduce the size of the new feature representation. However, after this step, the dimensionality of the features is still on the order of thousands, which may be undesirable for real-time systems.

For the specific classification task, some of the features may be correlated or redundant, leading to the overfitting problem [62]. We propose to utilize the l_1/l_q regularized multiclass logistic regression [63-66] $(q \geq 1)$ as shown in Equation (20) to select effective features for HSI classification,

$$\min_{\theta} \sum_{l=1}^{p} \sum_{i=1}^{m} w_{il} \log(1 + \exp(-y_{il}(\theta_l^T \alpha_{il} + c_l))) + \beta \|\theta\|_{l_1/l_q} \tag{20}$$

where α_{il} is the *i*'th feature for the *l*'th class, w_{il} is the weight for α_{il}, y_{il} is the response of α_{il}, c_l is the intercept (scalar) for the *l*'th class, θ is the obtained weight, β is a regularization parameter, *m* is the total number of the training samples and *p* is the total number of class. In this study, we set $l_q = l_2$ and $w_{il} = 1/(mp)$, to select the most effective features. It should be noted that the feature selection step was performed on training data only. The l_1/l_q regularized regression optimization, as described in Section 2.4, is a recently developed method that favors the group sparsity in the model [63-66].

Once we obtained the final representation vector for each pixel, we used two different classifiers, SVM and CKSVM described in Sections 2.3.2 and 2.3.3, to evaluate the efficiency of the SC-DFR method for HSI classification. In SVM, we used the linear kernel function $K(x_i, x_j) = x_i^T x_j$. In CKSVM, first, the composite kernels of the spectral and spatial features were calculated via the polynomial kernel in Equation (7) and gaussian kernel in Equation (8), respectively, then the weighted summation equation described in Section 2.3.3 was used to combine them.

3.4 SPECTRAL-BASED EXPERIMENTS

In this section, we reported the experiments of the spectral-based sparse coding framework proposed in Section 3.2 on a HSI data collected at the Kennedy Space Center (KSC) for land

cover classification [90]. In the experiments, we compared two dictionary learning methods: random selection and sparse coding solved with the coordinate descent algorithm [97]. Results from the proposed framework were also compared to those obtained by a recently proposed method; supervised locally linear embedding weighted *k*-nearest-neighbor (SLLE-W*k*NN) [98] classifier. The experiments of SC-DFR method will be reported in Section 3.5.

3.4.1 DATA DESCRIPTION

A hyperspectral data set collected by NASA airborne visible/infrared imaging spectrometer (AVIRIS) over Kennedy Space Center [99] in March 1996 was used in this experiment. One single band image from this data set is shown in Figure 17. This sensor can acquire 224 bands with an 18-m spatial resolution and a 10 nm spectral resolution over the range of 400-2500 nm. There are 5,211 out of 314,368 (512 × 614) pixels labeled. Details of the land cover classes are given in Table 1. After removing the water absorption and noisy bands, there are 176 bands left for this study.

Figure 17: A sample band of Kennedy Space Center HSI data.

Table 1: Class names and number of labeled KSC data.

Class Number	Class	Number/Percentage of Labeled Data
1	Scrub	761 (14.6%)
2	Willow swamp	243 (4.66%)
3	Cabbage palm hammock	256 (4.92%)
4	Cabbage/oak hammock	252 (4.84%)
5	Slash pine	161 (3.07%)
6	Oak/broadleaf hammock	229 (4.38%)
7	Hardwood swamp	105 (2.0%)
8	Graminoid marsh	431 (8.27%)
9	Spartina marsh	520 (9.9%)
10	Cattail marsh	404 (7.76%)
11	Salt marsh	419 (8.04%)
12	Mud flats	503 (9.66%)
13	Water	927 (17.8%)
	TOTAL	5,211

3.4.2 EXPERIMENTAL SETUP

For a fair comparison, we utilized the same configuration as that in [98] for our experiments. In detail, the parameters involved in our algorithm were tuned based on three-fold cross-validation using 50% of the labeled data. There were two parameters in the proposed method: (1) receptive field length, b, and (2) threshold, t, for encoding. After parameter selection, 30% and 40% of the data are randomly selected for testing and training, respectively. The testing dataset is disjoint from the data used for cross-validation. In our experiments, dictionary was learned or randomly selected from the training set.

In the followings, we first compare the two dictionary learning methods and then the final classification results are compared with a recent spectral based algorithm, (SLLE-W*k*NN) classifier [98].

3.4.3 RESULTS ON KSC DATA

One of our goals in this experiment is to compare the random selection method and the sparse coding technique for dictionary learning in HSI classification. With the best parameters (k = 6, b = 120 and t = 0.1) we obtained in cross-validation, we performed 10 replications of experiments with different number of basis functions in the dictionary to compare these two approaches. It should be noted that 40% of labeled pixels were used for training and 30% of labeled pixels were used for testing in each of the experiments. Table 2 shows the overall accuracy (OA) results, which measures the ratio of correctly classified pixels to all test pixels, of each dictionary methods. As seen from Table 2, the randomly selected dictionaries can achieve similar or slightly better results than those obtained by sparse coding. The success of the randomly selected dictionaries implies that we can implement a very efficient system for remote sensing.

We also compared our results with those obtained by a recently proposed method for HSI classification, SLLE-W*k*NN classifier, in terms of OA. In SLLE-W*k*NN classifier, a kernel function of locally linear embedding (LLE) algorithm is employed to determine the weights of a weighted *k*NN classifier. In other words, the local structure of the distribution of the manifold is learned by employing LLE algorithm in conjunction with the *k*NN classifier.

Table 2: Comparison of dictionary learning algorithms.

Number of basis functions (*K*)	**50**	**100**	**150**	**200**	**250**	**300**
OA of sparse coding (%)	94.18 ±0.69	94.64 ±0.51	94.73 ±0.46	94.87 ±0.57	94.85 ±0.59	94.69 ±0.55
OA of random functions (%)	94.27 ±0.72	94.80 ±0.54	95.09 ±0.52	95.15 ±0.48	95.21 ±0.45	95.12 ±0.49

In the comparison, we used the same training and testing datasets for different classifiers. The parameters of the proposed algorithm were b = 120 and t = 0.1. The parameters of SLLE-W*k*NN were set the same as those reported in [98]. In particular, the dimensionality of the manifold were fixed to the number of classes and 35 nearest neighbors took into account in SLLE-W*k*NN. Results from the 10 replications of experiments are shown in Table 3. It is observed that the proposed algorithm with random dictionary is significantly better than the SLLE-W*k*NN method and the system with a learned dictionary is also significantly better than the SLLE-W*k*NN method. The system with a learned dictionary performs similarly to that with a randomly selected dictionary.

Table 3: Comparison of proposed algorithm and SLLE-W*k*NN classifier.

Classifier	**Classification Accuracy (%)**	**Standard Deviation**
Linear SVM using sparse coding for dictionary learning	94.83	0.59
Linear SVM with random dictionary	95.21	0.45
SLLE-W*k*NN (*k*=35)	93.13	0.52

3.5 EXPERIMENTAL RESULTS OF SC-DFR METHOD

In this section, we provide experimental results to validate the effectiveness of the SC-DFR model described in Section 3.3 on three well-known HSIs: AVIRIS Indian Pines [100], University of Pavia and Center of Pavia [101]. We utilized SVM and CKSVM to perform the classification task based on the new representation from the SC-DFR model and compared our method with two baseline methods, where SVM and CKSVM were directly applied to HSI pixels for classification. We also compared the proposed method with the SOMP [12] method and the image fusion and recursive filtering (IFRF) [102] algorithm.

3.5.1 QUALITY INDEXES

For the performance evaluation, three quality indexes, overall accuracy (OA), average accuracy (AA) and kappa coefficient [103] were adopted. OA measures the ratio of correctly classified pixels to all test pixels, AA is the average accuracy for all classes, and kappa coefficient estimates the percentage of the correctly classified pixels with respect to the number of agreements that are correct by random guess.

3.5.2 EXPERIMENT SETUP AND PERFORMANCE EVALUATION

The hyperparameters for SVM and CKSVM (polynomial kernel degree, gaussian kernel parameter, regularization parameter, composite kernel weight, and window size) were optimized by five-fold cross-validation. The one-against-one strategy was used for multiclass classification in SVM and CKSVM. For the experiments with SVM, we used LIBSVM (Ref. 104) codes. For the SOMP algorithm, reported parameters (window size and sparsity level) in Ref. [12] were used. In IFRF [102], the dimension of the spectral data was reduced first with image fusion; then

the recursive filtering [105] was used to combine the spatial information and extract the final representation for a pixel. We used five-fold cross-validation to tune the parameters (number of features, spatial and range standard deviations of the recursive filter) in IFRF. The MATLAB® codes of the SOMP (Ref. 106) and IFRF (Ref. 107) used in the experiments are available online.

There are three parameters in the proposed method to tune: (1) length of each sub-band called the receptive field length, b, (2) threshold for encoding, t, and (3) the trade-off in Equation (16) to adjust the mixture of spatial and spectral information, μ. To evaluate different settings of the parameters on the classification performance, we created multiple dictionaries from the training data samples with 200 basis functions, k. This yielded 4,000 features for each pixel. To remove the redundant information, this number was decreased to 600 features in the feature selection step. The two parameters, threshold and trade-off, were optimized using cross-validation. In the following, we compare the proposed method with recently proposed HSI classification algorithms described above in terms of classification performance. In the experiment part, we also study the impact of these parameters on the classification performances of AVIRIS Indian Pines data. The highest classification accuracies are given in bold in Tables 5 to 10, 12, 13 and 15 for each individual classes and quality indices.

Experiments with AVIRIS Indian Pines Dataset

This dataset, one of the bands is shown in Figure 18, was collected by NASA AVIRIS over Indian Pine site in June 1992 [100]. The original data consisted of 220 bands with a 20-m spatial resolution and a 10-nm spectral resolution over the range of 200 to 2400 nm. After noise corrupted bands were removed, there were 200 bands left and those were used in this experiment. The ground truth was developed by labeling 10,366 pixels out of 21,025 (145 × 145) pixels. These data have 16 classes as shown in Figure 19(a) and Table 4. For each class, we

randomly chose 10% of the labeled pixels for training and the remaining labeled pixels for testing as shown in Table 4. One each of the training and test sets from one of the multiple runs is visually shown in Figure 19(b) and 19(c), respectively.

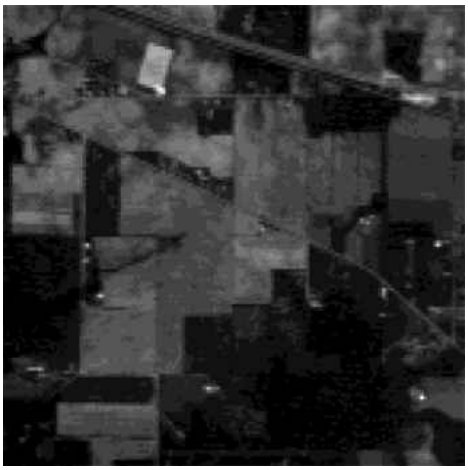


Figure 18: A sample band of Indian Pines HSI data.

Table 5 shows the AA of 10 runs for each class, OA, AA, and kappa obtained by the different methods on the test set. One of the obtained classification maps for each method is shown in Figure 19(d)-19(i). From Table 5, there are a few observations to be mentioned. First, all spectral-spatial algorithms outperformed SVM, a baseline method that was based only on spectral characteristics. This emphasizes the importance of the spatial information for HSI classification. Second, the proposed feature representation method improved the OA of SVM from 83.84 to 98.15% and OA of CKSVM from 94.11 to 98.31%. Third, the two feature extraction methods, IFRF and SC-DFR, have significantly higher accuracies for most of the classes, including classes 7 and 9, which have very limited samples for training. Fourth, the proposed method performed the best in terms of OA, AA, and kappa.

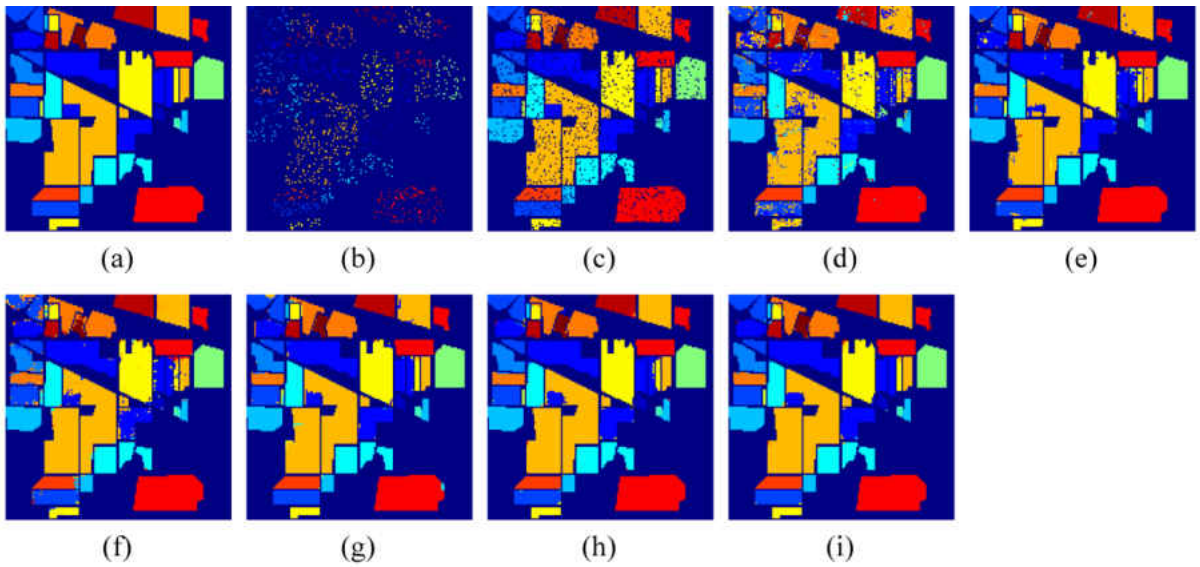


Figure 19: Pines image. (a) ground truth, (b) training data, and (c) test data. Classification maps obtained by (d) SVM, (e) CKSVM, (f) SOMP, (g) IFRF, (h) SC-DFR with SVM and (i) SC-DFR with CKSVM.

Table 4: Class names and number of labeled Indian Pines data.

Class		**Samples**	
No	**Name**	**Train**	**Test**
1	Alfalfa	6	48
2	Corn-notill	144	1,290
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	Soybeans-notill	97	871
11	Soybeans-mintill	247	2,221
12	Soybeans-clean	62	552
13	Wheat	22	190
14	Woods	130	1,164
15	Building-Grass-Trees-Drives	38	342
16	Stone-steel towers	10	85
	Total	1,043	9,323

Table 5: Classification accuracies (%) for Indiana Pines HSI.

Class	SVM	CKSVM	SOMP	IFRF	SC-DFR with SVM	SC-DFR with CKSVM
1	70.00	92.09	87.92	**99.58**	95.21	95.00
2	80.15	93.36	92.53	96.83	97.81	**98.12**
3	72.71	94.40	89.68	98.00	97.45	**98.31**
4	72.86	94.92	92.10	94.89	**98.95**	98.05
5	94.27	97.86	93.62	**99.41**	97.56	97.52
6	95.97	98.53	98.57	98.45	**98.75**	98.18
7	65.65	93.00	54.35	**100**	94.35	94.78
8	97.91	99.33	99.73	**100**	99.50	99.80
9	67.22	85.63	9.44	77.52	88.89	**89.44**
10	66.83	92.90	85.72	96.52	96.19	**97.05**
11	85.21	96.44	97.19	97.91	98.55	**98.80**
12	82.46	94.67	86.92	96.62	**97.55**	97.07
13	99.37	99.60	96.84	96.83	**99.53**	99.21
14	95.57	98.36	99.36	**99.91**	99.48	99.30
15	64.33	92.67	97.75	95.02	97.84	**99.18**
16	88.71	93.97	92.00	**97.26**	96.71	92.82
OA	83.84 ±0.79	95.86 ±0.38	94.11 ±0.43	97.74 ±0.39	98.15 ±0.23	**98.31 ±0.32**
AA	81.20 ±1.57	94.88 ±1.30	85.86 ±1.44	96.55 ±1.56	**97.14 ±1.77**	97.04 ±1.76
Kappa	81.53 ±0.89	95.28 ±0.43	93.27 ±0.49	97.43 ±0.45	97.89 ±0.26	**98.07 ±0.36**

We also conducted experiments using various rates of data for training (from 1 to 50%) to observe how the number of training samples will affect the classification performances of the proposed method. As shown in Table 6, when we compare these results with the results of SVM, CKSVM, SOMP, and IFRF algorithms, it is clear that the proposed algorithm had better performance than the other methods in all training data samples. Furthermore, the classification

accuracies with a limited number of training samples are significantly higher than all of the previous results we have found in literature for the considered HSI data.

Table 6: OA of the proposed algorithm with various training data samples.

Classifier	**OA (%)**					
	1%	**2%**	**5%**	**10%**	**30%**	**50%**
SVM	67.53 ±4.12	73.75 ±2.82	82.15 ±1.14	83.84 ±0.79	89.55 ±1.21	90.73 ±0.89
CKSVM	71.05 ±3.41	81.35 ±2.55	90.35 ±1.56	95.86 ±0.46	97.54 ±0.38	98.55 ±0.32
SOMP	70.56 ±2.60	77.80 ±1.91	89.05 ±1.62	94.11 ±0.43	97.13 ±0.31	97.32 ±0.28
IFRF	82.07 ±3.31	90.03 ±2.16	95.88 ±0.61	97.74 ±0.39	99.17 ±0.28	99.59 ±0.11
SC-DFR with SVM	**85.63** ±1.93	**91.43** ±0.75	96.11 ±0.56	98.15 ±0.23	99.43 ±0.09	99.68 ±0.07
SC-DFR with CKSVM	84.98 ±2.35	90.89 ±1.13	**96.48** ±0.67	**98.31** ±0.32	**99.53** ±0.13	**99.83** ±0.05

Next, we studied the impact of the parameters (receptive field length, threshold value in soft thresholding, and trade-off parameter in weighted summation) and feature selection step on the classification performances of the AVIRIS Indian Pines data. The length of each sub-band, receptive field length, b, can significantly influence the performance of the classification. It is notable that a larger receptive field could allow recognizing more bands in the pixel. On the other hand, a larger receptive field could extract a fewer number of sub-bands for each input HSI pixel, decreasing the discriminating power. The number of overlapped sub-bands that can be extracted for each HSI pixel is (B-b+1), where B is the number of bands and b is the receptive

field length. Before the encoding step, we had a (B-b+1)-by-b matrix to represent each input pixel. Table 7 summarizes the evaluation results with different receptive field lengths (varying from 60 to 200). It is observed that longer receptive field lengths (a fewer number of sub-bands) give better results compared to shorter lengths until a maximum value is reached. The best accuracy could be obtained if the receptive field length was 180. Note that experiments with the whole spectral data with a receptive field length of 200 did not need the construction and pooling steps to get the final representation. This table was generated based on the average performance of 10 randomly selected dictionaries. For the other parameters, we set the trade-off parameter μ = 0 and the threshold value in encoding t = 0.1 in this evaluation.

Table 7: Effect of receptive field length on performance.

Receptive field length (length of each sub-band) (b)	60	80	100	120	140	160	180	190	200
Number of sub-bands for each pixel (B-b+1)	141	121	101	81	61	41	21	11	1
Number of elements to represent a pixel (B-b+1)-by-b	8,460	9,680	10,100	9,720	8,540	6,560	3,780	2,090	200
Overall Accuracy (%)	95.53 ±0.52	94.95 ±0.54	95.97 ±0.35	96.40 ±0.39	96.99 ±0.30	97.08 ±0.39	**98.14** ±0.34	98.02 ±0.25	97.02 ±0.26

In the encoding step of the proposed algorithm, the soft threshold function was used to mimic the sparse coding. This function is also called the shrinkage function since it eliminates the insignificant representations of a pixel. We tested the threshold value t from 0 to 1 with a step size of 0.1. Table 8 illustrates the effect of this parameter with the other parameters being set as

μ = 0 and b = 180. It is observed that the performance was not sensitive to the threshold if it was small and t = 0.1 gave the best accuracy for this dataset.

Table 8: Effect of soft thresholding value, *t*.

Threshold for encoding (t)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Overall Accuracy (%)	97.73 ±0.38	**98.09** ±0.27	97.79 ±0.42	97.66 ±0.44	97.65 ±0.48	97.49 ±0.52	97.33 ±0.55	97.06 ±0.60	96.85 ±0.64	96.65 ±0.65	96.53 ±0.54

In the proposed algorithm, we aimed to enhance the classification accuracy by taking into account spatial information using Equation (19) as compared to the classical approaches that utilized spectral information only. In this experiment, we investigated the effect of the trade-off value, μ, in the range of [0, 1] on the classification performance. The spatial information was obtained by computing the mean of the neighbors in a window of size 9 × 9 for each pixel in each spectral band. As reported in Table 9, the spatial information enhanced the classification significantly. Smaller values of μ yielded better results compared to larger ones and weighted sum of the spectral and spatial features achieved the best overall performance when the weight of the spectral features was zero. Nevertheless, this does not mean the spectral content has no contribution to the performance. It should be noted that spatial features were aligned along the bands of HSI pixel. Thus, there were still spectral relationships among the spatial features of the pixel. The parameters were set as b = 180, and t = 0.1 in this experiment.

Table 9: Effect of tradeoff value μ on OA.

The trade-off value, μ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Overall Accuracy (%)	**98.07** ±0.27	97.60 ±0.33	96.90 ±0.31	95.50 ±0.30	93.49 ±0.29	91.04 ±0.41	88.29 ±0.44	85.37 ±0.52	82.50 ±0.51	80.23 ±0.52	77.24 ±0.75

After the encoding step, we obtained new representations for HSI pixels. However, the dimension of the final representation was usually huge, particularly when a large number of basis functions was employed, and most of the data may be redundant. In the study, we proposed the l_1/l_q regularized multiclass logistic regression model to obtain a compact feature representation that decreased computational time and memory requirements in the classification step.

In this experiment, we first analyzed the OA performances of 200 basis functions without feature selection. Then we applied the feature selection step to the same features to reduce their dimensions down to various sizes ranging from 100 to 1600. OA performances with and without feature selection are shown in Table 10. The dictionary sizes and feature representation dimensions are shown in the first and the second columns of the table, respectively. The third column shows the OA performance of the algorithm without the feature selection step. The remaining columns show the accuracy performances with various selected features. For most cases, the size reduced feature representation achieved similar performances, such as the 600 selected from the original 4000 features obtained almost the same result (98.26 vs. 98.15%). From the results, it can be concluded that the l_1/l_q regularized multiclass logistic regression model has the ability to eliminate redundant features. The other parameters were set as b = 180, μ = 0 and t = 0.1.

Table 10: Effect of feature selection.

The size of the dictionary (k)	The size of the representation of the pixels (without feature selection)	OA without feature selection (%)	OA with feature selection (%)											
			Number of selected features											
			1600	1400	1200	1000	800	700	600	500	400	300	200	100
200	4,000	98.14 ±0.39	98.21 ±0.41	98.18 ±0.42	98.17 ±0.40	98.15 ±0.40	98.10 ±0.39	98.09 ±0.39	98.12 ±0.38	98.03 ±0.42	97.93 ±0.44	97.87 ±0.45	97.71 ±0.48	96.85 ±0.53

Experiments with ROSIS University of Pavia Dataset

The University of Pavia dataset, one of the bands is shown in Figure 20, was obtained by ROSIS. This sensor generated 115 spectral bands ranging from 430 to 860 nm, and it has 1.3-m per pixel spatial resolution. The University of Pavia image consists of 610 × 340 pixels in which each pixel has 103 spectral bands after the noisy bands are removed. This dataset contains nine ground truth classes of interest as shown in Figure 21(a). We used the same training and testing data partition as that in Refs. 28, 12, and 13 where ~9% of the data were used for training and the rest was used for testing. The details of the dataset, the training pixels, and the test pixels are shown in Table 11, Figure 21(b) and 21(c), respectively.

Table 12 reports the classification accuracy for each class: the OA, AA and kappa obtained by different methods. The classification results are visually mapped in Figure 21(d)-21(i). The parameters of the proposed algorithm were set as b = 60, t = 0.4, and μ = 0.1. We used the neighbors of a pixel in a window size of 5 × 5 pixels per spectral band for the spatial information since this image contains an urban area with small buildings.

Figure 20: A sample band of the University of Pavia HSI data.

Table 11: Class names and number of labeled University of Pavia data.

Class		**Samples**	
No	**Name**	**Train**	**Test**
1	Asphalt	548	6,304
2	Meadows	540	18,146
3	Gravel	392	1,815
4	Trees	524	2,912
5	Metal sheets	265	1,113
6	Bare soil	532	4,572
7	Bitumen	375	981
8	Bricks	514	3,364
9	Shadows	231	795
	Total	3,921	40,002

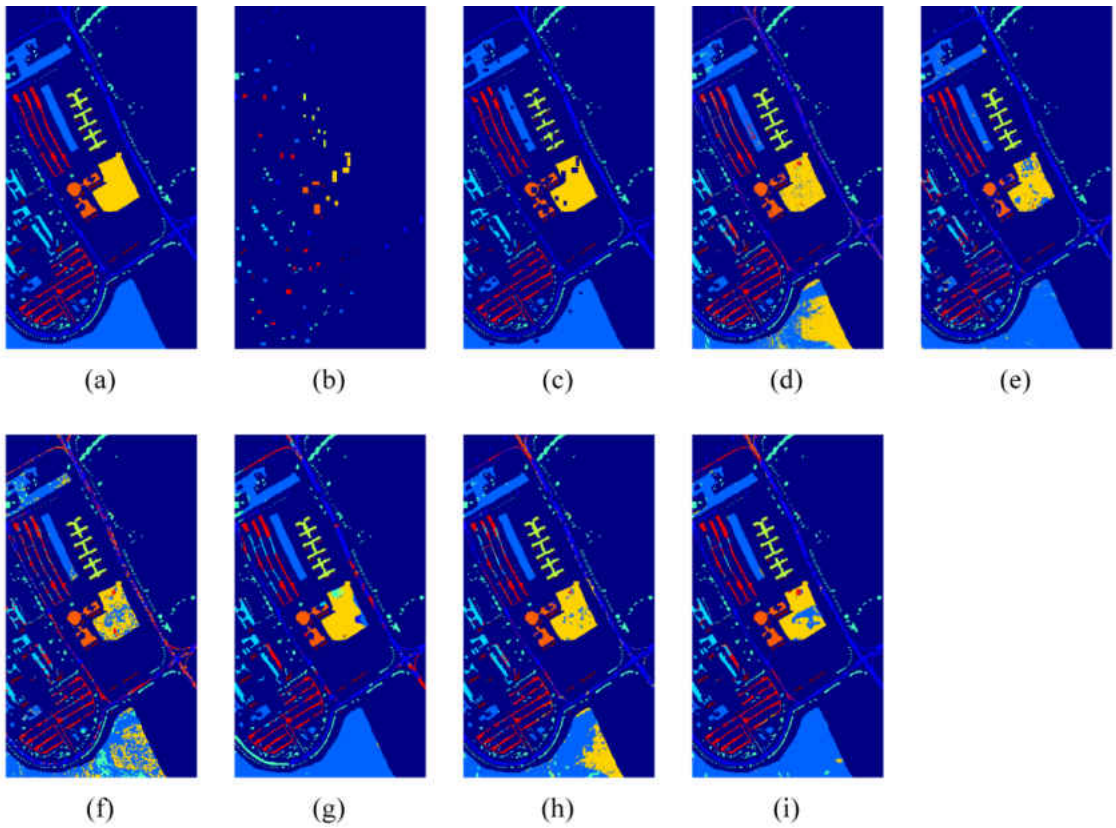


Figure 21: University of Pavia image. (a) ground truth, (b) training data, and (c) test data. Classification maps obtained by (d) SVM, (e) CKSVM, (f) SOMP, (g) IFRF, (h) SC-DFR with SVM and (i) SC-DFR with CKSVM.

As seen from Table 12, the proposed feature representation method improved the OA of SVM from 76.77 to 83.10% and OA of CKSVM from 89.74 to 91.31%. Spectral-based algorithm SVM and SOMP provided poor results as compared to the other classifiers. CKSVM, IFRF, and SC-DFR integrated with CKSVM performed better due to their capability to reveal both spectral and spatial information. Moreover, SC-DFR with CKSVM yielded the best OA, AA, and kappa for this dataset. Here, it should be noted that the spatial features were employed in both the feature representation learning and CKSVM processes.

Table 12: Classification accuracies (%) with fixed samples for University of Pavia HSI.

Class	SVM	CKSVM	SOMP	IFRF	SC-DFR with SVM	SC-DFR with CKSVM
1	79.54	88.02	50.57	**93.41**	80.26	86.12
2	66.23	92.19	71.68	**98.03**	77.55	94.95
3	73.55	80.66	73.77	66.33	**86.93**	83.25
4	93.68	**99.46**	96.94	71.94	97.76	98.17
5	99.82	75.31	99.91	90.00	99.97	**100.00**
6	91.86	85.12	64.72	**99.58**	92.24	80.73
7	85.83	**94.62**	95.41	91.08	87.10	91.04
8	82.82	**95.47**	83.32	64.60	80.98	90.11
9	84.91	95.47	94.21	**98.13**	97.84	97.11
OA	76.77	89.74	72.29	89.34	83.10 ±1.27	**91.31 ±1.06**
AA	84.25	89.78	81.17	85.90	88.96 ±0.44	**91.27 ±0.59**
Kappa	70.64	86.28	65.23	85.66	78.19 ±1.49	**88.32 ±1.33**

The accuracy results of this dataset were significantly worse than the Indian Pines and Center of Pavia datasets as confirmed by other studies [13] by using these fixed training and testing samples. It can be clearly observed in Figure 21 that the regions that have false positive pixels do not have training samples. For further evaluation, we randomly selected the same number of training and testing samples as specified in Table 11 and obtained results with different classification methods as illustrated in Table 13. As confirmed in Ref. 13, the performance of the random set was significantly better than those shown in Table 12.

Table 13: Classification accuracies (%) with random samples for University of Pavia HSI.

Performance Metrics	CKSVM	IFRF	SC-DFR with SVM	SC-DFR with CKSVM
OA	98.13 ±0.13	98.96 ±0.09	97.85 ±0.13	**99.07** ±0.08
AA	97.16 ±0.23	98.44 ±0.25	96.92 ±0.29	**98.79** ±0.15
Kappa	97.38 ±0.18	98.64 ±0.12	97.17 ±0.21	**98.78** ±0.11

Experiments with ROSIS Center of Pavia Dataset

The last HSI data studied in this study is Center of Pavia, which was obtained by ROSIS. Figure 22 shows one of the bands of the data. The captured 115 spectral bands were decreased to 102 spectral bands after removing 13 noisy ones. This HSI consists of 1096 × 492 pixels and nine ground truth classes of interest as shown in Figure 23(a). We used the same training and testing data partition as that in Refs. 28, 12, and 13, where the training set was ~5% of all labeled pixels and the rest was used for testing. The nine ground truth classes of the dataset, the training pixels, and the testing pixels are shown in Table 14, Figure 23(b) and 23(c), respectively.

Figure 22: A sample band of the Center of Pavia HSI data.

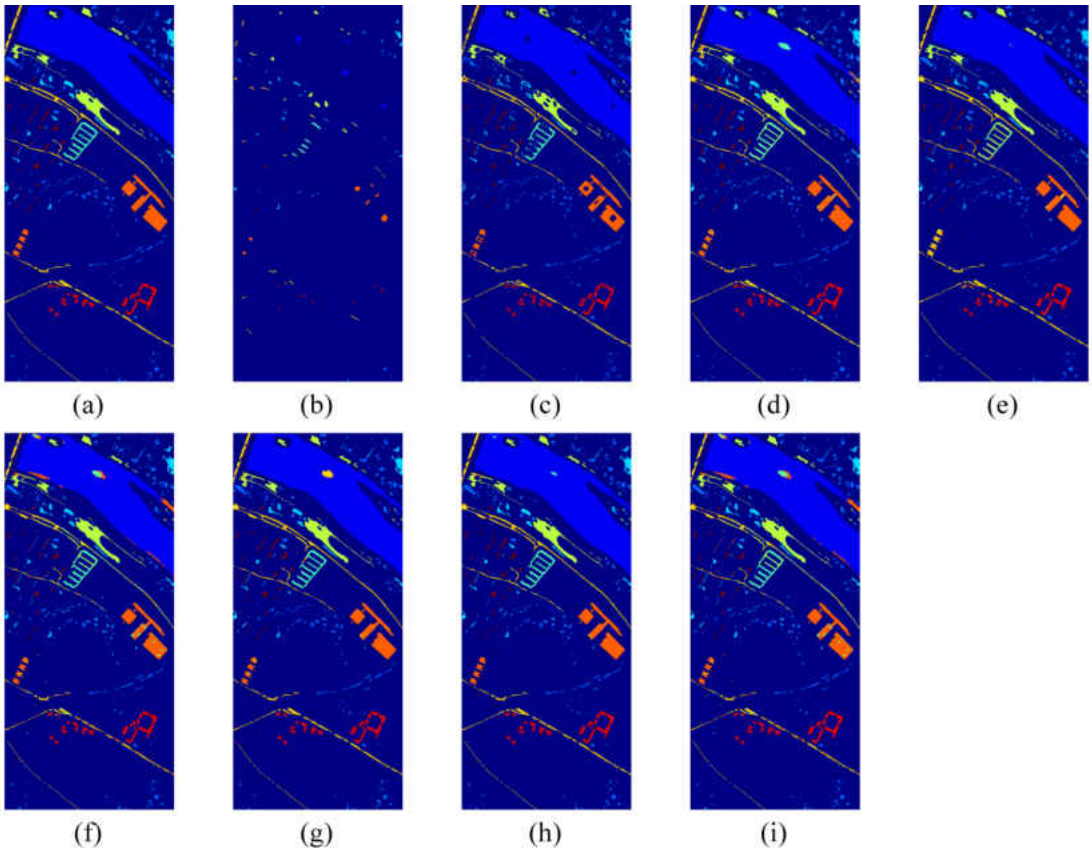


Figure 23: Center of Pavia image. (a) ground truth, (b) training data, and (c) test data. Classification maps obtained by (d) SVM, (e) CKSVM, (f) SOMP, (g) IFRF, (h) SC-DFR with SVM and (i) SC-DFR with CKSVM.

Table 14: Class names and number of labeled Center of Pavia data.

Class		Samples	
No	**Name**	**Train**	**Test**
1	Water	745	64,533
2	Trees	785	5,722
3	Meadows	797	2,094
4	Bricks	485	1,667
5	Soil	820	5,729
6	Asphalt	678	6,847
7	Bitumen	808	6,479
8	Tile	223	2,899
9	Shadow	195	1,970
	Total	5,536	97,940

The classification results for each class, OA, AA, and kappa obtained with different classifiers are summarized in Table 15 and the classification maps are shown in Figure 23(d)-23(i). The parameters of the proposed algorithm were set as $b = 60$, $t = 0.1$, $\mu = 0.4$. The spatial information was obtained from the average of the surrounding neighbors of a pixel in a 7×7 pixel window per each spectral band. The proposed feature representation method improved the OA of SVM from 95.02 to 98.56% and the OA of CKSVM from 96.11 to 98.70%. The proposed algorithm integrated with CKSVM yielded the best OA, AA, and kappa on this dataset.

Table 15: Classification accuracies (%) for Center of Pavia HSI.

Class	SVM	CKSVM	SOMP	IFRF	SC-DFR with SVM	SC-DFR with CKSVM
1	97.90	96.96	97.36	**99.96**	99.64	99.79
2	86.58	94.51	83.84	**98.53**	90.01	90.13
3	96.30	97.15	95.78	87.26	**97.61**	97.30
4	67.61	**99.04**	70.10	94.10	98.30	98.32
5	91.29	94.47	96.04	**98.86**	97.27	97.54
6	82.67	92.24	77.37	91.11	96.50	**96.54**
7	94.17	93.09	92.17	95.77	98.83	**99.10**
8	99.52	99.72	98.17	98.87	99.99	**100.00**
9	97.31	92.18	94.77	**100.00**	97.25	97.51
OA	95.02	96.11	94.22	98.44	98.56 ±0.10	**98.70 ±0.09**
AA	90.37	94.65	89.51	96.05	97.27 ±0.10	**97.36 ±0.09**
Kappa	91.07	93.17	89.52	97.17	97.38 ±0.18	**97.64 ±0.17**

3.6 DISCUSSIONS

In this chapter, we first proposed a spectral-based sparse coding framework for HSI classification and experimentally showed that a randomly selected dictionary can achieve slightly better results than the dictionary learned by an optimization problem. Then, we enhanced the algorithm considering the spatial-spectral structures of HSI pixels. In SC-DFR, first, sub-bands were constructed to populate a set of basis functions. Then new representations for pixels were obtained by projecting the spatial and spectral data onto the basis functions in a convolutional manner. A feature selection model was used to obtain a compact final representation for classification. We evaluated the proposed system on three HSI benchmark

datasets and compared it with four methods: SVM, CKSVM, SOMP, and IFRF. The proposed system outperformed the other methods on three datasets. One unique property of the proposed system is that the performance is much better than the others if labeled data is limited. This property gives the method special advantages because labeling HSI data pixel by pixel is difficult and time-consuming.

There are four steps in the proposed system: sub-band construction, dictionary learning, encoding, and feature selection. To obtain new representations, overlapping sub-bands in HSI were utilized. In dictionary learning, we showed that randomly selected dictionaries performed very well, which is of particular interest because the computational resources required for the random selection are negligible. In the encoding step, soft thresholding, which mimics the sparse coding method, was able to extract new features relatively fast. In addition, the feature selection step effectively reduced the dimensionality of the new pixel representation without performance degradation.

There are three main parameters in the proposed methods: (1) receptive field length, (2) threshold in soft thresholding, and (3) mixing coefficient between spectral and spatial information. The receptive field length is the size of the sub-band along the spectral direction in HSI data. We observed that for the data we tested, a very small sub-band may not be able to capture the global characteristics along the spectral direction and a long sub-band may miss the discriminative details. Thus, this parameter should be set by cross-validation. In soft thresholding, our experiments showed that the performance of the system is sensitive to the threshold value and a small value can usually provide good results. The mixing coefficient makes trade-offs between spectral and spatial information. In our experiments, a small mixing value (<0.4) usually achieved good results. Moreover, in the Indian Pine dataset, we observed

that the best overall performance was achieved when the weight of the spectral features was zero. We believe that spatial features also have spectral information since spatial features were aligned along the bands of HSI pixels. The spatial information may be more important for the discrimination of different materials in HSI data. However, from the classification maps, we also observed that misclassified pixels usually occurred at edges of class regions. The main reason of this misclassification is that the use of a square window in sparse coding resulted in mixed pixels at class boundaries. To correct the misclassified pixels at class boundaries, a neighborhood method driven by nearest neighbors to the current pixel can be used to calculate the spatial information as proposed in [108]. In addition, from the experiments with various sizes of basis functions on KSC data, we found that large dictionaries can usually improve the classification performance (Table 2). However, increasing the number of basis functions more than required is not necessary for our data since more basis functions will require more computational resources.

3.7 CONCLUSION

In summary, we have implemented an efficient sparsity-based feature representation model for HSI image classification. The proposed system utilized random dictionaries and then used the soft thresholding technique to efficiently obtain a new representation for each pixel in HSI data. In addition, a feature selection strategy was employed to further eliminate redundant features in the new presentation, achieving an efficient classification model for remote sensing applications. We evaluated the system on three HSI benchmark datasets and compared the results with state-of-the-art performances. The proposed system performed the best on three datasets, making it a good candidate for efficient HSI data classification.

Chapter 4

DETECTION OF SEAGRASS SCARS USING SPARSE CODING AND MORPHOLOGICAL FILTER

In this chapter, we apply the SC-DFR algorithm described in Section 3.3 to multispectral pan-sharpened bands to detect seafloor propeller seagrass scars in shallow water. This application incorporates multiple image processing methods including pan-sharpening, classification, and morphological operations. We first introduce the scar detection problem, the proposed method will be followed by experimental results.

4.1 INTRODUCTION OF SCAR DETECTION PROBLEM

Seagrasses (marine angiosperms) inhabit the shallow coastal waters of temperate and tropical seas throughout the world. They provide a critical habitat and food source for a variety of wildlife, improve water quality, stabilize coastal sediments and represent an important reservoir of organic carbon (blue carbon) in oceans [109]. However, seagrasses are particularly vulnerable to the human impact on coastal environments creating a global crisis in terms of their long-term sustainability [110]. Along heavily populated coasts, including the US Gulf of Mexico, mechanical damage from propeller scars caused by recreational boats operating in very shallow water can have significant impacts on seagrass distributions (Figure 24). It can take years for the scars to disappear and they often accelerate more seagrass loss by facilitating erosion at scar boundaries. Although remote sensing can provide an excellent tool for imaging propeller scars in shallow water, detection and quantification of scars has not been automated to enable rapid and reliable assessment of the damage they cause in coastal waters. There are many challenges for

accurate scar detection and quantification. Firstly, scars can have different shapes including straight lines, smooth curves, circles and ellipses. Secondly, scars can also have various lengths and widths that vary with seafloor bathymetry and age. Thirdly, although scars are usually defined as bright quasi-linear regions surrounded by dark vegetation, they can sometimes appear as dark lines in very shallow water, particularly at low tide when adjacent seagrass leaves fill the depressions. Finally, some structural elements on the seabed are as bright as scars. All of these variations pose significant challenges for the automatic detection and quantification of scars.

Figure 24: Examples of prop scar.

One effective way to track the distribution of seagrass scarring is to use remote sensing surveys [111-112]. In [111], aerial photography of Charlotte Harbor, Florida, gathered in 1999 was used to estimate scar distribution. Delineations of scarred habitat polygons were first outlined using ArcGIS 8.3 software. An image analyst then manually rated these scar polygons as light, moderate and severe scarring polygons, based on which areas in the generated map showed regions of light, moderate and severe scarring. Another similar project was performed in the Gulf Coast of Texas with digital images captured with a Leica ADS40 airborne digital

camera in 2004 [112]. The project is significant since it is one of the first successful applications that utilized automated techniques for benthic habitat mapping. In the project, an object-oriented classification scheme, which consists of a combination of field investigations, image segmentation software, and classification and regression tree (CART) analysis, was used. A recent study used the same classification approach to map the benthic habitat and identify boat propeller scars in seagrass habitats [113]. In the study, they compared the use of three digital airborne imageries, ADS40-52, DMC, and ULTRACAMD, in which images collected over Redfish Bay, Texas, with standard formats in terms of radiometric, spectral, spatial resolutions for a fair comparison. To benefit from both the spectral and spatial resolutions, they created pan-sharpened images using both panchromatic and multispectral images of the region. It was concluded that prop scar maps produced by the ADS40-52 sensor were the best among the three devices compared.

The studies described above successfully identified prop scars based on an object recognition method in which a segmentation process and a classification step were usually combined to detect the scars. However, the success of these methods depends on a segmentation scale that determines the maximum detectable scar size in images [114]. A large value for the scale may make some large non-scar objects being detected while a small value may miss true scars in images. In addition, the methods are not fully automatic, and usually need some manual editing such as field observations and injection of the classification rules, requiring human in the loop to obtain final detection results.

In this section, we develop an automatic system to detect both bright and dark prop scars in pan-sharpened images. Firstly, a set of features of scar and non-scar pixels are extracted using the SC-DFR method proposed in Section 3.3 and then a classifier categorizes each pixel in the

image as either scar or non-scar and generates an initial prop scar map. Secondly, morphological directional filters [115-117] are applied to the detected scar pixels to eliminate false positives. Directional filters process an image in a defined direction to remove objects according to their orientations. We applied the proposed algorithm to a pan-sharpened image obtained from the fusion of panchromatic and multispectral images collected by WorldView-2. Our results showed that the proposed method successfully mapped prop scars, resulting in an efficient scar detection system.

The contribution of this chapter consists of three parts: 1) we successfully detected both bright and dark-toned scars, the later have very similar pixel values with their neighborhood structures and are difficult to discriminate, 2) we developed a morphological filtering process to improve the detection results achieved by sparse coding, making the system very effective and 3) the proposed system is fully automatic once the model is trained, requiring no human intervention for scar detection.

4.2 REMOTE SENSING DATA

In this dissertation, we used panchromatic and multispectral images captured in May 2010 near Deckle Beach, Florida, USA by WorldView-2, an orbiting multispectral imaging system operated by DigitalGlobe, LLC. WorldView-2 is a remote-sensing satellite used to capture high-resolution images of the Earth. The WorldView-2 panchromatic band covers the range from 450 to 800 nm with 0.46-m spatial resolution and it has 8 multispectral bands ranging from 400 to 1050 nm with 1.84-m spatial resolution. The availability of a WorldView-2 panchromatic band, in conjunction with the multispectral bands, provides the opportunity to create a multispectral pan-sharpened image by fusing these images.

From the multispectral image, we observed propeller seagrass scars on the 3rd to 6th bands (green, yellow, red, red edge bands). To create a pan-sharpened data set for our analysis, the panchromatic image was fused with these 4 multispectral bands by using the principal component analysis (PCA)-based pan-sharpening method [118] as described in Section 4.3.1. Figure 25 shows the area near Hagen's Cove Park, FL we used to evaluate our proposed method. Each of the bands contains 720 by 820 pixels. As seen in Figure 25, our data set presents some challenges for accurate scar detection. 1) Although scars are usually represented by relatively brighter pixels, some parts of the scars consist of dark-toned pixels. 2) Brighter pixels do not indicate scar pixels only, but also some structural elements in the sea. 3) Scars have different shapes ranging from straight lines to smooth curves with various widths. To overcome these challenges, we propose a three-step algorithm in the following section.

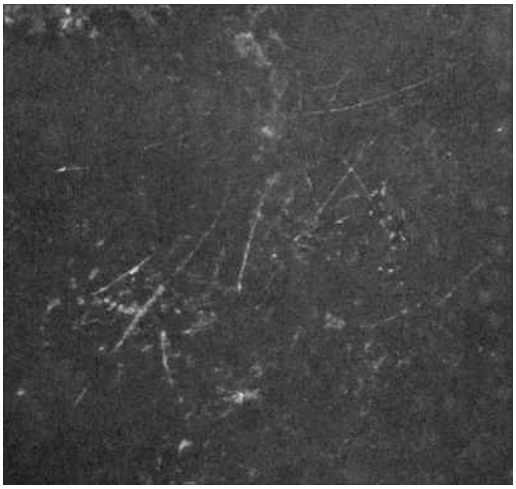


Figure 25: Panchromatic image obtained over shallow vegetated area near Hagen's Cove Park.

4.3 METHOD

Figure 26 shows the flowchart of the proposed method for scar detection. The proposed method consists of three main steps: 1) pan-sharpening of multispectral and panchromatic

images based on the PCA method, 2) classification of new representations of original pixels created by sparse coding into scar and non-scar categories and 3) derivation of local orientation (LO) of each scar pixel to eliminate false positives and to produce a final scar map. Each step is detailed in the following subsections.

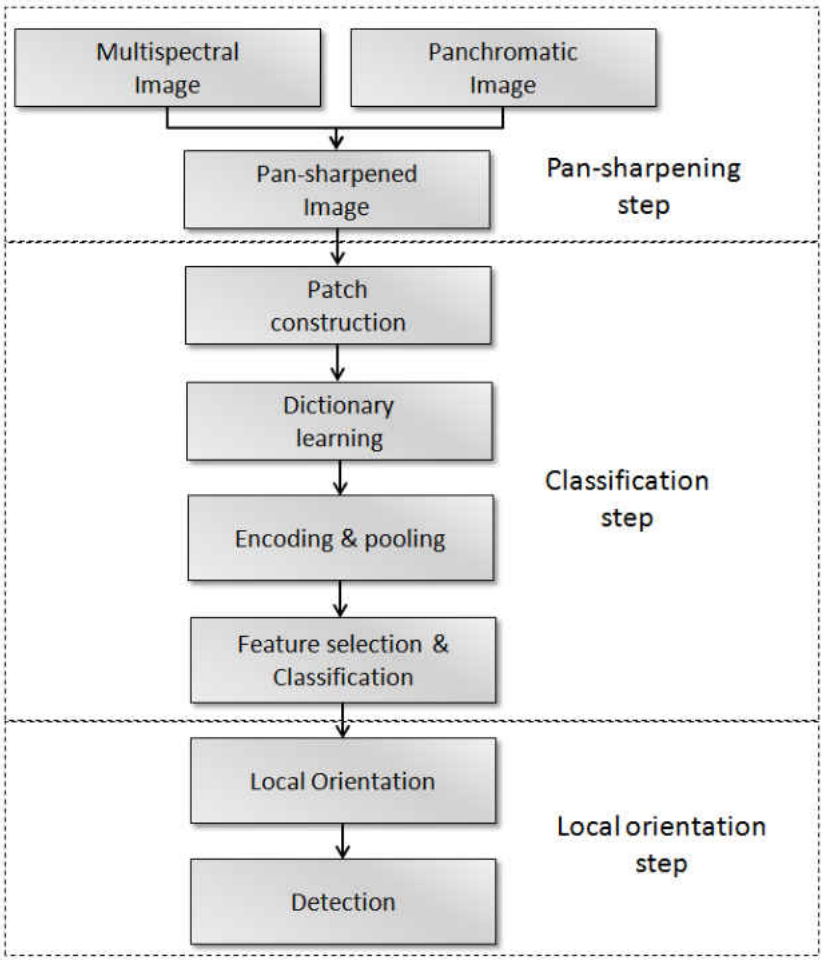
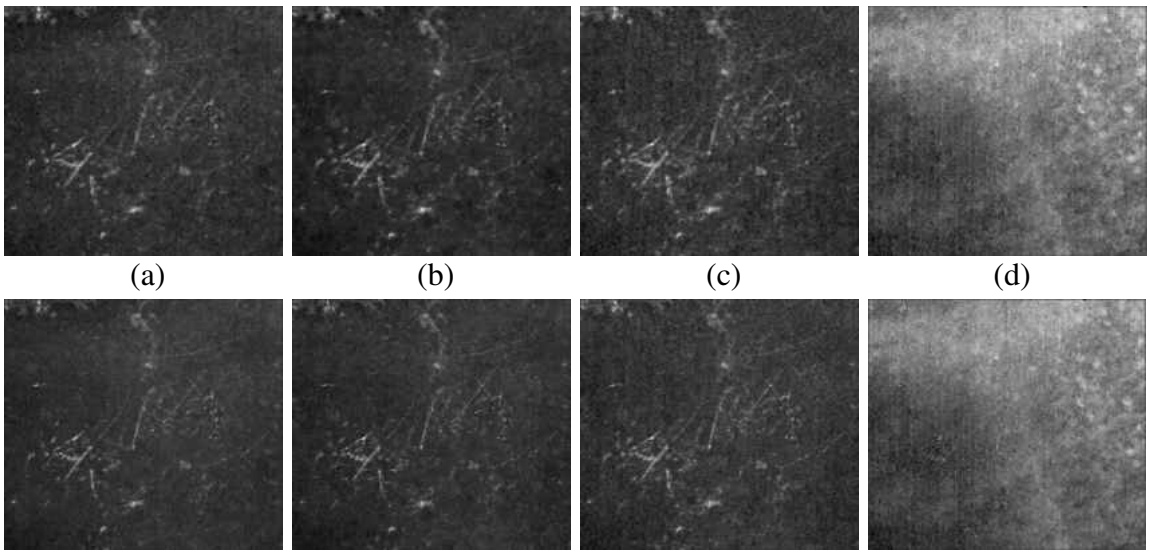


Figure 26: Flowchart of the proposed algorithm.

4.3.1 PCA based pan-sharpening method

A high spectral resolution of a multispectral image results in a low spatial resolution, and conversely, a high spatial resolution of a panchromatic image results in a low spectral resolution. To benefit from both the spectral and spatial resolutions, we fused 4 multispectral bands with the

panchromatic image using the PCA-based pan-sharpening method [118]. We first up-sampled the low spatial resolution multispectral bands to match the resolution of the panchromatic image, then we converted high correlated multispectral bands into a new set of uncorrelated components using the PCA method. We assumed that the first PC image with the largest variance contains the most significant information from the multispectral image and replaced it with the histogram-matched panchromatic image. We kept the remaining PCs considering they have spectral information. The pan-sharpened bands were obtained by performing inverse PCA on the modified panchromatic image and the remaining PCs. Figure 27 shows the original (Figure 27(a)-27(d)) and pan-sharpened multispectral bands (Figure 27(e)-27(h)).



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 27: (a) Green, (b) yellow, (c) red, (d) red edge bands of the multispectral image, and (e)-(h) pan-sharpened multispectral bands after fusion.

4.3.2 Classification of pixels using the sparse coding technique

We modified the SC-DFR method, described in Section 3.3, to get the new represenation of pan-sharpened image pixels. In this section, we describe the modified method shortly.

Patch construction

Figure 28 illustrates the construction of patches. To create our training data, first we created overlapped scar and non-scar blocks manually from the pan-sharpened image. Each block was assigned to a scar-class if the middle point (pixel) of the block represented a scar. Figure 29(a) and 29(b) show samples of scar and non-scar blocks, respectively. Each block has a dimension of $w \times w$ and has 4 bands. Then, we extracted non-overlapped patches with a dimension of $b \times b$ from the input blocks, where b is called "receptive field size." After that, we transformed each patch into a vector of dimension $B = b \times b \times 4$. Note that $x^{(i)} \epsilon R^B$ is the vector representing the i'th patch of a block. We chose $w = 36$ and $b = 6$ in this study. Then the patches were preprocessed as follows. First, each patch $x^{(i)}$ was normalized as in Equation (17). Then, the zero-phase components analysis (ZCA) whitening process was applied to patches to provide uncorrelated patches and to distribute more basis functions in the orthogonal directions.

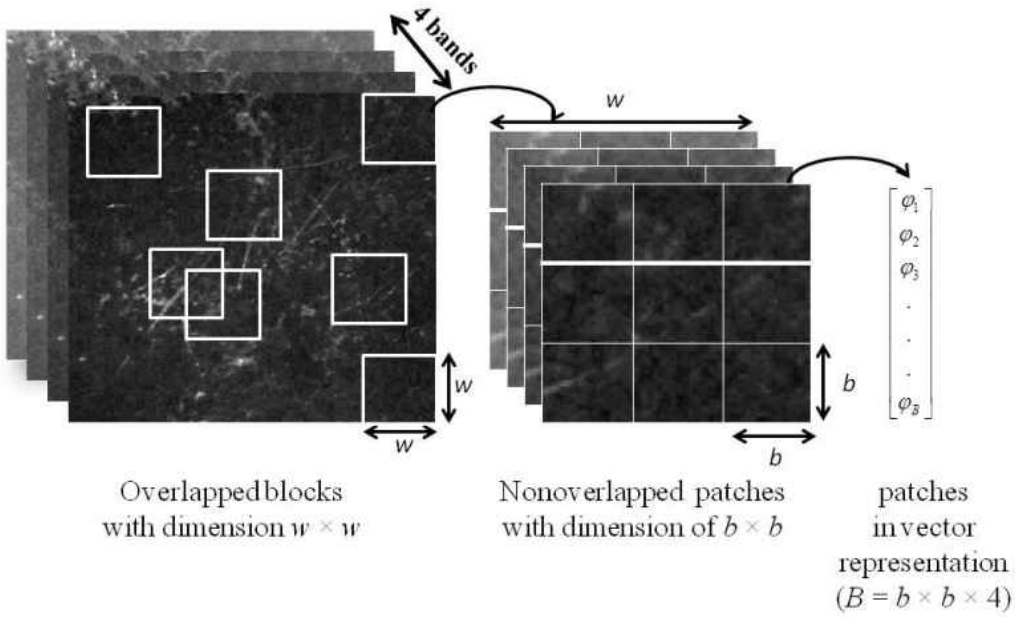


Figure 28: Construction of patches.

(a) (b)

Figure 29: Samples of (a) scar and (b) non-scar blocks.

Dictionary learning from image patches

Dictionary **D** is randomly selected from the extracted patches of the training set. If k is the total number of sampled patches for dictionary learning, we denote it as $\mathbf{D}=\{d^{(1)}, d^{(2)}, \ldots, d^{(k)}\}$, where $d^{(j)} \in R^{B}$.

Encoding and pooling

With the dictionary **D**, we obtained a new representation for a pixel as follows,

Patch construction. For a given test block of $w \times w$ pixels (with 4 bands), we constructed and preprocessed test patches in the same style as described patch construction part.

Encoding. For each patch, we obtained a new representation by applying the soft thresholding technique as described in Section 3.2.2. Then, we split the positive and negative components of the new representation into two separate features. Note that the dimension of the representation for the patch is $2 \times k$. We applied the encoding step to each patch of the block, thus, we had $(w/b)^2$-by-$(2 \times k)$ matrix for each pixel to be converted into a single feature representation in the pooling step.

Pooling. We split all representations from the $(w/b)^2$-by-$(2 \times k)$ features of the block into 6 equal-sized groups and summed them in each group to obtain 6 representations. The final representation, α, with a dimension of $12 \times k$, for the pixel was achieved by concatenating the 6 representations.

Feature selection and classification

After the encoding and pooling steps described above, the dimensionality of the new feature representation, $12 \times k$, may be undesirable. We utilized the l_1/l_q regularized multi-class logistic regression [63-66] $(q \geq 1)$ as shown in Equation (20) to select effective features for classification. It should be noted that the feature selection step was performed on training data only. The regularization parameters of the logistic regression were determined by cross-validation. Once we obtained the final representation vector for each pixel, we applied a weighted *k*nn classifier [98] to sort the pixels into scar and non-scar classes.

4.3.3 Derivation of local orientation of scar pixel for false positive reduction

After the classification step, some unwanted objects (false positives) were also classified as scars. To remove the false positives, we computed local orientations of candidate scar pixels using directional filters that define the morphological directional profile (MDP) for a pixel as described in the following subsections.

Morphological filters

Morphological filters process image objects based on the characteristics of a structuring element (SE). A SE is a binary matrix with arbitrary shape, size and orientation. Figure 30 shows four linear types of structural elements in size 5 at various orientations. Two fundamental morphological filters are dilation and erosion. A dilation filter adds pixels to the boundaries of objects in an image while an erosion filter removes pixels on object boundaries. The size and shape of the SE determines the number of pixels being added or removed. By concatenating the two fundamental morphological filters in different number and order, many complicated morphological filters can be implanted such as opening and closing. For simplicity, we only discuss opening filter in this section. The opening operation is an erosion filter followed by a dilation filter with the same SE. In the opening operation, erosion typically removes objects smaller than the SE and the subsequent dilation tends to restore the shape of the objects that remain. However, the accuracy of this restoration depends on the similarity between the shapes of the objects and the SE. On the other hand, "opening by reconstruction filter" [115, 119] can restore the original shapes of the objects as described in the following part.

0	0	0	0	0
0	0	0	0	0
1	1	1	1	1
0	0	0	0	0
0	0	0	0	0

(a)

0	0	0	0	1
0	0	0	1	0
0	0	1	0	0
0	1	0	0	0
1	0	0	0	0

(b)

0	0	1	0	0
0	0	1	0	0
0	0	1	0	0
0	0	1	0	0
0	0	1	0	0

(c)

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

(d)

Figure 30: Linear type of structural elements in size 5 and various orientations (a) 0°, (b) 45°, (c) 90°, (d) 135°.

MDP and local orientation

In the "opening by reconstruction" operation, first erosion filter removes or shrinks objects according to the characteristics of the SE and then reconstruction filter restores the shrunk objects as their original sizes. Thus, all the bright objects in an image smaller than the SE are removed. In morphological operations, directional filters, such as "opening by reconstruction" with a linear SE, can process an image in a defined direction. Thus, it is possible to preserve or remove objects with different orientations. Figure 31 shows an example of implementing "opening by reconstruction" filter with a linear SE and different orientations on a part of our binary classification map. When the orientations of the scars fit the orientation of the SE, the scars are preserved, otherwise removed from the image. This property yields to compute MDP and the LO of a pixel. The MDP of a pixel is the composition of the outputs of the directional filter with a varying orientation parameter Φ ranging from 0° to 180° [119, 120] (Figure 32). For a given pixel x, the local orientation LO(x) is defined as the difference between the maximum value and the minimum value of the MDP of a pixel [115]:

$$LO(x) = \max_{\Phi}\left(MDP(x)\right) - \min_{\Phi}\left(MDP(x)\right) \tag{19}$$

Here, it should be noted that LO of a pixel in a binary image can have either 1 or 0.

(a) (b) (c) (d)

Figure 31: Implementation of "opening by reconstruction" filter on (a) initial map with orientation (b) 30°, (c) 60° and (d) 110°.

In our algorithm, to remove false positives observed on the binary map of the classification step, we composed the MDP of the candidate scar pixels using "opening by reconstruction" filter. In the filter, we used a linear SE in size of 35 and orientation ranging from 0° to180° with a step size of 5. We observed that the oriented objects (usually scars) restored with at least one of the orientations in the range from 0° to 180°, thus the LO of the pixels that belong to these objects were computed as 1. Figure 32 shows a sample MDP of a scar pixel restored when the SE was in orientations of 15°, 20° and 25°. Conversely, the LO of the pixels that belong to a non-oriented object (usually false positives) were computed as 0 since the filter did not restore a non-oriented object in any orientation. Thus, the final scar map was obtained from 1's (oriented objects) eliminating false positives (0's) from the initial classification map.

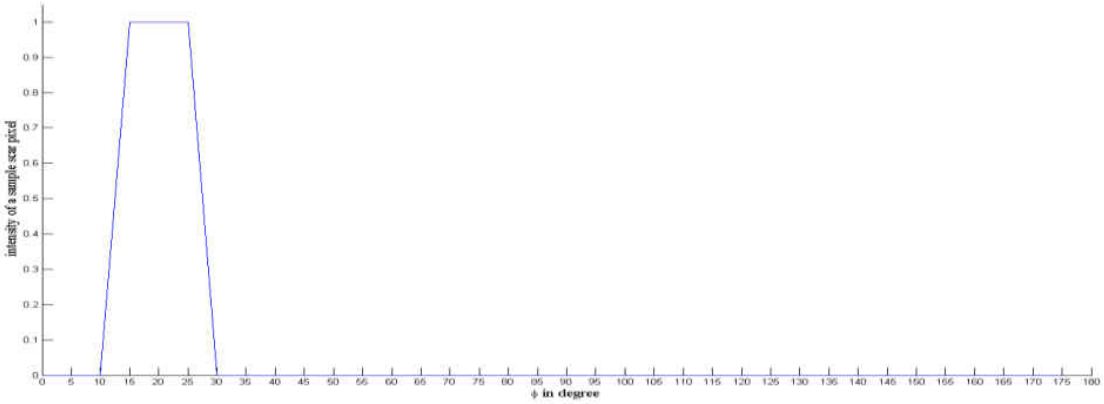


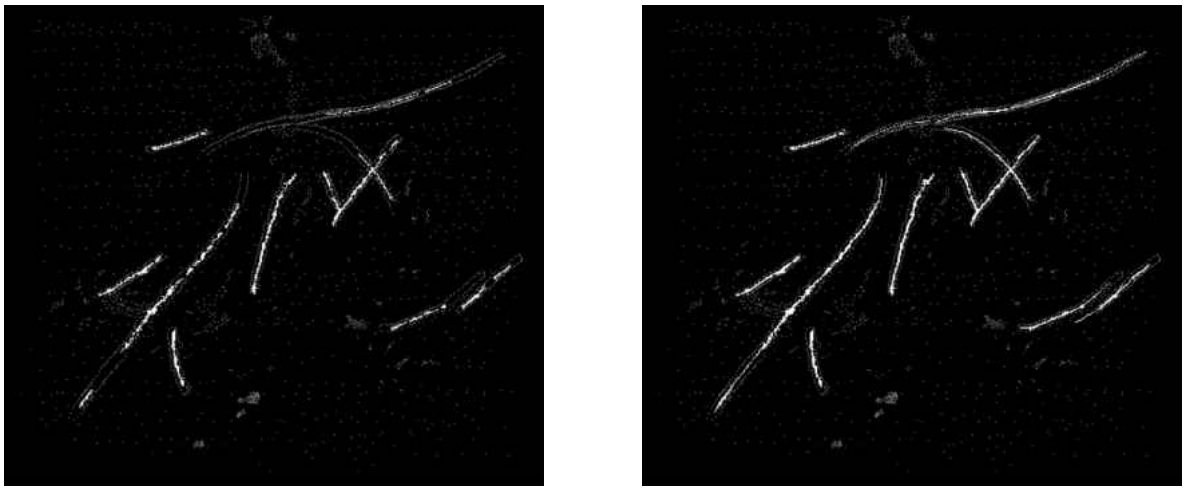
Figure 32: MDP of a scar pixel.

4.4 RESULTS

4.3.1 EXPERIMENTAL SETUP AND PERFORMANCE EVALUATION

The scars can be identified by their increased brightness over seagrass. However, sand patches, rocks or other structural elements can be also identified as bright pixels. On the contrary, scars may be identified as gray or black pixels when they are covered by seagrass

leaves. For our experiments, we created two training and test sets manually. In the first set, we labeled 1,956 bright pixels as scar pixels and 3,055 bright and dark pixels as non-scar pixels. In the second set, we added 1,308 gray and black scar pixels to the first set to investigate the effect of dark scar pixels on classification performance. The ground truth data of the first and second sets are shown in Figure 33(a) and 33(b), respectively. It should be noted we did not label all scar pixels appearing in Figure 25.



(a) (b)

Figure 33: The ground truth data of the (a) first and (b) second sets.

We conducted experiments using various amounts of data for training (from 10 to 75% of the available ground truth data) to study how the number of training samples affected the classification performance of the proposed method. We repeated the experiments ten times and computed the average performance of all ten runs. For performance metrics, we utilized classification accuracies of scar and non-scar pixels and overall accuracy (OA). OA measures the ratio of correctly classified pixels to all test pixels. There were five parameters in the proposed method: (1) number of basis functions, k, (2) size of each patch, also named receptive field size, b, (3) threshold for encoding, t, in the coding step, (4) number of neighbors in the weighted

classification algorithm, n, and (5) size of the linear SE. In this dissertation, we conducted the experiments many times and optimized those parameters. With those optimized parameters, we repeated the experiments and reported the results in the next section.

4.3.2 INITIAL PIXEL CLASSIFICATION RESULTS

Table 16 and 17 show the testing classification results of the first and second sets, respectively. In the experiment, the parameters of the proposed algorithm were set $k = 250$, $b = 6$, $t = 0.1$ and n = 15. Before the feature selection step, there were about 3,000 ($12 \times k$) features and the feature selection algorithm retained 500 of them.

Table 16: Classification results of the first data set using different amounts of data for training.

	Rate (%) of data for training			
	10%	25%	50%	75%
Accuracy of non-scar pixels	91.45 ±1.27	94.98 ±0.48	97.35 ±0.50	98.37 ±0.37
Accuracy of bright scar pixels	85.80 ±1.80	94.78 ±0.90	97.97 ±0.83	98.99 ±0.55
Overall Accuracy	89.28 ±0.87	94.90 ±0.38	97.66 ±0.42	98.60 ±0.40

Table 17: Classification results of the second data set using different amounts of data for training.

	Rate (%) of data for training			
	10%	25%	50%	75%
Accuracy of non-scar pixels	86.94 ±1.37	91.96 ±0.74	95.46 ±0.72	96.66 ±0.73
Accuracy of bright and dark scar pixels	87.54 ±1.66	94.03 ±0.85	97.47 ±0.68	98.54 ±0.55
Overall Accuracy	87.23 ±0.54	92.96 ±0.50	96.43 ±0.51	97.57 ±0.42

From the tables, it is clear that the proposed algorithm can achieve greater than 90% accuracy with the first and second sets, when a sufficient amount of training pixels (≥25%) were used. However, there were still some misclassified pixels, especially when we added the dark scar pixels to the data set. This issue will be discussed in subsection 4.4.

4.3.3 VISUALIZATION FOR THE INITIAL CLASSIFICATION

In this section, a qualitative analysis of the classification results over the whole image is presented. The goal of this analysis is to test and visualize the whole image (contains 720 by 820 pixels) after training various amounts of the ground truth data. Figure 34 and 35 show the classification maps from the second step of the proposed algorithm on the first and the second data sets, respectively.

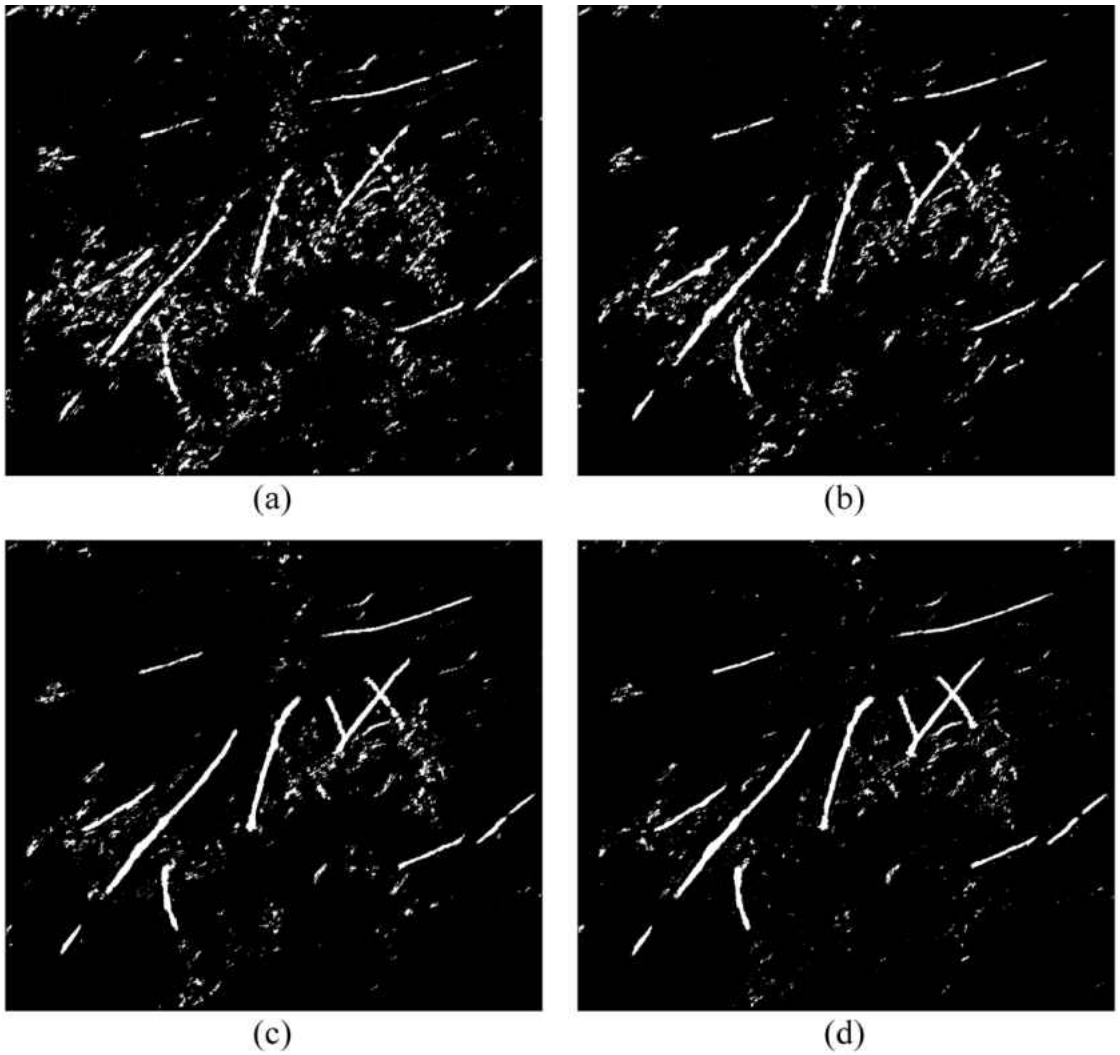


Figure 34: Classification maps of test image after the classification step. The maps were obtained with training (a) 10%, (b) 25%, (c) 50% and (d) 75% of the first ground truth set.

It is observed from Figure 34 that most of the scar pixels were identified correctly, especially when we used more than 25% of the data for training. There are some false positives (bright non-scar pixels) such as rocks or other structural elements in the sea that appear bright in the image. Notice that false positives were reduced when more data were used for training. For the second data set, Figure 35 shows the same trend except that the inclusion of dark scar pixels resulted in more false positives.

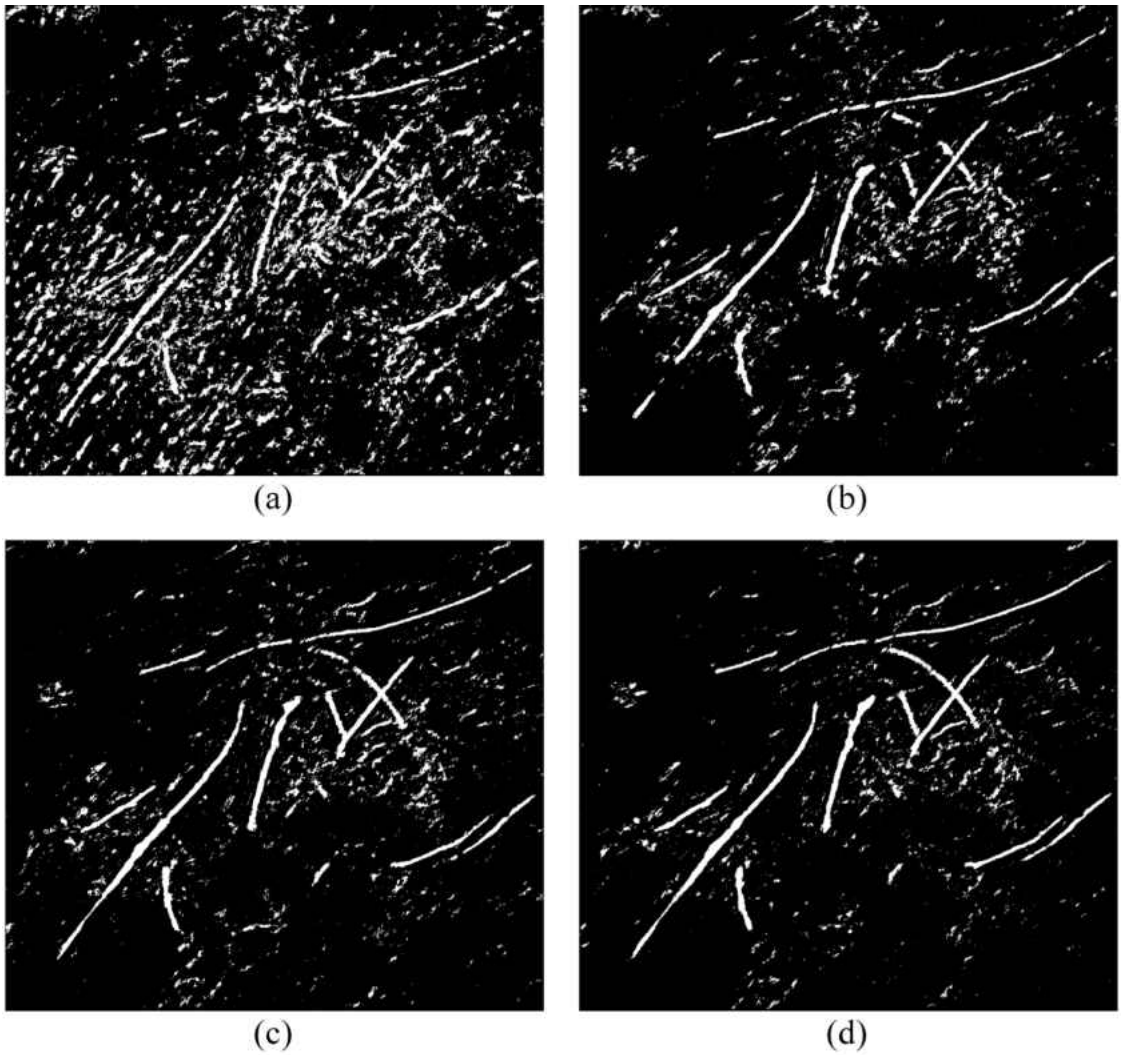


Figure 35: Classification maps of test image after the classification step. The maps were obtained with training (a) 10%, (b) 25%, (c) 50% and (d) 75% of the second ground truth set.

4.3.4 FALSE POSITIVE REDUCTION BASED ON PIXEL'S LOCAL ORIENTATION

To remove the unwanted objects, local orientations of scar pixel candidates were computed using the "opening by reconstruction" filter. The filter was tested with different SE sizes and the size of 35 was chosen in this dissertation. Figure 36 and 37 show the final scar maps obtained for the first and second sets, respectively, with various amounts of data used for training.

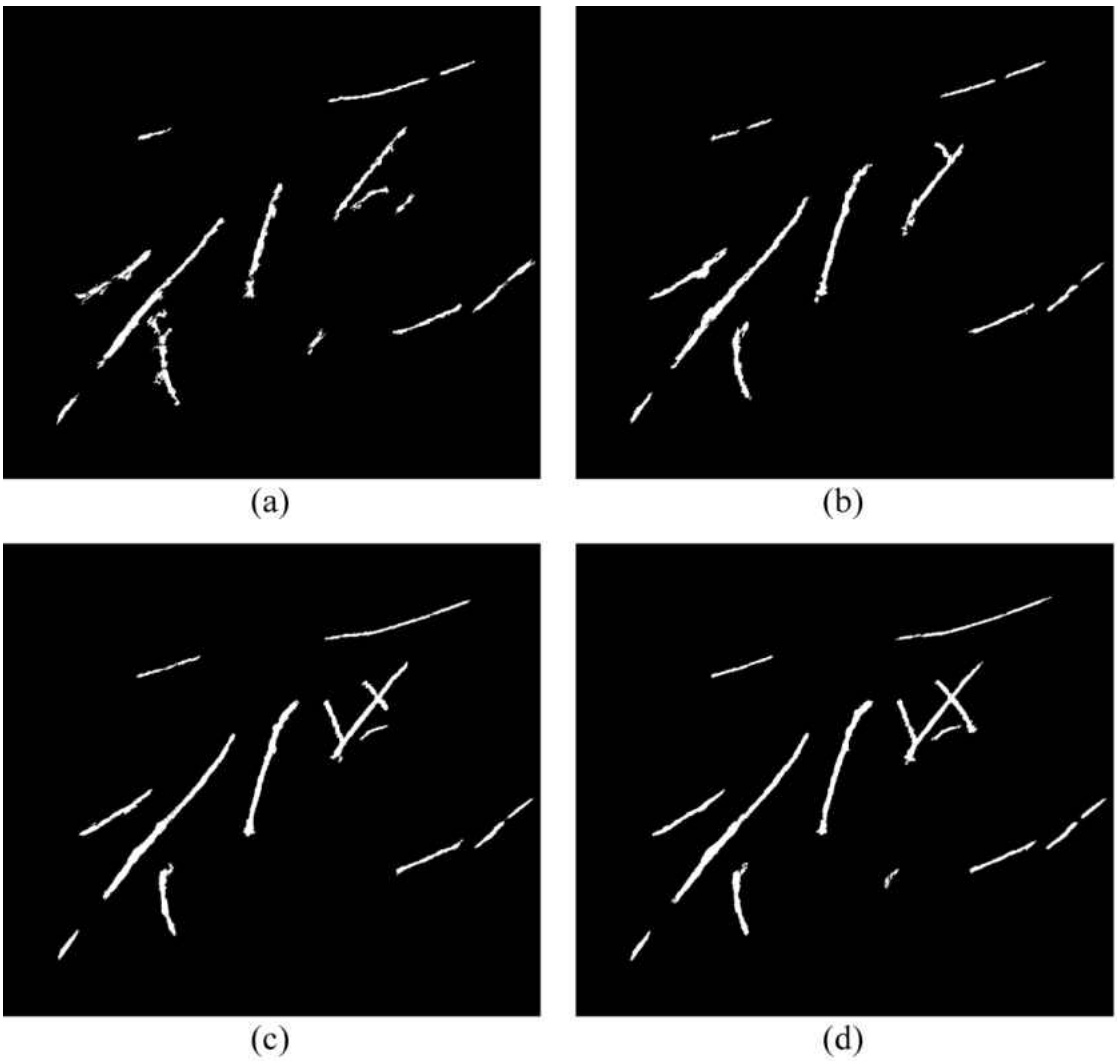


Figure 36: Final scar maps of test image after the local orientation step. The maps were obtained with training (a) 10%, (b) 25%, (c) 50% and (d) 75% of the first ground truth set.

It is observed from Figure 36 that most of the scar pixels were identified correctly, however, not all false positives were removed when a small amount of training pixels (10% and 25%) was used. Notice that all false positives were removed when more data were used for training. For the second data set, Figure 37 shows that training pixels at 25% of the data set is sufficient to remove almost all false positives. This amount also retains all of the scars, except for a small portion of the scar marked in red in Figure 37(b). We also observed that some small scars, which

were not trained in the classification step, were also detected as scars (marked in green in Figure 37).

In this study, we also calculated the numbers and the region occupied by the scars detected in the final scar map. From Table 18, when more data were used for training, we detected more scars. We ignored the results of 10% training data, since it contains too many false positives.

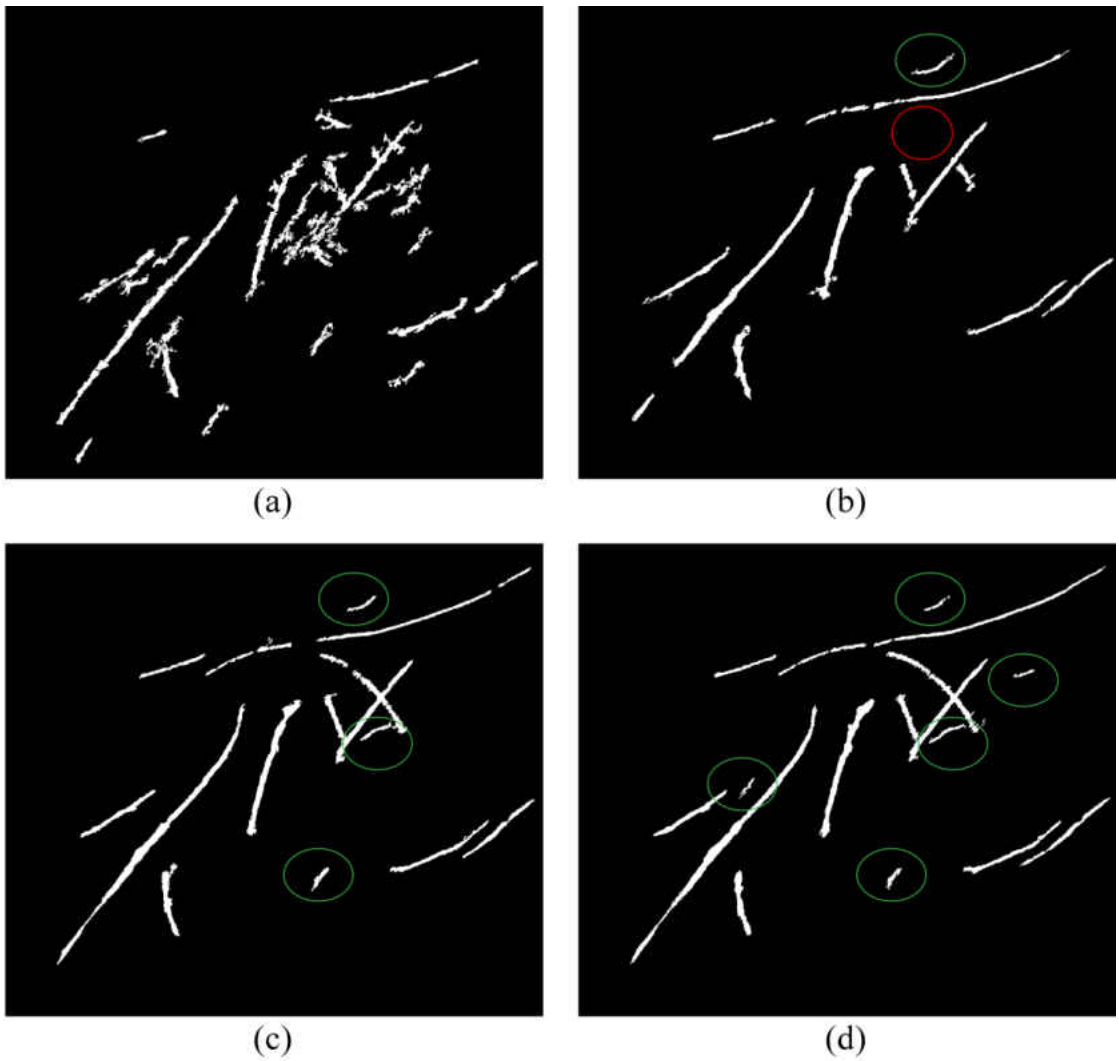


Figure 37: Final scar maps of test image after the local orientation step. The maps were obtained with training (a) 10%, (b) 25%, (c) 50% and (d) 75% of the second ground truth data set.

Table 18: Detected scars using different amounts of data for training (%).

	Rate (%) of data for training		
	25%	50%	75%
The numbers of detected scars	15	16	19
Area occupied by the detected scars in pixels	14,383	16,286	16,551
Area occupied by the detected scars in m^2 (1 pixel = 0.25 m^2)	3,595	4,071	4,138

4.4 Discussion and conclusion

Accurate detection of propeller scars on seagrass beds is challenging as described in Section 4.1. We proposed a three-step algorithm to address these challenges in this dissertation. The detection of the scars in the classification step depends on the number of basis functions, the size of patches, threshold value and number of neighbors in the weighted knn classification algorithm. We conducted experiments on two data sets to optimize these parameters and used various amounts of data samples for training. Experimental results showed that inclusion of dark-toned scar pixels (the second data set) will degrade the detection performance of the proposed system. For example, if 25% of the ground truth sets were used for training, the overall classification accuracy decreased from 94.90% (first set) to 92.96% (second set). This can also be confirmed in the classification maps (Figure 34-35). We believe the reason for this difference is that by adding the dark-toned scar pixels to the training set, there is a false detection of some non-scar regions as scar pixels. Another reason of for the presence false positive scar pixels in the initial scar map is caused by structural elements in the sea being represented by bright pixels.

From the experiments, we also observed that the accuracy of classification scar pixels with multispectral pan-sharpened bands is higher than that obtained by the original multispectral data.

In our previous study [121], we experimented the proposed algorithm on the panchromatic image of the same region to detect the scars. In that experiment, we obtained 91.80% overall accuracy by training 2073 samples. The experiment on the multispectral pan-sharpened bands achieved an overall accuracy of 94.90% with 1580 samples.

We utilized morphological directional profiles to eliminate false positives and generate the final scar detection map. The size of linear SE is an important parameter for the formation of the morphological directional profiles. We analyzed the effect of SE size on the classification map in our previous study [121]. Relatively small size of SE can still produce false results. Conversely, large sizes of SE cause some thin scars to be eliminated on the final scar detection map. For example, although thin and discontinuous parts of upper scar were detected as scars in classification step (Figure 35(b)), they were not mapped in final scar map (Figure 37(b)) since they were too short for the size of SE optimized in this paper. The SE size in the algorithm could be reduced to detect this part; however, this would lead more non-scar regions being mapped as scar regions. The size of the SE, therefore, depends on the purpose of the detection; a greater size will detect longer scars and miss some short scars, whereas a smaller size will detect most of the scars but also non-scars. Aside from this aspect, the local orientation step can successfully distinguish the scars from false positives. The optimized size of SE can retain the scars while eliminating unwanted small objects, especially with the training rate of more than 25%.

It is also interesting to note the results of the two sets. As seen from Table 16-17 and Figure 34-35, the amount of false positives in the experiment of the first set is much more than that of the second set before the local orientation step. However, the local orientation step can successfully eliminate false positives, thus, it is more reasonable to include dark-toned scar

pixels into the training set. This illustration also shows that local orientation is a very powerful tool to eliminate false positives from scar pixels.

We conclude that this promising algorithm could be implemented on a regular basis to monitor changes in habitat characteristics in coastal waters. It represents an important new capability for coastal management.

Chapter 5

CONCLUSION AND FUTURE WORK

This dissertation studied methods for feature representation of hyperspectral and multispectral images based on the sparse coding algorithm. Specifically, the dissertation addressed two challenge problems in sparse coding framework: the complex optimization problem and very high dimensional feature representation. A spectral based sparse coding framework was introduced for the representation of the features. We proved that a set of random sub-bands as a dictionary and the soft-thresholding function as an encoder are sufficient to obtain high performance on the classification of HSI data we tested instead of solving the optimization problems. For further improvement on the classification performance with a compact representation, we proposed two different ways: to incorporate the spatial information and to reduce the dimension of the new representation before the classification step. The enhanced framework, SC-DFR, was applied to three well known HSI data and was compared against the standard and recently developed methods. Experiments on real HSI data confirmed the effectiveness of the proposed model for HSI classification. From the classification experiments, the proposed method showed improvement over the other representation methods. It is also worth to note that the performance of our method is much better than the others in scenarios with a relatively low number of labeled training data. This property was very important since labeling HSI data is time-consuming and difficult.

In this dissertation, we also addressed the detection of seagrass scars in shallow waters. For this purpose, the SC-DFR method was modified for a pan-sharpening image and systematically

integrated with different image processing applications: pan-sharpening and morphological operations. Experiments showed that the proposed system can detect the scars that even have very similar pixel values as their surrounding structures. We believe that this fully automatic model provides valuable baseline information for assessment of scar recovery.

A number of directions for the remaining work are possible. The work completed has been performed using just the mean per spectral band of HSI pixel as the spatial information. Another spatial feature, such as standard deviation, or a smarter algorithm that makes excessive use of both spectral and spatial features of HSI would be appealing for more powerful representation systems. Another future research option may be the use of more smooth and differentiable functions, such as "double tanh" function [122] for thresholding. This function could discriminate very similar structures while creating sparse codes.

REFERENCES

[1] Tuia, D., Pacifici, F., Kanevski, M., Emery, W.J., "Classification of Very High Spatial Resolution Imagery Using Mathematical Morphology and Support Vector Machines," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 11, pp. 3866-3879, 2009.

[2] Palsson, F., Sveinsson, J.R., Benediktsson, J.A., Aanaes, H., "Classification of Pansharpened Urban Satellite Images," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal*, vol. 5, no. 1, pp. 281-297, 2012.

[3] Alparone L., Wald L., Chanussot J., Thomas C., Gamba P., Bruce L. M., "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp.3012–3021, 2007.

[4] H.Chen, C. Lin; S. Chen, C. Wen, C. Chen, Y. Ouyang, C. Chang, "PPI-SVM-Iterative FLDA Approach to Unsupervised Multispectral Image Classification," in *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* , vol. 6, no. 4, pp. 1834-1842, 2013.

[5] A. Gualtieri et al., "Support vector machine classifiers as applied to AVIRIS data," in *Summaries of the Eighth JPL Airborne Earth Science Workshop*, pp. 217–227, JPL Pub., 1999.

[6] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.

[7] H. Yang, “A back-propagation neural network for mineralogical mapping from AVIRIS data,” *Int. J. Remote Sens.*, vol. 20, no. 1, pp. 97-110, 1999.

[8] G. Camps-Valls, T. Bandos Marsheva and D. Zhou, "Semi-Supervised Graph-Based Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044-3054, 2007.

[9] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation”, *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311-4322, 2006.

[10] A. S. Charles, B. A. Olshausen and C. J. Rozell, “Learning sparse codes for hyperspectral imagery,” *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 5**,** no. 5, pp. 963–978, 2011.

[11] A. Castrodad et al., “Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.,* vol. 49, no. 11, pp. 4263-4281, 2011.

[12] Y. Chen, N.M. Nasrabadi, T.D. Tran, "Hyperspectral Image Classification Using Dictionary-Based Sparse Representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973-3985, 2011.

[13] A.Soltani-Farani, H.R.Rabiee, S.A.Hosseini, "Spatial-Aware Dictionary Learning for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 527-541, 2015.

[14] M. D. Iordache, J. Bioucas-Dias and A. Plaza, “Sparse unmixing of hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014-2039, 2011.

[15] J. Yang et al., "Linear spatial pyramid matching using sparse coding for image classification," *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, 1794-1801, IEEE, Miami, 2009.

[16] A. Coates, A. Y. Ng and S. Mall, "The Importance of Encoding Versus Training with Sparse Coding and Vector Quantization," *Proceedings of the 28th International Conference on Machine Learning*, *ICML-11*, 921-928, ACM, New York, 2011.

[17] I. Gkioulekas and T. Zickler, "Dimensionality Reduction Using the Sparse Linear Model," *Advances in Neural Information Processing Systems 24, NIPS-2011,* pp. 271-279, 2011.

[18] U. Srinivas et al., "Exploiting Sparsity in Hyperspectral Image Classification via Graphical Models," *IEEE Geosci. and Remote Sens.Letters*, vol. 10, no. 3, pp. 505-509, 2013.

[19] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, 2013.

[20] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, pp. 297–301, 1965.

[21] D. J. Field, "What is the goal of sensory coding?," Neural Computation, vol. 6, no. 4, pp. 559–601, 1994.

[22] A. K. Jain, Fundamentals of digital image processing. Prentice-Hall, 1989.

[23] S. Mallat, A wavelet tour of signal processing, third ed. Academic Press, 2009.

[24] I. T. Jolliffe, Principal component analysis, second ed. Springer, New York, 2002.

[25] R. Rubinstein, A.M. Bruckstein, M. Elad, "Dictionaries for Sparse Representation Modeling," *in Proceedings of the IEEE*, vol.98, no.6, pp.1045-1057, 2010.

[26] J. B. Allen and L. R. Rabiner, “A unified approach to short-time Fourier analysis and synthesis,” *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, 1977.

[27] W. B. Pennebaker and J. L. Mitchell, JPEG still image data compression standard. Springer, New York, 1993.

[28] D. Gabor, “Theory of communication,” J. Inst. Electr. *Eng*, vol. 93, no. 26, pp. 429–457, 1946.

[29] J. Wexler and S. Raz, “Discrete gabor expansions,” Signal processing, vol. 21, no. 3, pp. 207–221, 1990.

[30] S. Qian and D. Chen, “Discrete gabor transform,” *IEEE Trans. Signal Process*., vol. 41, no. 7, pp. 2429 - 2438, 1993.

[31] J. G. Daugman, “Two-dimensional spectral analysis of cortical receptive field profiles.,” Vision *research*, vol. 20, no. 10, pp. 847–856, 1980.

[32] J. G. Daugman, “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters,” *J. Opt. Soc. Am. A*, vol. 2, no. 7, pp. 1160–1169, 1985.

[33] J. G. Daugman, “Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression,” *IEEE Trans. Acoust. Speech Sigal Process.*, vol. 36, no. 7, pp. 1169–1179, 1988.

[34] M. Porat and Y. Y. Zeevi, “The generalized Gabor scheme of image representation in biological and machine vision,” *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 10, no. 4, pp. 452–468, 1988.

[35] P. Burt and E. Adelson, “The Laplacian pyramid as a compact image code,” *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, 1983.

[36] I. Daubechies, Ten lectures on wavelets. Society for Industrial Mathematics, 1992.

[37] Y. Meyer and D. Salinger, Wavelets and operators. Cambridge University Press, 1995.

[38] J. Morlet and A. Grossman, “Decomposition of Hardy functions into square integrable wavelets of constant shape,” SIAM J. Math. Anal., vol. 15, pp. 723–736, 1984.

[39] Y. Meyer, “Principe dincertitude, bases hilbertiennes et alg`ebres dop´erateurs,” S´eminaire Bourbaki, no. 662, 1985-86.

[40] S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[41] S. Mallat and W. L. Hwang, “Singularity detection and processing with wavelets,” *IEEE Trans. Inf. Theo.*, vol. 38, no. 2, pp. 617–643, 1992.

[42] S. Mallat and S. Zhong, “Characterization of signals from multiscale edges,” *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 14, no. 7, pp. 710–732, 1992.

[43] R. R. Coifman, Y. Meyer, and V. Wickerhauser, “Wavelet analysis and signal processing,” Wavelets and their Applications, pp. 153–178, 1992.

[44] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transforms,” *IEEE Trans. Inf. Theory*, vol. 38, no. 2 part 2, pp. 587–607, 1992.

[45] S. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” IEEE Trans. Signal Process., vol. 41, no. 12, pp. 3397–3415, 1993.

[46] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” Technical Report – Statistics, Stanford, 1995.

[47] I. S. Dhillon and D. M. Modha. "Concept decompositions for large sparse text data using clustering." Machine Learning, vol. 42, no. 1, 2001.

[48] Y. C.Pati, R.Rezaifar, and P. S.Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition”, *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference*, vol. 1, pp. 40-44, IEEE, Pacific Grove, CA, 1993.

[49] B. A. Olshausen and D. J. Field. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images” *Nature*, vol. 381, no. 6583, pp.607-609, 1996.

[50] K. Jarrett et al., “What is the best multi-stage architecture for object recognition?” *IEEE 12th International Conference Computer Vision-2009*, pp. 2146-2153, IEEE, Kyoto, 2009.

[51] K. Kavukcuoglu, M. Ranzato and Y. LeCun, “Fast inference in sparse coding algorithms with applications to object recognition,” *CoRR*, abs/1010.3467, 2010.

[52] A. Krizhevsky, “Convolutional Deep Belief Networks on CIFAR-10”. Unpublished manuscript, 2010.

[53] B. A. Olshausen and D. J. Field, “Sparse coding with an overcomplete basis set: A strategy employed by V1?”, *Vision Research*, vol.37, p:3311–3325, 1997.

[54] H. Lee, A. Battle, R. Raina, and A. Y. Ng., “Efficient sparse coding algorithms”, NIPS, 2007.

[55] Y. Sohn, NS. Rebello, "Supervised and unsupervised spectral angle classifiers" *Photogram Eng Remote Sens*, vol. 68, no. 12, pp. 1271–1280, 2002.

[56] C.H. Chen, T.M. Tu, "Computation reduction of the maximum likelihood classifier using the winograd identity" *Pattern Recognit*, vol. 29, no. 7, pp.1213–1220, 1996.

[57] G.Camps-Valls et al., "Composite kernels for hyperspectral image classification," *IEEE Geosci. and Remote Sens.Letters*, vol. 3**,** no. 1, pp. 93-97 2006.

[58] T. M. Cover, P. E. Hart, "Nearest neighbor pattern classification", IEEE *Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.

[59] S. A. Dudani, "The distance-weighted k-nearest neighbor rule", IEEE *Transactions on System, Man, and Cybernetics*, vol. 6, pp. 325-327, 1976.

[60] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[61] C. Cortes, V.Vapnik, "Support-vector networks". *Machine Learning,* vol. 20, no. 3, pp. 273, 1995.

[62] C. M. Bishop, "*Neural Networks for Pattern Recognition*," Clarendon Press, Oxford, 1995.

[63] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection" *J. Mach. Learn. Res.* vol. 3, pp.1157-1182, 2003.

[64] J. Liu, S. Ji and J. Ye, "Multi-task feature learning via efficient L2,1-norm minimization", *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, *UAI '09*, pp. 339-348, AUAI Press, Montreal, 2009.

[65] I. Guyon et al., "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389-422, 2002.

[66] J. Liu , J. Chen and J. Ye , "Large-scale sparse logistic regression" , *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining*, pp. 547 -556 , 2009

[67] Tibshirani R., "Regression shrinkage and selection via the lasso". *J. R. Statist. Soc. Series B* , vol. 58, pp. 267–288, 1996.

[68] J. Liu, S. Ji, and J. Ye. "SLEP: Sparse Learning with Efficient Projections. Arizona State University, 2009", http://www.public.asu.edu/~jye02/ Software/SLEP.

[69] G. Yuan, C. Ho, C. Lin, "Recent Advances of Large-Scale Linear Classification" *Proceedings of the IEEE*, vol. 100, no. 9, pp. 2584-2603, 2012.

[70] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables", *Journal Of The Royal Statistical Society Series B*, vol. 68, no. 1, pp.49–67, 2006.

[71] P. Zhao, G. Rocha, and B. Yu. The composite absolute penalties family for grouped and hierarchical variable selection. *Annals of Statistics*, vol. 37, no. 6A, pp. 3468–3497, 2009.

[72] C. Chang, *Hyperspectral Data Exploitation: Theory and Applications*, Wiley, New York 2007.

[73] G. Jun and J. Ghosh, "Spatially adaptive classification of land cover with remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2662-2673 2011.

[74] R. M. Haralick and K. D. L. Shanmugam, “Textural features for image classification,” *IEEE Trans. Systems Man and Cybernetics*, vol. 3, no. 6, pp. 610-621, 1973.

[75] Y. Tarabalka et al, "Multiple Spectral–Spatial Classification Approach for Hyperspectral Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122-4132, 2010.

[76] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, “Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, 2009.

[77] J. Li, J. M. Bioucas-Dias, A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields," *IEEE Trans. Geosci. Remote Sens.*" vol. 50, no. 3, pp. 809-823, 2011.

[78] S. Geman and D. Geman, “Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. and Machine Intell*, vol. 6**,** no. 6, pp. 721-741, 1982.

[79] H. Derin and H. Elliott, “Modeling and segmentation of noisy and textured images using gibbs random fields,” *IEEE Trans. Pattern Anal. and Machine Intell*, vol. 9, no. 1, pp. 39–55, 1987.

[80] Z. He, Q. Wang, Y. Shen, M. Sun, "Kernel Sparse Multitask Learning for Hyperspectral Image Classification With Empirical Mode Decomposition and Morphological Wavelet-Based Features," *IEEE Trans. Geosci. Remote Sens.*, vol. 52**,** no. 8, pp. 5150-5163, 2013.

[81] L. Bruzzone, L.Carlin, "A multilevel context-based system for clasification of very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587-2600, 2006.

[82] M. Fauvel et al., "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804-3814, 2008.

[83] A. Plaza et al., "Advanced Processing of Hyperspectral Images," *IEEE International Conference Geoscience and Remote Sensing Symposium 2006*, *IGARSS 2006*, pp. 1974-1978, IEEE, Denver, 2006.

[84] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, “Classification of hyperspectral data from urban areas based on extended morphological profiles,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480-491, 2005.

[85] M. Fauvel, J. Chanussot, J.A.Benediktsson, "A spatial-spectral kernel- based approach for the classication of remote-sensing image," *Pattern Recognition*, vol. 45, no. 1, pp. 381-392, 2012.

[86] S. Velasco-Forero and J. Angulo, "Classification of hyperspectral images by tensor modeling and additive morphological decomposition," *Pattern Recognition*, vol. 46, no. 2, pp. 566-577, 2013.

[87] Q. Sami ul Haq, L. Tao, F. Sun, and S. Yang, "A fast and robust sparse approach for hyperspectral data classification using a few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2287–2302, 2012.

[88] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no.4, pp. 2276-2291, 2013.

[89] J.A.Tropp, A.C. Gilbert, and M.J. Strauss, "Simultaneous sparse approximation via greedy pursuit", *IEEE International Conference Acoustics, Speech, and Signal Processing*, *2005, ICASSP '05*, vol. 5, pp. 721-724, Philadelphia , 2005.

[90] E. Oguslu, K. Iftekharuddin and J. Li, "Sparse coding for hyperspectral images using random dictionary and soft thresholding", *Proc. SPIE 8399, Visual Information Processing XXI 83990A*, *SPIE Proceedings*, 8399, Baltimore, 2012.

[91] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411-430, 2000.

[92] A. Coates, L. Honlak and A. Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," *14th International Conference on AI and Statistics*, pp.215-223, 2011.

[93] A. Coates and A.Y. Ng, "Learning Feature Representations with K-means", *Chapter Neural Networks: Tricks of the Trade, Springer Lecture Notes in Computer Science*, 7700, pp. 561-580, Springer, 2012.

[94] Y-Lan Boureau, Nicolas Le Roux, Francis Bach, Jean Ponce, and Yann LeCun, "Ask the locals: multi-way local pooling for image recognition", *Proc. International Conference on Computer Vision (ICCV'11),* 2011.

[95] E. Oguslu, K. Iftekharuddin and J. Li, "Sparse coding for hyperspectral images using random dictionary and soft thresholding", *Proc. SPIE 8399, Visual Information Processing XXI 83990A*, *SPIE Proceedings*, 8399, Baltimore, 2012.

[96] E. Oguslu, G. Zhou and J. Li, "Hyperspectral image classification using a spectral-spatial sparse coding model," *Proc. SPIE 8892, Image and Signal Processing for Remote Sensing XIX, 88920R*, Dresden, Germany, 2013.

[97] Wu, T.T., and Lange, K., "Coordinate descent algorithms for lasso penalized regression," *Annals of Applied Statistics* vol. 2, no.1, 2008.

[98] L. Ma, M. M. Crawford and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 48, no.11, pp. 4099-4109, 2010.

[99] Ham, J., Yangchi, C., Crawford, M.M., Ghosh, J., "Investigation of the random forest framework for classification of hyperspectral data," *Geoscience and Remote Sensing IEEE Transactions* vol. 43, no.3, pp. 492-501, 2005.

[100] Baumgardner, M., Biehl, L., Landgrebe, D. "220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3". Purdue University Research Repository, (accessed 22 February 2016), [https://purr.purdue.edu/publications/1947]

[101] A. Plaza *et al.*, “Recent advances in techniques for hyperspectral image processing,” *Remote Sens. Environ.*, vol. 113, (S1), S110-S122, 2009.

[102] Xudong Kang; Shutao Li; Benediktsson, J.A., "Feature Extraction of Hyperspectral Images With Image Fusion and Recursive Filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742-3752, 2014.

[103] J. A. Richards, *Remote Sensing Digital Image Analysis*. Springer-Verlag, New York, NY, 2012.

[104] C. Chih-Chung and L. Chih-Jen, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2**,** no. 3, pp. 27:1-27:27, 2011.

[105] E. S. L. Gastal andM.M. Oliveira, “Domain transform for edge-aware image and video processing,” *ACM Trans. Graph.*, vol. 30**,** no. 4, pp. 69:1-69:12, 2011.

[106] Digital Media Laboratory, “Spatial-aware dictionary learning,” 18 July 2011, http://ssp.dml.ir/research/sadl (17 November 2015).

[107] X. Kang, “Xudong Kang’s homepage,” 11 January 2013, http://xudongkang.weebly.com (17 November 2015).

[108] Menon, V. Prasad, S. Fowler, J.E., "Hyperspectral classification using a composite kernel driven by nearest-neighbor spatial features," in *Image Processing (ICIP), 2015 IEEE International Conference on* , vol., no., pp. 2100-2104, 2015

[109] Larkum, A., R. Orth, and C. Duarte, Seagrasses: Biology, Ecology and Conservation. Springer, Dordrecht, 2006.

[110] Orth, R., T. Carruthers, W. Dennison, C. Duarte, J. Fourqurean, K. Heck, A. Hughes, G. Kendrick, W. Kenworthy, S. Olyarnik, F. Short, M. Waycott, and S. Williams., "A global crisis for seagrass ecosystems." *BioScience,* vol. 56, pp. 987-996, 2006.

[111] Madley K., Krolick J., Sargent B. "Assessment of Boat Propeller Scar Damage within the Greater Charlotte Harbor Region." *Fish and Wildlife Research Institute*, St. Petersburg, FL, 2004

[112] Green K., Lopez C., "Using Object-Oriented Classification of ADS40 Data to Map the Benthic Habitats of the State of Texas," *Photogrammetric Engineering and Remote Sensing,* vol. 73, no. 8, pp. 861-864, 2007.

[113] Green K., Tukman M., Finkbeiner M., "Comparison of DMC, UltraCam, and ADS40 Imagery for Benthic Habitat and Propeller Scar Mapping," *Photogrammetric Engineering & Remote Sensing*, pp.589-599, 2011.

[114] S.W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, Q. Weng, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery", *Remote Sensing of Environment*, vol. 115, no. 5, pp. 1145-1161, 2011.

[115] M. Fauvel, B. Arbelot, J.A. Benediktsson, D. Sheeren and J. Chanussot, "Detection of Hedges in a Rural Landscape Using a Local Orientation Feature: From Linear Opening to Path Opening", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 6, no. 1, pp. 15-26, 2013.

[116] P. Soille, Morphological Image Analysis: Principles and Applications, Second Edition, Springer, Berlin, 2003.

[117] P. Soille and H. Talbot, “Directional morphological filtering,” *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 23, no.11, pp. 1313 -1329, 2001.

[118] V. K. Shettigara, “A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set,” *Photogramm. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[119] P. Soille and M. Pesaresi, “Advances in mathematical morphology applied to geoscience and remote sensing,” *IEEE Trans. Geosci. Remote Sens*., vol. 40, no.9, pp. 2042–2055, 2002.

[120] M. Pesaresi and J. A. Benediktsson, “A new approach for the morphological segmentation of high-resolution satellite imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, 2001.

[121] E. Oguslu, S. Erkanli, V. J. Hill, V. J. Hill, W. P. Bissett, R. C. Zimmerman and J. Li, "Detection of seagrass scars using sparse coding and morphological filter", *Proceedings of SPIE*, vol. 9240, 92400G, 2014.

[122] K. Gregor, Y.LeCun, "Learning fast approximations of sparse coding.", *In: International Conference on Machine Learning*, ICML, vol. 7, 2010.

VITA

Ender OGUSLU

Department of Electrical and Computer Engineering

Old Dominion University

Norfolk, VA 23529

Ender Oguslu received his B.S. degree in Electronics Engineering from Turkish Air Force Academy, Istanbul, Turkey, in 1995, and his M.S. degree in Electrical and Computer Engineering from Old Dominion University, Norfolk, USA, in 2006. In 2010, he began to continue his education at Old Dominion University in pursuit of a Ph.D. Degree in Electrical and Computer Engineering. He was supported by Turkish Air Force during his Ph.D. work.

LIST OF PUBLICATIONS

DRAFT MANUSCRIPT

E. Oguslu, V. J. Hill, W. P. Bissett, R. C. Zimmerman and J. Li, "Detection of seagrass scars using sparse coding and morphological filter", (2016).

JOURNAL ARTICLE

E. Oguslu; G. Zhou; Z. Zheng; K. Iftekharuddin and J. Li, "Sparse coding based dense feature representation model for hyperspectral image classification", J. Electron. Imaging. 24(6), 063009, 2015.

CONFERENCE PUBLICATIONS

E. Oguslu, K. Iftekharuddin and J. Li, "Sparse coding for hyperspectral images using random dictionary and soft thresholding", *Proc. SPIE 8399, Visual Information Processing XXI 83990A*, *SPIE Proceedings*, 8399, Baltimore, (2012).

E. Oguslu, G. Zhou and J. Li, “Hyperspectral image classification using a spectral-spatial sparse coding model,” *Proc. SPIE 8892, Image and Signal Processing for Remote Sensing XIX, 88920R*, Dresden, Germany, 2013.

E. Oguslu, S. Erkanli, V. J. Hill, V. J. Hill, W. P. Bissett, R. C. Zimmerman and J. Li, "Detection of seagrass scars using sparse coding and morphological filter", Proceedings of SPIE Vol. 9240, 92400G (2014).