1-1-2002

Linear Unmixing of Hyperspectral Signals via Wavelet Feature Extraction

Jiang Li

LINEAR UNMIXING OF HYPERSPECTRAL SIGNALS VIA WAVELET FEATURE EXTRACTION

By

Jiang Li

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Electrical Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2002

LINEAR UNMIXING OF HYPERSPECTRAL SIGNALS VIA

WAVELET FEATURE EXTRACTION

By

Jiang Li

Approved:

Dr. Lori M. Bruce Assistant Professor of Electrical and Computer Engineering (Major Professor and Director of Dissertation)	Dr. James E. Fowler Associate Professor of Electrical and Computer Engineering (Committee Member)
Dr. Roger L. King Professor of Electrical and Computer Engineering (Committee Member)	Dr. Nicholas H. Younan Professor of Electrical and Computer Engineering (Committee Member and Graduate Coordinator)
Dr. James C. Harden Acting Department Head of Electrical and Computer Engineering	Dr. A. Wayne Bennett Dean of the College of Engineering

Name: Jiang Li

Date of Degree: December 13, 2002

Institution: Mississippi State University

Major Field: Electrical Engineering

Major Professor: Dr. Lori Mann Bruce

Title of Study: LINEAR UNMIXING OF HYPERSPECTRAL SIGNALS VIA WAVELET FEATURE EXTRACTION

Pages in Study: 152

Candidate for Degree of Doctor of Philosophy

A pixel in remotely sensed hyperspectral imagery is typically a mixture of multiple electromagnetic radiances from various ground cover materials. Spectral unmixing is a quantitative analysis procedure used to recognize constituent ground cover materials (or endmembers) and obtain their mixing proportions (or abundances) from a mixed pixel. The abundances are typically estimated using the least squares estimation (LSE) method based on the linear mixture model (LMM).

This dissertation provides a complete investigation on how the use of appropriate features can improve the LSE of endmember abundances using remotely sensed hyperspectral signals. The dissertation shows how features based on signal classification approaches, such as discrete wavelet transform (DWT), outperform features based on conventional signal representation methods for dimensionality reduction, such as

principal component analysis (PCA), for the LSE of endmember abundances. Both experimental and theoretical analyses are reported in the dissertation.

A DWT-based linear unmixing system is designed specially for the abundance estimation. The system utilizes the DWT as a pre-processing step for the feature extraction. Based on DWT-based features, the system utilizes the constrained LSE for the abundance estimation. Experimental results show that the use of DWT-based features reduces the abundance estimation deviation by 30-50% on average, as compared to the use of original hyperspectral signals or conventional PCA-based features.

Based on the LMM and the LSE method, a series of theoretical analyses are derived to reveal the fundamental reasons why the use of the appropriate features, such as DWT-based features, can improve the LSE of endmember abundances. Under reasonable assumptions, the dissertation derives a generalized mathematical relationship between the abundance estimation error and the endmember separabilty. It is proven that the abundance estimation error can be reduced through increasing the endmember separability. The use of DWT-based features provides a potential to increase the endmember separability, and consequently improves the LSE of endmember abundances.

The stability of the LSE of endmember abundances is also analyzed using the concept of the condition number. Analysis results show that the use of DWT-based features not only improves the LSE of endmember abundances, but also improves the LSE stability.

DEDICATION

To my wife and parents

ACKNOWLEDGMENTS

I would like to thank my major advisor, Dr. Lori Bruce, for her advising, support and help that finally lead to the successful completion of this dissertation. I would also like to thank my committee members, Dr. James Fowler, Dr. Roger King, and Dr. Nicholas Younan, for their encouragement and support.

I am grateful for the study and research environment provided by Department of Electrical and Computer Engineering (ECE), Engineering Research Centers (ERC), and Remote Sensing Technologies Center (RSTC), at Mississippi State University. I acknowledge all of the faculty and staff, working in the ECE department, the ERC and the RSTC, for their help and support.

Finally, but not least, I would like to thank my full family, specially my wife and parents, for their caring and love.

TABLE OF CONTENTS

Page

DEDICATION ii

ACKNOWLEDGMENTS iii

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER

I. INTRODUCTION 1

1.1. Study Background of The Dissertation 1
1.2. Motivation and Concerns of The Dissertation 7
1.3. Organization of The Dissertation 17
1.4. Contributions of The Dissertation 18

II. LEAST SQUARES ANALYSIS IN LINEAR UNMIXING OF HYPERSPECTRAL SIGNALS 20

2.1. Least Squares Estimation of Abundances 21
2.2. Error Analysis of Abundance Estimation 24
2.3. Simplification of Abundance Error Analysis Results 29
2.4. Stability Analysis of Abundance Estimation 33

III. FEATURE EXTRACTION USING DISCRETE WAVELET TRANSFORM FOR LINEAR UNMIXING OF HYPERSPECTRAL SIGNALS 37

3.1. Background on Wavelet Transform 39
3.2. DWT-Based Feature Extraction for Linear Unmixing of Hyperspectral Signals 49

CHAPTER Page

IV. A DISCRETE WAVELET TRANSFORM BASED LINEAR UNMIXING SYSTEM FOR LEAST SQUARES ESTIMATION OF ABUNDANCES 55
4.1. Comparison of Various Pre-processing Methods 57
4.2. Constrained Least Squares Estimation 64
4.3. System Performance Evaluation 66

V. EXPERIMENTS, RESULTS AND DISCUSSION 70

5.1. Preparation of Experiment Data 71
5.2. Abundance Estimation Results for Experiment I: Soybean versus Soil ... 79
5.2.1. System Training Results and Feature Selection for Experiment I 79
5.2.2. System Testing Results and Performance Evaluation for Experiment I 82
5.3. Abundance Estimation Results for Experiment II: Soybean versus Grass 86
5.3.1. System Training Results and Feature Selection for Experiment II 87
5.3.2. System Testing Results and Performance Evaluation for Experiment II 88
5.4. Abundance Estimation Results for Experiment III: Soybean, Grass versus Soil 92
5.4.1. System Training Results and Feature Selection for Experiment III 92
5.4.2. System Testing Results and Performance Evaluation for Experiment III 94
5.5. Discussion about Linearity of Features 101
5.6. Discussion about Stability of Least Squares Solution 107
5.7. Discussion about Assumptions in Mathematical Derivations of Abundance Estimation 110
5.8. Discussion about Alternative DCT and PCA Features 113

VI. CONCLUSIONS 123

APPENDIX

A. PROOF OF EXISTENCE OF $(A^T A)^{-1}$ IN EQUATION (2.7) 131

B. DERIVATION OF EQUATION (2.14) 133

C. DERIVATION OF EQUATIONS (2.22) AND (2.27) 135

D. DERIVATION OF EQUATION (2.25) 139

APPENDIX Page

E. DERIVATION OF EQUATION (2.30) 141

F. DERIVATION OF EQUATIONS (2.39) AND (2.40) 143

BIBLIOGRAPHY 146

LIST OF TABLES

TABLE Page

5.1. Information summary of the three sets of experiments 71

5.2. RMSE of abundances estimation based on training data set for feature selection in Experiment I 80

5.3. RMSE of abundances estimation based on training data set for feature selection in Experiment II 87

5.4. RMSE of abundances estimation based on training data set for feature selection in Experiment III 93

5.5. Experimental values of parameters in the theoretical analysis of abundance estimation, based on testing data set, for Experiment I 108

5.6. Experimental values of parameters in the theoretical analysis of abundance estimation, based on testing data set, for Experiment II 108

5.7. Experimental values of parameters in the theoretical analysis of abundance estimation, based on testing data set, for Experiment III 109

5.8. Optimal window parameters when using the sliding window method for DCT- and PCA-based feature extraction in Experiment I, II and III 119

LIST OF FIGURES

FIGURE Page

1.1. (a) HYDICE image; and (b) hyperspectral electromagnetic radiance signals of HYDICE 5

3.1. The dyadic (2-channel) forward and inverse FWT 42

3.2. The *M*-channel forward FWT 44

3.3. An example of the dyadic (2-channel) forward FWT 47

3.4. The reconstruction of wavelet coefficients in Figure 3.3. 48

4.1. A DWT-based linear unmixing system block diagram 56

4.2. An example of the DCT of a hyperspectral signal 59

4.3. An example of the PCA of a hyperspectral signal 63

5.1. An example of ASD's hyperspectral reflectance signal 73

5.2. The ASD's hyperspectral reflectance signals of soybean, grass and soil utilized for the linear unmixing analysis in the dissertation 74

5.3. Endmember spectra of soybean, grass and soil utilized for the linear unmixing analysis in the dissertation 78

5.4. RMSE of abundance estimation based on testing data set in Experiment I 83

5.5. Confidence of abundance estimation based on testing data set in Experiment I 85

5.6. Abundance distribution diagram based on testing data set in Experiment I 86

5.7. RMSE of abundance estimation based on testing data set in Experiment II 90

FIGURE Page

5.8. Confidence of abundance estimation based on testing data set in Experiment II 90

5.9. Abundance distribution diagram based on testing data set in Experiment II ... 91

5.10. RMSE of abundance estimation based on testing data set in Experiment III 95

5.11. Confidence of abundance estimation based on testing data set in Experiment III 96

5.12. Abundance distribution diagrams for soybean based on testing data set in Experiment III 98

5.13. Abundance distribution diagrams for grass based on testing data set in Experiment III 99

5.14. Abundance distribution diagrams for soil based on testing data set in Experiment III 100

5.15. Abundance distribution diagrams for soybean, based on testing data set, in Experiment I when using all DWT, PCA, or DCT coefficients as features 102

5.16. RMSE of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment I 104

5.17. Confidence of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment I 104

5.18. RMSE of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment II 105

5.19. Confidence of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment II 105

FIGURE Page

5.20. RMSE of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment III 106

5.21. Confidence of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment III 106

5.22. RMSE of abundance estimation based on training data set when using the sliding window method for DCT-based feature extraction in Experiment I 116

5.23. RMSE of abundance estimation based on training data set when using the sliding window method for PCA-based feature extraction in Experiment I 116

5.24. RMSE of abundance estimation based on training data set when using the sliding window method for DCT-based feature extraction in Experiment II 117

5.25. RMSE of abundance estimation based on training data set when using the sliding window method for PCA-based feature extraction in Experiment II 117

5.26. RMSE of abundance estimation based on training data set when using the sliding window method for DCT-based feature extraction in Experiment III 118

5.27. RMSE of abundance estimation based on training data set when using the sliding window method for PCA-based feature extraction in Experiment III 118

5.28. RMSE of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment I 120

5.29. Confidence of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment I 120

5.30. RMSE of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment II 121

FIGURE Page

5.31. Confidence of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment II 121

5.32. RMSE of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment III 122

5.33. Confidence of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment III 122

CHAPTER I

INTRODUCTION

1.1. Study Background of The Dissertation

One of the important applications of the remote sensing technology is target detection and classification. The airborne and spaceborne remote sensors allow us to rapidly acquire the large area information of the Earth's surface with a relatively low cost. This great advantage makes it become reality to implement the target detection and classification in a large ground cover range. The early remote sensing applications focused on the use of spatial information, *i.e.*, imagery, because it was the easiest and the most direct information for people to utilize. However, later researchers realized the substantial limitations of using only spatial information, particularly the limitation of spatial resolution. Using spatial information for target detection and classification is to investigate spatial relationships among objects in images. In order to identify an object on the ground, a remote sensing image has to have enough high spatial resolution, which means a remote sensor has to have enough high spatial resolution. For example, to recognize a building on the ground directly using a remote sensing image, it would require a remote sensor with a spatial resolution on the order of one meter. To recognize certain vegetation in agriculture applications, it would require a spatial resolution on the order of one centimeter. Unfortunately, however, the improvement of the spatial resolution of a remote sensor could be one of the most expensive factors in remote

sensing, due to the fact that it is very expensive to manufacture and maintain a remote sensor with a very high spatial resolution. As a result, remote sensing data acquired by such a high spatial resolution sensor could be too expensive to most users. Moreover, the amount of data could be too huge to be manipulated easily, because for the same size of ground cover area, higher spatial resolution data corresponds to a larger image size. These inevitable problems contradict the great advantages of economy and efficiency using the remote sensing technology. Hence, only utilizing the spatial information acquired by remote sensors severely restricts the extensive applications of the remote sensing technology.

Realizing the fact that optical remote sensing signals are typically electromagnetic radiance spectra of materials on the surface of the Earth received by remote sensors, researchers began to concentrate in using spectral information in remote sensing signals, instead of only using spatial information, for target detection and classification. Depending upon the electromagnetic energy sources, the remote sensing is typically divided into two broad categories: passive and active remote sensing. A passive remote sensing is such that the electromagnetic energy stems from a natural source such as the Sun. Many typical remote sensor systems, such as NASA's Landsat Multispectral Scanner System (MSS) and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) system, fall into the category of the passive remote sensing. An active remote sensing is such that a remote sensor itself provides the electromagnetic energy sources, instead of using the natural source. The synthetic aperture radar (SAR) system is an

example of the active remote sensing [1, 2]. For this study, we are focusing on the passive remote sensing.

Typically, the number of spectral bands of remote sensors determines the amount of spectral information that remote sensors can acquire. Early remote sensors only have several spectral bands, and thus limited spectral information can be obtained from such remote sensors. These remote sensors are typically called multispectral sensors. For example, the first Landsat MSS launched in July 1972 had 4 spectral bands with 80m spatial resolution and a 6-bit data system. The current Landsat 7 system launched in April 1999 has 7 spectral bands with 30m spatial resolution and an 8-bit data system. The French Système pour l'Observation de la Terre (SPOT) Haute Résolution Visible (HRV) system launched in February 1986, January 1990 and September 1993 had 3 spectral bands with 20m spatial resolution and an 8-bit data system, and 1 panchromatic band with 10m resolution and an 8-bit data system [1, 2]. With the ongoing development of the remote sensor technology, currently many remote sensors can operate with more than one hundred spectral bands. These remote sensors are typically called hyperspectral sensors. For example, the airborne Hyperspectral Digital Image Collection Experiment (HYDICE) system had 210 spectral bands with wavelengths from 400 to 2500nm. The AVIRIS system had 224 spectral bands with wavelengths from 400 to 2500nm. The spaceborne Hyperion system recently launched in November 2000 had 220 spectral bands with wavelengths from 400 to 2500nm and 30m spatial resolution. Compared to multispectral signals, hyperspectral remote sensing signals include much more spectral information, and are able to measure more detailed electromagnetic radiance characteristics of

materials. Thus, using hyperspectral signals provide a great potential for target detection and classification using the remote sensing technology.

Using spectral information for target detection and classification is to investigate electromagnetic radiance characteristics of materials. It is possible because of the fact that different materials have their own electromagnetic reflectance characteristics. That is, target materials can be identified via analyzing their spectral characteristics. Figure 1.1 shows an example of HYDICE image and corresponding hyperspectral curves. The HYDICE data were provided courtesy of the Spectral Information Technology Applications Center (SITAC) for our previous research [3-6], funded by the Hyperspectral Algorithms Research Center (HyMARC) at the Department of Energy (DOE) Remote Sensing Lab in Nevada. The image shown in Figure 1.1(a) is a pseudo-color RGB image. Images from three spectral bands (band 60, band 35 and band 15, lying in the near infrared, red and green regions, respectively) are used as red (R), green (G) and blue (B) channels to form the pseudo-color image. For a pixel in the HYDICE image, a corresponding hyperspectral curve can be formed by connecting samples of the pixel at each spectral band. That is, HYDICE system has 210 spectral bands and the hyperspectral curve of the HYDICE consists of 210 samples. Figure 1.1(b) shows four different hyperspectral curves of the HYDICE. These curves represent electromagnetic radiance of four different ground cover materials: car, tree, roof and road. It can be seen that these materials have their own electromagnetic reflectance characteristics. When using the spectral information for target detection and classification, it is not necessary to have pixels with a high spatial resolution, because the recognition of targets is based on

the electromagnetic radiance characteristics of the target materials. Thus, it is possible to implement a target detection and classification system using remote sensing data with low spatial resolution, which greatly degrades the limitation requiring high spatial resolution sensors. It is this advantage of using spectral information that leads to the use of hyperspectral remote sensing technology for a large variety of target detection and classification applications.

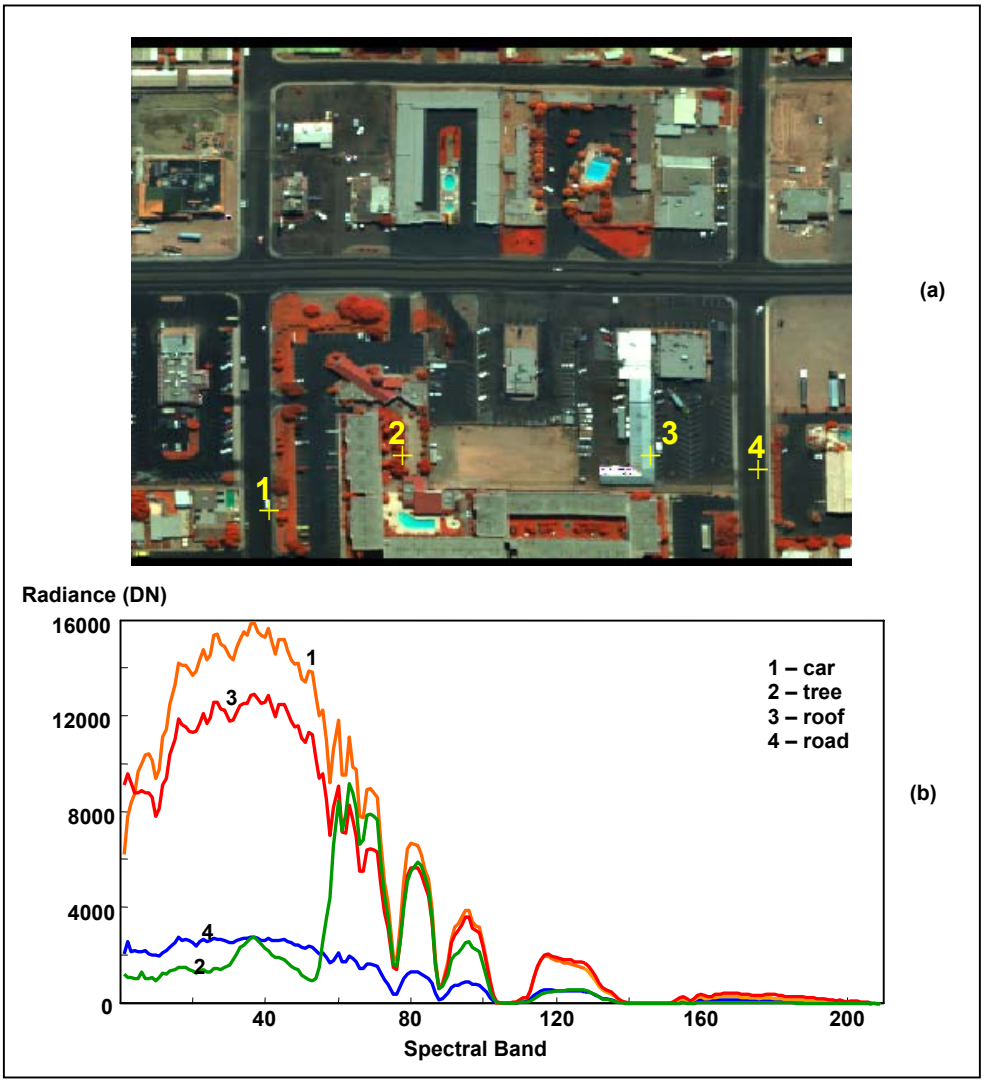

Figure 1.1. (a) HYDICE image; and (b) hyperspectral electromagnetic radiance signals of HYDICE.

Unfortunately, however, the problem of the limited spatial resolution still exists even when using spectral information for target detection and classification. In the target detection and classification problem, it is typically assumed that a pixel in a remotely sensed image represents a certain ground cover material, and each pixel can be uniquely assigned to a ground cover class. However, this assumption cannot be guaranteed because of the fact that a pixel in a remotely sensed image corresponds to a certain size of ground cover due to the limited spatial resolution of remote sensors. Consequently, it is possible that a pixel consists of several different ground cover materials. For example, Landsat 7 system has 30m spatial resolution, which means a pixel in the Landsat 7 image corresponds to a ground cover region of 30m by 30m. It is possible that the whole region of 30m by 30m is covered only by one certain type of ground material. However, for many practical cases, the region of 30m by 30m is inevitably covered by more than one type of material. That is, pixels in remotely sensed imagery are typically mixed pixels consisting of electromagnetic radiance spectra of several distinct ground cover materials. Note that no matter how much we improve the spatial resolution, most targets will result in a mixed pixel. This is due to the heterogeneous nature of most target surfaces. In other words, because of the heterogeneous surface, the electromagnetic radiance measured by remote sensors inevitably consists of radiance from multiple ground cover materials, which result in a mixed pixel. Therefore, question is how to implement the target detection and classification when a pixel's corresponding spectrum is composed of a mixture of multiple materials. It becomes necessary to investigate the sub-pixel

information, in order to achieve a better performance of target detection and classification in the mixed-pixel case. This leads to a research area referred to as spectral unmixing.

1.2. Motivation and Concerns of The Dissertation

Spectral unmixing is generally described as a quantitative analysis procedure used to recognize constituent ground cover materials (or endmembers) and obtain their mixing proportions (or abundances) from a mixed pixel. That is, the sub-pixel information of endmembers and their abundances can be obtained through the spectral unmixing process. Therefore, the target detection and classification can be implemented at a sub-pixel level. A simple example is the production of a herbicide map in the precision agriculture application, where pixels typically are a mixture of crops, weeds and soil. Using the spectral unmixing technique, mixing proportions of crops, weeds and soil in a certain area could be obtained. Based on the abundances of the weed endmember, a herbicide map can be produced.

The spectral unmixing problem has caused concerns and been extensively investigated for the past two decades. A general analysis approach for spectral unmixing is first to build a mathematical model of the spectral mixture. Then based on the mathematical model, certain techniques are applied to implement spectral unmixing. In general, mathematical models for spectral unmixing are divided into two broad categories: linear mixture model (LMM) and nonlinear mixture models (NLMM). The LMM assumes that each ground cover material only produces a single radiance, and the mixed spectrum is a linear combination of ground cover radiance spectra. The NLMM takes into account the multiple radiances of the ground cover materials, and thus the

mixture is no longer linear. The NLMM typically has a relatively more accurate simulation of physical phenomena [7], but the model is usually complicated and application dependent [8, 9]. For example, Mustard *et al.* analyzed the mixture of materials on the lunar surface using both the LMM and the NLMM [7], and showed that the NLMM produced a more accurate analysis. In [8], Mustard *et al.* implemented a quantitative analysis of mineral mixture spectra using a NLMM based on the bidirectional reflectance spectroscopy theory [10]. In [9], Borel *et al.* investigated the vegetation and soil surfaces using a NLMM based on the radiosity method [11]. In short, typically there is not a simple and generic NLMM that can be utilized in various spectral unmixing applications. This disadvantage of the NLMM greatly limits its extensive application.

In contrast, the LMM is simpler and more generic, and it has been proven successful in various remote sensing applications, such as geological applications [12], the forest studies [13-15], and the vegetation studies [16-18]. For example, Adams *et al.* utilized the LMM to determine the mineral types and abundances [12]. Using the LMM, Cross *et al.* estimated the proportions of forest covers in regions with small forest patches and convoluted clearance patterns [13], Gong *et al.* determined the forest species and canopy closure for forest ecological studies and forest management [14], and Hlavka *et al.* mapped forest clearcuts and monitored forest regrowth in certain areas [15]. Using the LMM, Smith *et al.* monitored vegetation covers in a large desert region [16], Quarmby *et al.* estimated crop areas at regional scales [17], and Garcia-Haro *et al.* estimated and monitored the amount of vegetation in the semi-arid area [18]. It is because of the

advantage of simplicity and generality that the LMM has become a dominant mathematical model for the spectral unmixing analysis.

Another major reason why the LMM has been broadly accepted for the spectral unmixing analysis is that the linear mixture assumption allows many mature mathematical skills and algorithms, such as least squares estimation (LSE) [19-21], to be easily applied to the spectral unmixing problem. As a matter of fact, the LSE method has been commonly accepted as an algorithm to solve the generalized linear spectral unmixing problem [12-18, 22-25]. Methods like orthogonal subspace projection (OSP) [26] have also been investigated for addressing the LMM [27, 28, 29]. While the OSP method is closely related to the LSE method, there exist fundamental differences. The LSE method results in the abundance estimation of all underlying endmembers, but the OSP method is used to estimate the presence of a single target endmember.

In general, the LMM is described using a linear equation,

$$\vec{y} = A\vec{x}_o + \vec{e}_o, \tag{1.1}$$

where

$$\vec{y} = [y_1, y_2, \cdots, y_N]^T \tag{1.2}$$

represents a mixed pixel spectrum;

$$\vec{x}_o = [x_{o1}, x_{o2}, \cdots, x_{oM}]^T \tag{1.3}$$

represents true abundances of endmembers;

$$\vec{e}_o = [e_{o1}, e_{o2}, \cdots, e_{oN}]^T \tag{1.4}$$

represents a random measurement error; and

$$A = [\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_M] \tag{1.5}$$

contains endmember spectra. Each column vector of matrix A represents an endmember spectrum,

$$\vec{a}_i = [a_{i1}, a_{i2}, \cdots, a_{iN}]^T, \text{ for } i = 1,2,\cdots,M \,. \tag{1.6}$$

In equations (1.1) to (1.6), the superscript T stands for a vector or matrix transpose; N is the number of spectral bands; and M is the number of endmembers. The LMM described in equation (1.1) is equivalent to the classic linear regression model (LRM) in the statistics field [20], where A is referred to as the regressor variable, $\vec{y}$ is referred to as the response variable, and $\vec{x}_o$ is referred to as the regression coefficients. The linear regression analysis based on the LRM is a classic statistical analysis technique. A major objective of the linear regression analysis is to estimate the unknown parameters $\vec{x}_o$ in the LRM described in equation (1.1). An important technique to reach this objective of linear regression analysis is the LSE. Both the objective and the technique of linear regression analysis are what we need to solve the linear unmixing problem in the remote sensing field. Based on the LMM (or LRM), the estimation of abundances (or unknown parameters), $\vec{x}_o$, using the LSE method can be described as follows: given mixed pixel vector $\vec{y}$ and endmember matrix A, obtain an optimum estimate, $\vec{x}_{LS}$, of abundance vector $\vec{x}_o$, such that the total error energy, $\|\vec{e}_o\|^2 = \|\vec{y} - A\vec{x}_o\|^2$, is minimized, where the symbol $\|\cdot\|$ stands for a vector norm operation.

A drawback of the LSE method is its sensitivity to the error outlier [20]. Given an element in the error vector $\vec{e}_o$, for example e_{oj}, if it has much greater energy than others, then it is defined as an error outlier. The total error energy $\|\vec{e}_o\|^2$ is greatly affected by such an error outlier. Since the LSE method is to minimize the total error energy, it is greatly affected by the error outlier. As a result, the LSE solution, $\vec{x}_{LS}$, becomes inaccurate. To overcome this drawback, robust approaches have been investigated [30]. One of them is the least median of squares (LMedS) method and has been reported in the spectral unmixing application [31]. Instead of minimizing the total error energy, the LMedS method minimizes the median of the error energy and thus could suppress the influence of error outliers. It has been shown that the LMedS method could produce reasonable estimates as long as the number of outlier data is less than half of the dataset. However, the fact that the LSE method is sensitive to error outliers does not degrade its significance, and the LSE is still the dominant approach for the linear spectral unmixing analysis.

One requirement for implementing the abundance estimation using the LSE method is that the number of spectral bands must be greater than the number of endmembers. This is called the "condition of identifiability" [23, 32], which essentially stems from solving the linear equation (1.1). For the linear spectral unmixing problem, equation (1.1) represents a set of N equations with M unknown variables, where N is the number of spectral bands, and M is the number of endmembers. In general, based on the linear equation theory, there exist three cases: i) when $N = M$, if the solution exists

then it is unique; ii) when $N < M$, which is called an underdetermined system, theoretically there exist infinite nontrivial solutions; iii) when $N > M$, which is called an overdetermined system, there does not exist an exact solution, but it is possible that there exists a unique LSE solution. Case i) seems perfect, but it is not the case for the linear spectral unmixing because of the existence of the random measurement error $\vec{e}_o$ in equation (1.1). Case ii) obviously is not acceptable because of the existence of infinite solutions. Case iii) is suitable for the linear spectral unmixing problem and leads to the "condition of identifiability".

To a certain extent, the "condition of identifiability" limits the use of multispectral data for the linear spectral unmixing problem. Multispectral data typically have only a few spectral bands. For instance, Landsat Thematic Mapper (TM) data have 7 bands and SPOT data have 3 bands. Thus, when the number of endmembers, M, increases, the "condition of identifiability" no longer holds and the LSE method fails. One way to solve the problem is to increase the number of equations, N. For example, Bosdogianni *et al.* incorporated higher order moment characteristics into the LMM equation (1.1) to increase the number of equations and thus alleviate the sufferance from the "condition of identifiability" [33]. Another way to solve the problem is to utilize hyperspectral data, which typically have hundreds of spectral bands. For instance, HYDICE data have 210 spectral bands, AVIRIS data have 224 spectral bands and Hyperion data have 220 spectral bands. The problem of "condition of identifiability" seems easily solved by utilizing hyperspectral data. However, it is questionable that simply using all hyperspectral bands for linear unmixing can lead to a satisfactory result.

For the classification and target detection problem using hyperspectral data, it has been realized that classification performance can be greatly improved by using effective features extracted from the original hyperspectral data [5, 6, 34-36]. Naturally, the question is whether or not linear unmixing performance, particularly abundance estimation performance, can be improved by using appropriate hyperspectral features. Answering this question becomes a major concern of this dissertation.

A variety of advanced techniques have been reported for the feature extraction of hyperspectral signals in the classification and target detection application [5, 6, 34-36]. For example, Jia *et al.* suggested a segmented principal components analysis (PCA) method for feature extraction and classification [34]. The method was based on the conventional PCA [2], but implemented the PCA on the segmented image, not the whole image. The useful features were extracted from principal components in each sub-image. The main advantage of the method was reducing the huge computational cost of the conventional PCA. Jimenez *et al.* utilized a projection pursuit method to extract features and reduce the dimensionality of hyperspectral data [35]. The method implemented a low-dimensional projection of high-dimensional data using a transformation matrix. The transformation matrix was found by optimizing a projection index, which typically was a class separability measurement such as the Bhattacharyya distance measurement [2]. Bruce *et al.* suggested a wavelet transform method for feature extraction and dimensionality reduction [5, 6]. Wavelet transform is a signal processing tool for implementing a multiresolution analysis of signals [37]. Using the wavelet transform, the original signal is represented by a set of wavelet transform coefficients, and features are

extracted from these coefficients. Since the feature space typically has a much lower dimensionality than the original data space, the data dimensionality is reduced. Huang *et al.* investigated a brushlet transform method for feature extraction and dimensionality reduction [36]. Brushlet transform is another multiresolution analysis tool and can be implemented by expanding the Fourier transform onto a series of windowed Fourier bases [38], and thus the brushlet transform coefficients are complex. Similar to the wavelet-based dimensionality reduction, the brushlet-based method transforms the original data into the brushlet domain, features are extracted from the transform coefficients, and the data dimensionality is reduced. Utilizing certain transformations together with feature extraction is a common characteristic of these methods for improving the classification performance.

Among these techniques of feature extraction, the wavelet transform has been our research focus [3-6, 39-41]. As a generic signal processing tool, the wavelet transform also has been extensively applied to various areas of remote sensing applications ranging from image compression [42-44] and image fusion [45-47] to atmospheric correction [48] and pattern recognition [49-51]. Our research interests focus on the pattern recognition task using the wavelet transform, for which the wavelet-based automated classification and target detection systems are designed and implemented for various applications. Generally, in these systems the wavelet transform is utilized to implement the multiresolution analysis of remotely sensed hyperspectral remote sensing signals. The wavelet-based scalar energy features are extracted from the wavelet transform coefficients. The features could be further optimized using the Fisher's linear

discriminant analysis [52]. Using the optimized features, classic statistical classifiers, such as the maximum likelihood classifier [2], are used to implement the classification and target detection. The system performance is evaluated in the sense of classification accuracy. The system performance is also evaluated using a receiver operating characteristics (ROC) analysis method [53]. Various experiments have shown promising results that using wavelet-based features greatly improves the performance of classification and target detection systems. It is these successful applications of the wavelet transform that promote the research topic of this dissertation. That is, we are interested in knowing how the wavelet-based feature extraction can help the linear spectral unmixing problem, and whether or not the feature extraction using the wavelet transform can improve the performance of hyperspectral linear unmixing, or abundance estimation.

Preliminary experimental results in the dissertation show that the feature extraction, particularly the wavelet-based feature extraction, can improve the linear spectral unmixing performance. For example, for a two-endmember abundance estimation problem where the two endmembers are soybean and soil, the average deviation of abundance estimation from true abundances is ~0.05 when using the wavelet-based features extracted from hyperspectral signals. Compared to directly using the original hyperspectral signals for the abundance estimation where the average estimation deviation is ~0.1, the estimation deviation is reduced by ~50%. Results from a three-endmember abundance estimation problem, where the three endmembers are soybean, grass and soil, show that the estimation deviation can be reduced by ~30%.

Specifically, the average deviation of abundance estimation is reduced from ~0.14, where the original hyperspectal signals are directly utilized for the abundance estimation, to ~0.1, where the wavelet-based features are utilized for the abundance estimation. These preliminary experiment results are very promising. Furthermore, we wonder whether there is any fundamental reason that leads to these promising results. Thus, in this dissertation, a series of mathematical analyses are performed. As a result, it is theoretically proven that there do exist some basic rules, according to which using the appropriate features extracted from hyperspectral signals is being able to improve the abundance estimation performance.

Feature extraction approaches based on PCA and discrete cosine transform (DCT) are also investigated for the endmember abundance estimation in the dissertation, as a comparison with the DWT-based method. Note that the PCA-based method has been reported for the spectral unmixing problem [54, 55, 56, 57], but the aim is to reduce the dimensionality of hyperspectral data and the computational expenses. This is a conventional way to use PCA for feature extraction. In general, when dealing with hyperspectral signals, traditionally the dimensionality reduction has been based on methods that provide superior energy compaction, such as PCA and DCT. The reduction of dimensionality has stemmed from the use of only the first few transform coefficients. The approach works well when the aim is signal representation, such as the case of signal compression. However, this approach may be misguided when the aim is signal classification, which is also the aim of the spectral unmixing. In this case, differences between signals take on importance, and simply using the first few large-amplitude

transform coefficients may not be adequate. For this reason, alternative approaches to dimensionality reduction (or feature extraction) based on PCA and DCT are investigated in the dissertation.

1.3. Organization of The Dissertation

The dissertation is organized as follows. Chapter II first provides the background information on the LSE and its application to the hyperspectral linear unmixing problem. Then the error of abundance estimation using the LSE is mathematically derived and analyzed, which leads to a theoretical analysis of how and why feature extraction could improve the performance of abundance estimation using the LSE. The stability of the abundance estimation using the LSE is also discussed in this chapter. Chapter III provides the background information on the wavelet transform and its application to the analysis of hyperspectral signals. Particularly, the implementation of the DWT and the feature extraction approaches based on DWT coefficients are introduced. Chapter IV proposes a DWT-based linear unmixing system of hyperspectral signals. The system consists of two modules. One is called the pre-processing module, in which the DWT is implemented and DWT-based features are extracted. For comparison purposes, two other pre-processing approaches, PCA-based and DCT-based methods, are introduced as well. The other module is the abundance estimation, where a constrained LSE approach is proposed for the abundance estimation of endmembers. To evaluate the abundance estimation performance of the proposed unmixing system, three quantitative evaluation metrics are introduced in this chapter as well. Based on the proposed DWT-based linear unmixing system, Chapter V presents a series of experiments designed for the practical

testing and evaluating of the system. Experiment results, as well as result analyses and discussions, are provided in this chapter. Finally, Chapter VI draws conclusions from the experimental and theoretical analysis and recommends some future research topics.

1.4. Contributions of The Dissertation

The contributions of the dissertation include:

(1) The dissertation demonstrates the need for the use of non-traditional feature extraction (or dimensionality reduction) methods for linear unmixing of hyperspectral signals. It concludes that rather than using feature extraction methods that are based on signal representation, such as conventional PCA and DCT approaches, the remote sensing community needs to investigate feature extraction methods that are based on signal classification for linear unmixing problems.

(2) The dissertation introduces the wavelet transform to the abundance estimation of endmembers using hyperspectral signals. Feature extraction based on the discrete wavelet transform (DWT) is proposed for improving the abundance estimation performance. The dissertation designs and implements a DWT-based linear unmixing system for improving the abundance estimation of endmembers using hyperspectral signals. The system consists of a pre-processing module and an abundance estimation module. The DWT-based features are first extracted from the original hyperspectral signals in the pre-processing module. Then the abundance estimation of endmembers is completed using the DWT-based features, instead of the original hyperspectral signals.

(3) The dissertation both experimentally and theoretically analyzes how the use of appropriate features can improve the abundance estimation performance. Experimental

results show that the use of the DWT-based features extracted from the original hyperspectral signals greatly improves the endmember abundance estimation. Experimental results also show that the use of non-traditional DCT-based features, which are not based on signal representation, improves the endmember abundance estimation. Theoretical results reveal the fundamental reasons why the use of appropriate features, such as the DWT-based features, can improve the abundance estimation of endmembers. Moreover, the theoretical analysis results present a generic criterion to design a feature extraction technique for improving the endmember abundance estimation.

CHAPTER II

LEAST SQUARES ANALYSIS IN LINEAR UNMIXING OF HYPERSPECTRAL SIGNALS

Least squares estimation (LSE) is a mathematical and statistical technique that is used to implement an optimum estimation of parameters based on certain known information [19-21]. The estimation is regarded as optimum in the sense of minimizing the total energy of estimation errors. For the linear spectral unmixing problem, the LSE is used to implement an optimum estimation of abundance, given the information of mixed pixels and constituent pure pixels. The LSE is implemented, in the context of linear spectral unmixing, under an assumption of linear mixture represented in equation (1.1). That is, it is assumed that the mixed pixel is a linear combination of the constituent pure pixels. With this in mind, the major concern of this dissertation is to investigate whether or not the performance of abundance estimation using the LSE can be improved by using specific features extracted from the given mixed-pixel spectrum and constituent pure-pixel spectra. That is, instead of using the original mixed-pixel spectrum and constituent pure-pixel spectra, we use the extracted features to solve the linear mixture equation (1.1) for abundances and analyze the results to determine whether or not abundance estimation is improved. In the dissertation, both theoretical and experimental analysis results show a positive answer to the question. This chapter presents the theoretical analysis results.

2.1. Least Squares Estimation of Abundances

Starting with the linear mixture model of equation (1.1), the objective of the LSE of abundances is to obtain an optimum estimate, $\vec{x}_{LS}$, of abundances $\vec{x}_o$, given mixed-pixel spectrum $\vec{y}$ and endmember spectra A, such that the total error energy,

$$\left\|\vec{e}_o\right\|^2 = \left\|\vec{y} - A\vec{x}_o\right\|^2, \tag{2.1}$$

is minimized, where $\vec{y}$, $\vec{x}_o$, $\vec{e}_o$ and A are defined in equations (1.2) to (1.5), respectively. The symbol $\|\cdot\|$ stands for a vector norm operation, which is defined as the inner product of the vector. That is, equation (2.1) can be rewritten as,

$$\begin{aligned} \vec{e}_o{}^T\vec{e}_o &= (\vec{y} - A\vec{x}_o)^T(\vec{y} - A\vec{x}_o) \\ &= \vec{y}^T\vec{y} - 2\vec{x}_o{}^T A^T\vec{y} + \vec{x}_o{}^T A^T A\vec{x}_o \end{aligned}, \tag{2.2}$$

where superscript T refers to a vector or matrix transpose operation. Defining the total error energy, J_o, of error $\vec{e}_o$,

$$J_o = \left\|\vec{e}_o\right\|^2 = \vec{e}_o{}^T\vec{e}_o, \tag{2.3}$$

a LSE of abundances can be obtained by setting the partial derivative of J_o to zero,

$$\left.\frac{\partial J_o}{\partial \vec{x}_o}\right|_{\vec{x}_{LS}} = \vec{0}. \tag{2.4}$$

Substituting equations (2.2) and (2.3) into equation (2.4), a set of linear equations are obtained,

$$-2A^T\vec{y} + 2A^T A\vec{x}_{LS} = \vec{0}. \tag{2.5}$$

Equation (2.5) can be further simplified as,

$$A^T A\vec{x}_{LS} = A^T \vec{y}\,, \tag{2.6}$$

This is the well-known normal equation of LSE [20], where $A^T A$ is an auto-correlation matrix of endmember spectra and $A^T \vec{y}$ is a cross-correlation vector of endmember spectra and mixed-pixel spectrum. Solving the normal equation (2.6), an optimum estimation, $\vec{x}_{LS}$, of abundance $\vec{x}_o$, can be obtained,

$$\vec{x}_{LS} = (A^T A)^{-1} A^T \vec{y}\,, \tag{2.7}$$

provided that $(A^T A)^{-1}$ exists. Assuming that the endmember spectra are linearly independent, *i.e.*, in equation (1.5) the column vectors in endmember matrix A are linearly independent, it can be shown that the inverse of the auto-correlation matrix, $A^T A$, always exists. The proof is provided in Appendix A.

The objective is to investigate whether or not the abundance estimate, $\vec{x}_{LS}$, can be improved by adjusting the endmember spectra A and mixed-pixel spectrum $\vec{y}$, particularly in this study, by utilizing the features extracted from the endmember spectra A and mixed-pixel spectrum $\vec{y}$. First of all, in order to quantitatively evaluate whether or not the abundance estimation performance is improved, a quantitative criterion needs to be detemined. For this study, mean square error (MSE) of abundance estimate is utilized as such a quantitative criterion. Firstly, an error of abundance estimate, $\vec{e}_x$, is defined as,

$$\vec{e}_x = \vec{x}_{LS} - \vec{x}_o\,, \tag{2.8}$$

where $\vec{x}_o$ refers to the true abundance, defined in equation (1.3). The total error energy, J_x, of error $\vec{e}_x$, can be computed as,

$$J_x = \vec{e}_x{}^T \vec{e}_x . \tag{2.9}$$

Then, the MSE of abundance estimate, Γ_x, can be defined as the mathematical expectation of the average value of the total error energy,

$$\Gamma_x = E\left[\frac{J_x}{M}\right] = E\left[\frac{\vec{e}_x{}^T \vec{e}_x}{M}\right], \tag{2.10}$$

where $E[\cdot]$ represents an operation of mathematical expectation and M is the number of endmembers. Note that an advantage of using the average value of the total error energy, $\frac{J_x}{M}$, is that the MSE is no longer scaled by the number of endmembers. Thus, it allows for a fair comparison of the results of the MSE among different experiments with various numbers of endmembers.

Starting with the definition of the MSE, Γ_x, of abundance estimate in equation (2.10), it is necessary to further represent it using endmember spectra, A, and mixed-pixel spectrum, $\vec{y}$, because the objective is to investigate how the feature extraction from endmember spectra and mixed-pixel spectrum affects the performance of abundance estimation using the LSE. The further derivations will be provided in the next section to build a relationship among the three parameters: (i) the MSE of abundance estimate, Γ_x; (ii) the endmember spectra, A; and (iii) the mixed-pixel spectrum, $\vec{y}$.

2.2. Error Analysis of Abundance Estimation

Substituting equation (1.1) into equation (2.7), $\vec{x}_{LS}$ can be derived as,

$$\vec{x}_{LS} = \vec{x}_o + (A^T A)^{-1} A^T \vec{e}_o, \tag{2.11}$$

where $\vec{x}_o$ is the true abundance and $\vec{e}_o$ is the random measurement error, defined in equations (1.3) and (1.4), respectively. Substituting equation (2.11) into equation (2.8), the abundance estimation error, $\vec{e}_x$, can be expressed as,

$$\vec{e}_x = (A^T A)^{-1} A^T \vec{e}_o. \tag{2.12}$$

Utilizing equation (2.10) and the definition of the trace of a matrix, the MSE of abundance estimate, Γ_x, can be derived as,

$$\Gamma_x = E\left[\frac{\vec{e}_x{}^T \vec{e}_x}{M}\right] = E\left[Tr\left[\frac{\vec{e}_x{}^T \vec{e}_x}{M}\right]\right] = \frac{1}{M} E\left[Tr[\vec{e}_x{}^T \vec{e}_x]\right], \tag{2.13}$$

where $Tr[\cdot]$ is a matrix trace operation and defined as a summation of main diagonal elements of a matrix. Substituting equation (2.12) into equation (2.13), Γ_x can be further derived as,

$$\Gamma_x = \frac{1}{M} Tr\left[(A^+)^T A^+ R_o\right], \tag{2.14}$$

where $A^+ = (A^T A)^{-1} A^T$ is defined as a pseudo-inverse of matrix A and $R_o = E[\vec{e}_o \vec{e}_o{}^T]$ is an auto-correlation matrix of random measurement error vector $\vec{e}_o$. The derivation of equation (2.14) is provided in Appendix B.

Suppose that the random measurement error vector, $\vec{e}_o$, defined in equation (1.4), has the following two statistical characteristics:

(i) *Assumption 2.1.1:* $\vec{e}_o$ has zero mean,

$$E[\vec{e}_o] = \vec{0}; \tag{2.15}$$

(ii) *Assumption 2.1.2:* Random elements, e_{oj} (for $j = 1,2,\cdots,N$), in $\vec{e}_o$ are uncorrelated and have different variances of ${\sigma_{oj}}^2$, respectively, where N is the number of spectral bands.

Then, the auto-correlation matrix (or a covariance matrix due to $E[\vec{e}_o] = \vec{0}$), R_o, in equation (2.14) can be computed as,

$$R_o = E[\vec{e}_o \vec{e}_o{}^T] = \begin{bmatrix} {\sigma_{o1}}^2 & 0 & \cdots & 0 \\ 0 & {\sigma_{o2}}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & {\sigma_{oN}}^2 \end{bmatrix}, \tag{2.16}$$

Note that R_o is a diagonal matrix with ${\sigma_{oj}}^2$ being main diagonal elements. Substituting equation (2.16) into equation (2.14), the MSE of abundance estimate, Γ_x, can be further computed as,

$$\Gamma_x = \frac{1}{M} \sum_{j=1}^{N} p_j {\sigma_{oj}}^2, \tag{2.17}$$

where p_j (for $j = 1,2,\cdots,N$) are the main diagonal elements in the symmetric matrix, $(A^+)^T A^+$. Note that both p_j and ${\sigma_{oj}}^2$ are nonnegative, and thus each $p_j {\sigma_{oj}}^2$ is nonnegative.

To solve equation (1.1) for abundances, the endmember spectra, A, defined in equation (1.5), have to be determined *a priori*. Typically this is accomplished by using a

library of endmember spectra. It is inevitable that there exist differences between the library endmember spectra and the true endmember spectra constituting the mixed-pixels. Thus, it is reasonable to assume that the random measurement error, $\vec{e}_o$, defined in equation (1.4), stems from this difference, that is,

Assumption 2.2: The random measurement error in LMM stems from the difference between the library endmember spectra and the true endmember spectra constituting the mixed-pixels.

Based on *Assumption 2.2*, equation (1.1) can be rewritten as,

$$\vec{y} = (A + \Delta_A)\vec{x}_o = A\vec{x}_o + \Delta_A \vec{x}_o, \tag{2.18}$$

where matrix A represents the library endmember spectra and matrix $\Delta_A = [\vec{\delta}_1, \vec{\delta}_2, \cdots, \vec{\delta}_M]$ represents the differences between the library endmember spectra and the true endmember spectra constituting the mixed-pixels. Note that in matrix Δ_A, the i^{th} column vectors, $\vec{\delta}_i = [\delta_{i1}, \delta_{i2}, \cdots, \delta_{iN}]^T$ (for $i = 1,2,\cdots,M$), represents the difference between the i^{th} library and true endmember spectra, where M is the number of endmembers and N is the number of spectral bands. Comparing equations (1.1) and (2.18), the random measurement error $\vec{e}_o$ can be expressed, in terms of the random endmember spectrum differences,

$$\vec{e}_o = \Delta_A \vec{x}_o, \tag{2.19}$$

where $\vec{x}_o$ is the true abundance vector, as defined in equation (1.3).

Furthermore, suppose that the random errors, $\vec{\delta}_i$ (for $i = 1,2,\cdots,M$), have the following three statistical characteristics:

(i) *Assumption 2.3.1:* $\vec{\delta}_i$ are linearly independent;

(ii) *Assumption 2.3.2:* Each of the $\vec{\delta}_i$ has zero mean,

$$E[\vec{\delta}_i] = \vec{0}; \tag{2.20}$$

(iii) *Assumption 2.3.3:* Random elements, δ_{ij} (for $i = 1,2,\cdots,M$ and $j = 1,2,\cdots,N$), in each $\vec{\delta}_i$ are uncorrelated and have different variances of ${\sigma_{ij}}^2$, respectively, where M is the number of the endmembers and N is the number of spectral bands.

Then there exists a correlation matrix (or a covariance matrix due to $E[\vec{\delta}_i] = \vec{0}$), R_i, for each $\vec{\delta}_i$,

$$R_i = E[\vec{\delta}_i \vec{\delta}_i^{\,T}] = \begin{bmatrix} {\sigma_{i1}}^2 & 0 & \cdots & 0 \\ 0 & {\sigma_{i2}}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & {\sigma_{iN}}^2 \end{bmatrix}, \tag{2.21}$$

Note that each R_i is a diagonal matrix with ${\sigma_{ij}}^2$ being main diagonal elements. Based on these three assumptions and equation (2.19), ${\sigma_{oj}}^2$ in equation (2.17) can be further derived as,

$${\sigma_{oj}}^2 = \sum_{i=1}^{M} {\sigma_{ij}}^2 {x_{oi}}^2, \quad (j = 1,2,\cdots,N), \tag{2.22}$$

where each of x_{oi}, as defined in equation (1.3), is a true abundance corresponding to each of the endmembers. The derivation of equation (2.22) is provided in Appendix C.

Substituting equation (2.22) into equation (2.17), the MSE of the abundance estimate, Γ_x, can be further computed as,

$$\Gamma_x = \frac{1}{M}\sum_{j=1}^{N}\sum_{i=1}^{M} p_j \sigma_{ij}{}^2 x_{oi}{}^2, \tag{2.23}$$

Finally, equation (2.23) builds a relationship among the three parameters: (i) the MSE of abundance estimate, Γ_x; (ii) the endmember spectra, A; and (iii) the mixed-pixel spectrum, $\vec{y}$. Note that in equation (2.23), M, N and x_{oi} are all constants for the given endmember spectra and mixed-pixel spectrum. Thus, the value of Γ_x is only affected by the values of two variable terms, p_j and $\sigma_{ij}{}^2$. Based on equation (2.17), the variable term p_j (for $j = 1,2,\cdots,N$) are main diagonal elements in the symmetric matrix $(A^+)^T A^+$, and thus uniquely determined by the endmember spectra, A. The variable term $\sigma_{ij}{}^2$ (for $i = 1,2,\cdots,M$ and $j = 1,2,\cdots,N$) indicates the variances of endmember spectra. It can be called the within-endmember variances.

According to equation (2.23), if we can reduce the endmember spectrum variance $\sigma_{ij}{}^2$, and at the same time adjust the values of p_j such that the product of p_j and $\sigma_{ij}{}^2$ is reduced, then we can reduce the MSE of abundance estimate, Γ_x. Based on our previous research experience of classification and target detection using hyperspectral signals [5, 6, 39-41], it is possible to reduce the within-endmember (or within-class for the classification problem) variances, $\sigma_{ij}{}^2$, by using the feature extraction approach based on discrete wavelet transform (DWT). However, at this point, it is not apparent how the

value of p_j is affected by the use of feature extraction methods, such as those based on the DWT. Therefore, some further results are derived as shown in the next section.

2.3. Simplification of Abundance Error Analysis Results

The result of equation (2.23) can be simplified by making further assumptions. First of all, note that in *Assumption 2.1.2* concerning the random measurement error vector $\vec{e}_o$, it is assumed that each of the random elements e_{oj} (for $j = 1,2,\cdots,N$) in $\vec{e}_o$ has its own variance $\sigma_{oj}^{\ 2}$. Thus the resulting correlation matrix, R_o, is a diagonal matrix, as shown in equation (2.16). Furthermore, suppose that

Assumption 2.4: All the random elements e_{oj} (for $j = 1,2,\cdots,N$) in $\vec{e}_o$ have the same variance $\sigma_o^{\ 2}$, where N is the number of spectral bands.

Then the correlation matrix, R_o, can be reduced to an identity matrix multiplied by a constant factor, $\sigma_o^{\ 2}$, *i.e.*,

$$R_o = E[\vec{e}_o \vec{e}_o^{\ T}] = \sigma_o^{\ 2} I\,, \tag{2.24}$$

where I is a N by N identity matrix and N is the number of spectral bands. Substituting equation (2.24) into equation (2.14), a simplified expression about the MSE of abundance estimate, Γ_x, can be derived as,

$$\Gamma_{xs} = \frac{1}{M} Tr\left[R^{-1}\right] \sigma_o^{\ 2}\,, \tag{2.25}$$

where Γ_{xs} represents a simplified version of Γ_x, $R = A^T A$ is an auto-correlation matrix of endmember spectra, and M is the number of endmembers. The derivation of equation (2.25) is provided in Appendix D.

Secondly, note that in *Assumption 2.3.3* concerning the differences of endmember spectra, $\vec{\delta}_i$, it is assumed that each of the random elements, δ_{ij} (for $i = 1,2,\cdots,M$ and $j = 1,2,\cdots,N$), in $\vec{\delta}_i$ has its own variance $\sigma_{ij}^{\ 2}$. Thus the resulting correlation matrix, R_i (for $i = 1,2,\cdots,M$), is a diagonal matrix, as shown in equation (2.21). Given that *Assumptions 2.1.1*, *2.3.1* and *2.3.2* concerning $\vec{e}_o$ and $\vec{\delta}_i$ (for $i = 1,2,\cdots,M$) hold, in order to match *Assumption 2.4* that all the random elements in $\vec{e}_o$ has the same variance $\sigma_o^{\ 2}$, we make the following further assumption:

Assumption 2.5: All the random elements δ_{ij} (for $i = 1,2,\cdots,M$ and $j = 1,2,\cdots,N$) in $\vec{\delta}_i$ has the same variance $\sigma_i^{\ 2}$, where M is the number of the endmembers and N is the number of spectral bands.

Then the correlation matrix, R_i, can be reduced to an identity matrix multiplied by a constant factor, $\sigma_i^{\ 2}$, *i.e.*,

$$R_i = E[\vec{\delta}_i \vec{\delta}_i^{\ T}] = \sigma_i^{\ 2} I\,, \tag{2.26}$$

where I is a N by N identity matrix and N is the number of spectral bands. Based on the two further assumption, *Assumptions 2.4* and *2.5*, using equations (2.19), (2.24) and (2.26), $\sigma_o^{\ 2}$ can be derived as,

$$\sigma_o^{\,2} = \sum_{i=1}^{M} \sigma_i^{\,2} x_{oi}^{\,2} \,, \tag{2.27}$$

where x_{oi} are true abundances corresponding to each of endmembers, as defined in equation (1.3). The derivation of equation (2.27) is provided in Appendix C. Substituting equation (2.27) into equation (2.25), the simplified version of the MSE of abundance estimate, Γ_{xs}, can be computed as,

$$\Gamma_{xs} = \frac{1}{M} Tr\left[R^{-1}\right] \sum_{i=1}^{M} \sigma_i^{\,2} x_{oi}^{\,2} \,, \tag{2.28}$$

Equation (2.28) is a simplified version of equation (2.23), under the two further assumptions, *Assumptions 2.4* and *2.5*. Note that these two assumptions may not hold for some cases. For example, it will be experimentally shown in the following chapters that these assumptions do not hold when using the original hyperspectral signals for abundance estimation, but do hold when using features based on the DWT for abundance estimation. The simplified expression of the MSE of abundance estimate, Γ_{xs}, provides further insights into how the feature extraction can affect the abundance estimation using the LSE. Note that in equation (2.28) Γ_{xs} is represented by two separated terms multiplied by a constant factor $\frac{1}{M}$. One term is a weighted sum of the within-endmember variance, $\sigma_i^{\,2}$, which is similar to $\sigma_{ij}^{\,2}$ in equation (2.23) except that now all $\sigma_{ij}^{\,2}$ are the same and equal to $\sigma_i^{\,2}$. The other term is $Tr[R^{-1}]$, which is a simplified alternative of p_j term in equation (2.23).

Note that $R = A^T A$ used in equation (2.28) is an auto-correlation matrix of endmember spectra. Thus, intuitively there should exist certain relationship between $Tr[R^{-1}]$ and the correlation among the endmember spectra. Experimentally it can be shown that the value of $Tr[R^{-1}]$ can be reduced when increasing the separability among endmember spectra. Even though this trend is not a simple monotonic relationship, it does provide a possibility to reduce the value of $Tr[R^{-1}]$ via adjusting the endmember spectra, *e.g.*, via feature extraction. Again, based on our previous research on classification and target detection using hyperspectral signals [5, 6, 39-41], it is possible to increase the separability among classes by using the feature extraction approach based on the DWT, where the separability is measured by the between-class (or between-endmember for the linear unmixing problem) variances. That is, it is possible to reduce the value of $Tr[R^{-1}]$ by increasing the between-endmember variance by using the DWT-based features.

In summary, it is possible to reduce the MSE of abundance estimate, Γ_{xs}, by using appropriate features, such as DWT-based features, extracted from the original endmember spectra and mixed-pixel spectra, because the use of appropriate features can reduce the within-endmember variance and increase the between-endmember variance, and consequently reduce the values of σ_i^2 and $Tr[R^{-1}]$, as well as the value of Γ_{xs} according to equation (2.28).

2.4. Stability Analysis of Abundance Estimation

The auto-correlation matrix of endmember spectra, $R = A^T A$ used in equation (2.28), is real and symmetric, and can be represented in terms of its eigenvalues and eigenvectors [19],

$$R = Q\Lambda Q^T, \tag{2.29}$$

where Q is a matrix with each column being an eigenvector of R, and Λ is a diagonal matrix with diagonal elements being eigenvalues of R. Based on equation (2.29), it can be shown that,

$$Tr[R^{-1}] = Tr[\Lambda^{-1}] = \sum_{i=1}^{M} \frac{1}{\lambda_i}, \tag{2.30}$$

where M is the number of endmembers and λ_i are M distinct eigenvalues of R. The derivation of equation (2.30) is provided in Appendix E. Since the auto-correlation matrix, R, is positive definite (the proof is provided in Appendix A), the values of λ_i are all positive. According to equation (2.30), the value of $Tr[R^{-1}]$ is directly related to the values of λ_i. If there exists an extremely small value among λ_i, which indicates that the matrix R is ill-conditioned [19], then it will result in an extremely large value of $Tr[R^{-1}]$, and thus possibly large value of Γ_{xs} according to equation (2.28). Based on this analysis, an ill-conditioned R should be avoided in order to avoid a large MSE of abundance estimation.

The ill-condition or the well-condition of the matrix, R, is measured by the condition number of a matrix [19], which is computed as,

$$CN(R) = \|R\|\|R^{-1}\|, \tag{2.31}$$

where $\|R\|$ is the norm of matrix R, which is defined as,

$$\|R\| = \max_{\vec{z}\neq\vec{0}}\left\{\frac{\|R\vec{z}\|}{\|\vec{z}\|}\right\}, \tag{2.32}$$

where $\vec{z}$ is any non-zero vector. Equation (2.32) also implies the following fact that [19],

$$\|R\vec{z}\| \le \|R\|\|\vec{z}\|. \tag{2.33}$$

That is, the norm of matrix R, $\|R\|$, bounds the amplification power of matrix R in a linear transformation $R\vec{z}$. For a positive definite matrix, such as the auto-correlation matrix of endmember spectra R in the linear unmixing problem, the condition number can be simply computed as [19],

$$CN(R) = \frac{\lambda_{\max}}{\lambda_{\min}}, \tag{2.34}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum eigenvalues of matrix R, respectively. Based on the definition of the condition number, shown in equation (2.31) or (2.34), an ill-conditioned matrix has a large condition number, and the best condition number is 1, *i.e.*,

$$CN(R) \ge 1. \tag{2.35}$$

This can be easily shown using the definition of the condition number, equation (2.31) or (2.34), that is,

$$CN(R) = \|R\|\|R^{-1}\| \ge \|RR^{-1}\| = \|I\| = 1, \tag{2.36}$$

where I is referred to as an identity matrix, or,

$$CN(R) = \frac{\lambda_{\max}}{\lambda_{\min}} \geq 1. \tag{2.37}$$

For the LSE problem, the concept of the condition number also can be utilized to measure the stability of the least squares solution, $\vec{x}_{LS}$, in equation (2.7). Rewriting the normal equation (2.6) in the following form,

$$R\vec{x}_{LS} = \vec{d}, \tag{2.38}$$

where $R = A^T A$ is an auto-correlation matrix of endmember spectra, and $\vec{d} = A^T \vec{y}$ is a cross-correlation vector of endmember spectra and mixed-pixel spectrum. Obviously, equation (2.38) is a linear equation and $R\vec{x}_{LS}$ is a linear transformation of $\vec{x}_{LS}$. The solution of $\vec{x}_{LS}$ is completely determined by the auto-correlation matrix, R, and the cross-correlation vector, $\vec{d}$. Then, the question is how the least squares solution, $\vec{x}_{LS}$, is influenced, if there exists some disturbance in R or $\vec{d}$.

Based on equations (2.33) and (2.34), it can be shown that there exist the following relationships between the disturbance in R or $\vec{d}$ and the disturbance in the least squares solution $\vec{x}_{LS}$:

$$\frac{\left\|\vec{\delta}_{\vec{x}_{LS}}\right\|}{\left\|\vec{x}_{LS}\right\|} \leq CN(R) \frac{\left\|\vec{\delta}_{\vec{d}}\right\|}{\left\|\vec{d}\right\|}, \tag{2.39}$$

and

$$\frac{\left\|\vec{\delta}_{\vec{x}_{LS}}\right\|}{\left\|\vec{x}_{LS}\right\|} \leq CN(R) \frac{\left\|\delta_R\right\|}{\left\|R\right\|}, \tag{2.40}$$

where $CN(R)$ is the condition number of matrix R defined in equation (2.31), δ_R and $\vec{\delta}_{\vec{d}}$ are referred to as the disturbances in R and $\vec{d}$, respectively, and $\vec{\delta}_{\vec{x}_{LS}}$ is referred to as an error associated with $\vec{x}_{LS}$, caused by the disturbances δ_R and $\vec{\delta}_{\vec{d}}$. The derivations of equations (2.39) and (2.40) are provided in Appendix F. These results indicate that the relative error in the solution $\vec{x}_{LS}$ caused by the disturbances in R and $\vec{d}$ is bounded by the condition number of matrix R. A matrix is said to be ill-conditioned if a small disturbance in R or $\vec{d}$ results in a large disturbance in the solution $\vec{x}_{LS}$, which also implies a large condition number of matrix R, according to equations (2.39) and (2.40). Therefore, in order to obtain a stable solution of $\vec{x}_{LS}$, we need to reduce the condition number of matrix R and avoid an ill-conditioned matrix R.

Finally, in summarizing the theoretical analysis in this chapter, we propose the use of appropriate feature extraction approaches, particularly the feature extraction method based on the discrete wavelet transform, to (i) reduce within-endmember variances and increase between-endmember variances in order to reduce the abundance estimation error using the LSE, and (ii) reduce the condition number of the auto-correlation matrix R of endmember spectra, $CN(R)$, in order to avoid an ill-conditioned matrix R, which could result in an extremely large abundance estimation error and an unstable solution of abundance estimation.

CHAPTER III

FEATURE EXTRACTON USING DISCRETE WAVELET TRANSFORM FOR LINEAR UNMIXING OF HYPERSPECTRAL SIGNALS

Feature extraction plays a significant role in target detection and classification applications using remotely sensed hypersepctral signals. The use of appropriate features not only can reduce the data volume (or dimensionality) of hyperspectral signals and thus computational costs of hyperspectral signal analysis, but also can improve the performance of target detection and classification. The reduction of data dimensionality and computational costs results from the fact that the original high-dimensional hyperspectral signals are represented in a lower-dimensional space using fewer features. The improvement of target detection and classification performance stems from the possibility that feature extraction can make different classes more separable in feature spaces, and thus lead to a better classification performance. The separability is typically measured by within-class variances and between-class variances [2, 6, 52]. The basic idea is that the class separability is increased when within-class variances are decreased and between-class variances are increased. The fundamental results from the theoretical analysis in chapter II also shows that this idea is suitable to the linear unmixing problem of hyperspectral signals. That is, using appropriate features can make different endmembers more separable in feature space, and thus result in an improvement of abundance estimation performance. The increase of endmember separability is achieved

by feature extraction to decrease within-endmember variances and increase between-endmember variances.

This chapter will introduce feature extraction approaches based on the discrete wavelet transform (DWT) for the linear unmixing problem of hyperspectral signals. The DWT is a mathematical technique that is used to implement a multiresolution analysis (MRA) of signals [37]. The fine-scale and large-scale information of a hyperspectral signal can be simultaneously investigated by projecting the signal onto a set of wavelet basis functions with various scales. Extracting appropriate features from the wavelet-based multiresolution information provides a potential to effectively discriminate hyperspectral signals in feature space. The idea has been verified in our previous research [3-6, 39-41] for target detection and classification problems using hyperspectral signals. A large variety of choices of features could be extracted from the DWT-based multiresolution information, including the DWT coefficients themselves or any combination of the coefficients. According to characteristics of linear unmixing problem, linear wavelet-based features are preferred. In this chapter, the DWT-based feature extraction approaches and corresponding properties will be described and discussed. The discussion starts with a fundamental and brief introduction to the wavelet transform.

3.1. Background on Wavelet Transform

The wavelet transform of a signal, $f(x)$, is defined as an inner product of the signal and wavelet bases, *i.e.*,

$$W(s,b) =< f(x), \psi_{s,b}(x) >, \tag{3.1}$$

where $\psi_{s.b}(x)$ are referred to as wavelet bases and $W(s,b)$ are referred to as wavelet transform coefficients of signal $f(x)$. The wavelet bases, $\psi_{s.b}(x)$, can be formed from a basic wavelet (or mother wavelet), $\psi(x)$, by a series of scaling and shifting operations. The wavelet bases formed by this manner are usually called the first generation wavelets [58, 59, 60, 61]. The mother wavelet, $\psi(x)$, can be any real function, satisfying the following admissibility condition [62, 63],

$$\int_{-\infty}^{+\infty} \frac{\left|\Psi(\omega)]\right|^2}{|\omega|} d\omega < \infty , \tag{3.2}$$

where $\Psi(\omega)$ is the Fourier transform of $\psi(x)$, and ω is the Fourier domain variable. Note that since ω is in the denominator of the integrand in equation (3.2), it is necessary that $\Psi(0) = 0$, which results in,

$$\int_{-\infty}^{+\infty} \psi(x) dx = 0 . \tag{3.3}$$

Equation (3.3) indicates that the mother wavelet must oscillate and have an average value of zero.

From the mother wavelet, $\psi(x)$, the wavelet bases, $\psi_{s.b}(x)$, are formed by a series of scaling and shifting operations,

$$\psi_{s,b}(x) = \frac{1}{\sqrt{s}}\psi\left(\frac{x-b}{s}\right), \tag{3.4}$$

where $s > 0$ and b are any real numbers. The variable s indicates the scale (or width) of a particular basis function, and the variable b specifies its shifted position. Using wavelet bases in equation (3.4), the wavelet transform defined in equation (3.1) can be computed as,

$$W(s,b) = \frac{1}{\sqrt{s}} \int_{x=-\infty}^{\infty} f(x)\psi\left(\frac{x-b}{s}\right)dx, \tag{3.5}$$

provided that the wavelet bases, $\psi_{s.b}(x)$, are real. Equation (3.5) is a continuous version of the transform, generally called continuous wavelet transform (CWT).

If we discretize the scale and shift parameters, s and b, in equation (3.4) to integer values, *i.e.*, let $s = {s_0}^j$ and $b = {s_0}^j k$, where $s_0 \geq 2$ and $-\infty < j,k < \infty$ are any integers, then discrete wavelet bases, $\psi_{j,k}(n)$, can be formed as,

$$\psi_{j,k}(n) = \frac{1}{\sqrt{{s_0}^j}}\psi\left(\frac{n - {s_0}^j k}{{s_0}^j}\right). \tag{3.6}$$

Based on the definition of wavelet transform in equation (3.1), using the discrete wavelet bases, a discrete wavelet transform (DWT) of a finite energy sequence with N samples, $f(n)$, can be computed as,

$$W(j,k) = \frac{1}{\sqrt{{s_0}^j}} \sum_{n=0}^{N-1} f(n)\psi\left(\frac{n - {s_0}^j k}{{s_0}^j}\right), \tag{3.7}$$

where $W(j,k)$ are referred to as the DWT coefficients of sequence $f(n)$. When $s_0 = 2$, equation (3.7) defines a well-known two-channel (or dyadic) DWT, *i.e.*,

$$W(j,k)\big|_{s_0=2} = 2^{\frac{-j}{2}} \sum_{n=0}^{N-1} f(n)\psi\left(2^{-j}n - k\right). \tag{3.8}$$

In practice, instead of computing the inner product defined in equation (3.8), there exist computationally efficient algorithms to implement the dyadic DWT [37, 59]. One such algorithm is illustrated in the upper part of Figure 3.1. The algorithm is usually referred to as the fast wavelet transform (FWT). The basic idea behind the FWT algorithm is to represent the mother wavelet as a set of high-pass and low-pass filters in a filter bank [64, 65]. Then the signal is passed through the filter bank. Following the filtering, the signal is decimated by a factor of 2. The outputs of the low-pass branch are called wavelet approximation coefficients, and the outputs of the high-pass branch are called wavelet detail coefficients. This filtering process followed by decimation is referred to as single-step wavelet decomposition. The single-step wavelet decomposition can be performed iteratively. At each iteration step, the wavelet approximation coefficients from the previous scale are used as the input of the filter bank. Initially, the original signal is the input of the filter bank. While theoretically this iteration could be continued *ad infinitum*, in many practical settings the iteration is performed until a pre-defined scale (a set maximum) is reached. Oftentimes, the maximum scale is set to be the scale at which the number of the wavelet approximation coefficients is just less than the filter length, in order to avoid the trivial filtering operation. Note that the FWT algorithm implements the dyadic DWT from the bottom up in an iterative manner, *i.e.*, computing fine-scale coefficients first.

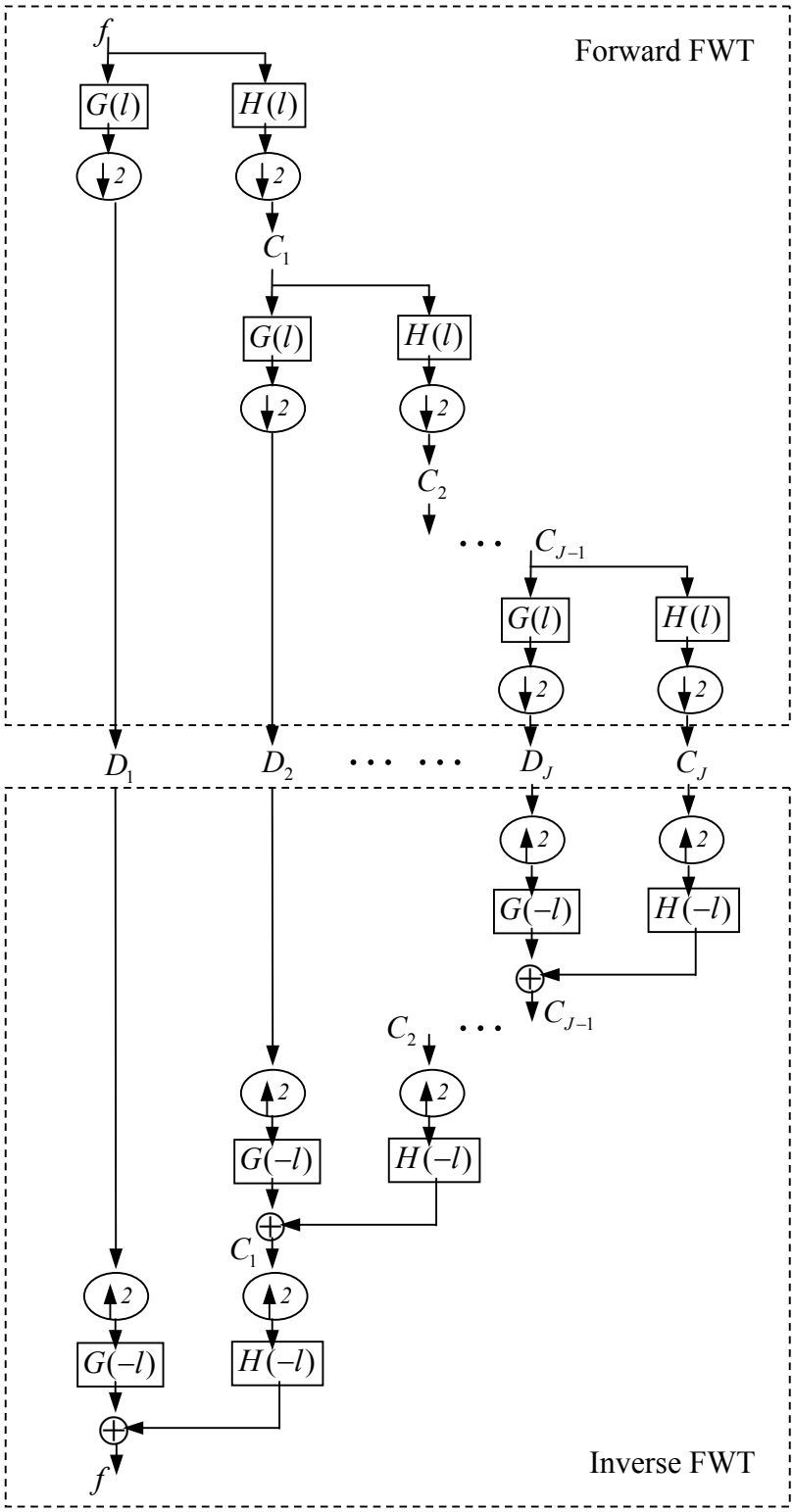

Figure 3.1. The dyadic (2-channel) forward and inverse FWT.

In general, the FWT algorithm can be described using the following iteration expressions,

$$D_{j+1}(k) = \sum_{l=0}^{L-1} G(l) \cdot C_j(2k+l) \ , \tag{3.9}$$

$$C_{j+1}(k) = \sum_{l=0}^{L-1} H(l) \cdot C_j(2k+l) , \tag{3.10}$$

where D_j and C_j are the wavelet detail and approximation coefficients at scale j, respectively. The functions G and H are referred to as the high-pass and low-pass decomposition filter impulse responses, respectively, and L is the filter length. Initially, C_0 (with $j = 0$) is equivalent to the original finite-length sequence, $f(n)$.

So far, we have discussed the two-channel FWT algorithm. It can be easily extended to the multi-channel case. According to the definition of DWT in equation (3.7), when $s_0 = M$, where $M \geq 2$ is any constant integer, a M-channel DWT can be computed as,

$$W(j,k)\big|_{s_0=M} = M^{\frac{-j}{2}} \sum_{n=0}^{N-1} f(n)\psi\left(M^{-j}n - k\right). \tag{3.11}$$

If we represent the mother wavelet as a set of high-pass, band-pass, and low-pass filters in a multi-channel filter bank [66, 67, 68], as illustrated in Figure 3.2, then a M-channel FWT algorithm can be described using the following iteration expressions,

$$D_{j+1,m}(k) = \sum_{l=0}^{L-1} G_m(l)\, C_j(Mk+l) \ , \tag{3.12}$$

$$C_{j+1}(k) = \sum_{l=0}^{L-1} H(l) C_j(Mk+l) . \tag{3.13}$$

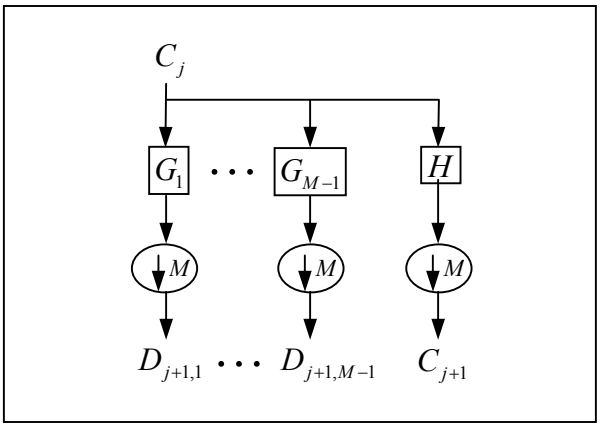

Figure 3.2. The M -channel forward FWT.

In equations (3.12) and (3,13), $D_{j,m}$ are the wavelet detail coefficients at scale j and channel m, and C_j are the wavelet approximation coefficients at scale j. The functions G_m are $M-1$ high-pass and band-pass decomposition filters and the function H is the low-pass decomposition filter. Similar to the two-channel FWT, the single-step M -channel wavelet decomposition consists of a filtering operation and a following decimation of factor M.

The signal can be perfectly reconstructed from the wavelet approximation and detail coefficients when applying the FWT algorithm inversely [59, 64-67]. As an example, a two-channel inverse FWT (IFWT) as illustrated in lower part of Figure 3.1, where the wavelet approximation and detail coefficients are firstly up-sampled by inserting zeros between any two DWT coefficients. Then, the up-sampled wavelet coefficients are filtered using a set of reconstruction filters. Note that the reconstruction filters are simply time reversals of the decomposition fitlers when the corresponding

wavelet bases are orthogonal [58, 59, 69]. Finally, the filter outputs are summed. This process of up-sampling, filtering and summation is referred to as single-step wavelet reconstruction. The single-step wavelet reconstruction can be iteratively performed, starting from the wavelet approximation and detail coefficient at the maximum decomposition scale, until the original signal is completely reconstructed.

Figure 3.3 shows an example of the wavelet analysis of a signal using the FWT algorithm. The signal, f, is a typical vegetation hyperspectral reflectance curve, consisting of 1400 samples which represents 1400 spectral bands between 354nm and 1753nm wavelengths. The mother wavelet utilized for the DWT is the Haar wavelet, which is defined as,

$$\psi(n) = \begin{cases} 1, & 0 \le n \le 1/2 \\ -1, & 1/2 \le n \le 1 \\ 0, & else \end{cases}. \tag{3.14}$$

The associated high-pass and low-pass decomposition filter impulse responses have only two samples,

$$G = [-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}], \tag{3.15}$$

$$H = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}], \tag{3.16}$$

which are the shortest possible wavelet filters. Note that the Haar wavelet bases are orthogonal [58, 59], thus the corresponding wavelet decomposition and reconstruction filters are the same. In this example, the filter G and H defined in equations (3.15) and (3.16) are used in the FWT algorithm, and the signal is decomposed until the 5th scale.

Figure 3.3 shows the wavelet decomposition coefficients at various scales based on the FWT algorithm. Figure 3.4 shows the various-scale reconstructed signals, which is obtained by using the inverse FWT algorithm described in Figure 3.2 and using only the wavelet coefficients at the corresponding scale while setting other coefficients to zeros. That is, the reconstructed signals, d_j, include only information in wavelet detail coefficients, D_j, for $j = 1,2,\cdots,5$. The reconstructed signal, c_5, includes only information in wavelet approximation coefficients, C_5. It can be seen that the fine-scale and large-scale information in the signal is extracted simultaneously after the DWT analysis. It is also interesting to notice that there exist the following relationships,

$$f = c_5 + d_5 + d_4 + \cdots + d_1. \tag{3.17}$$

That is, the original signal can also be perfectly reconstructed by summing the various-scale reconstructed signals.

In summary, the DWT allows a simultaneous investigation of fine-scale and large-scale information in the signal without losing any information. Thus it provides a potential to extract appropriate features from the DWT coefficients for improving the classification or linear unmixing performance. Furthermore, the FWT algorithm implements the DWT in a computationally efficient manner, and makes it possible for the DWT to be effectively used in practical applications.

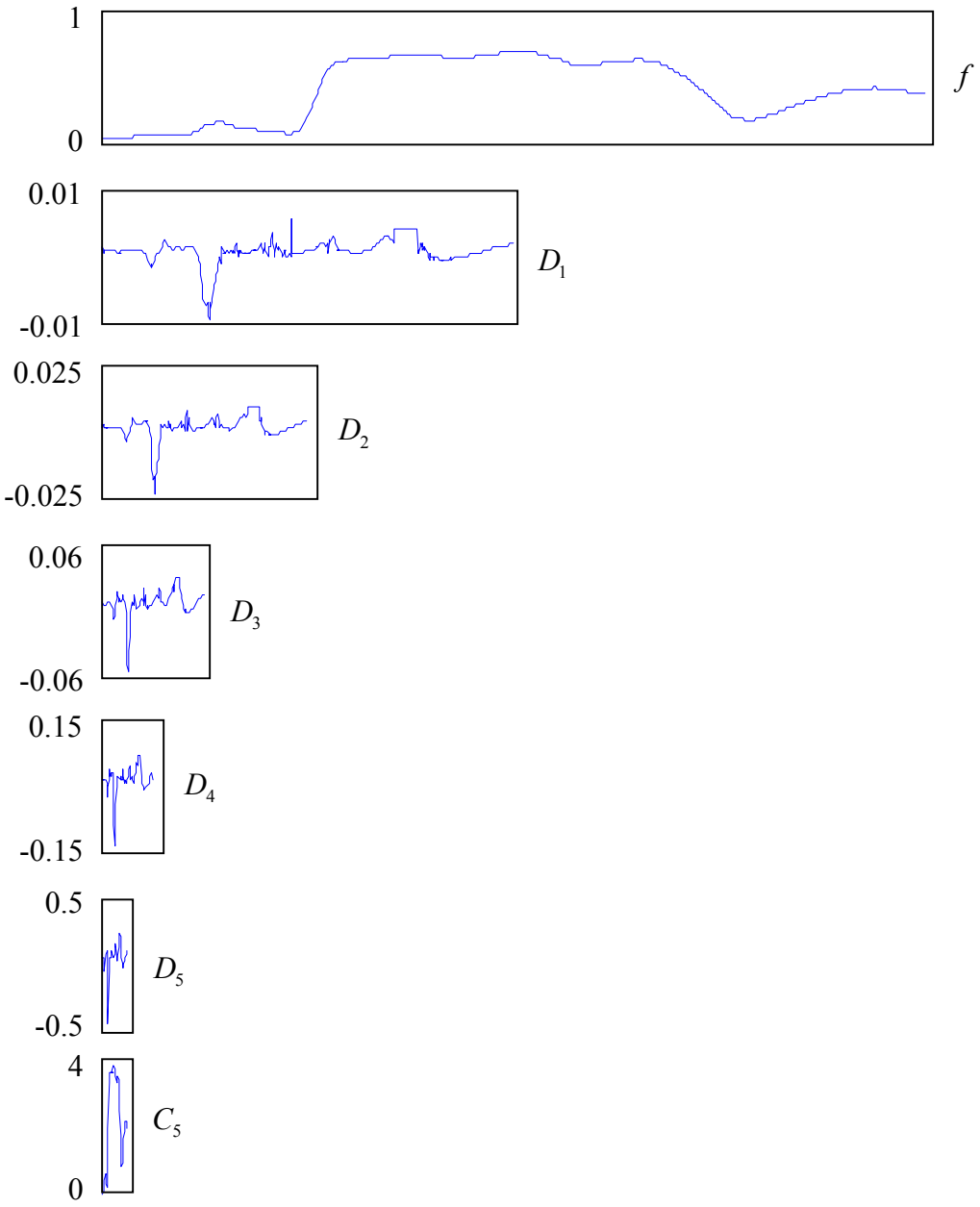

Figure 3.3. An example of the dyadic (2-channel) forward FWT.

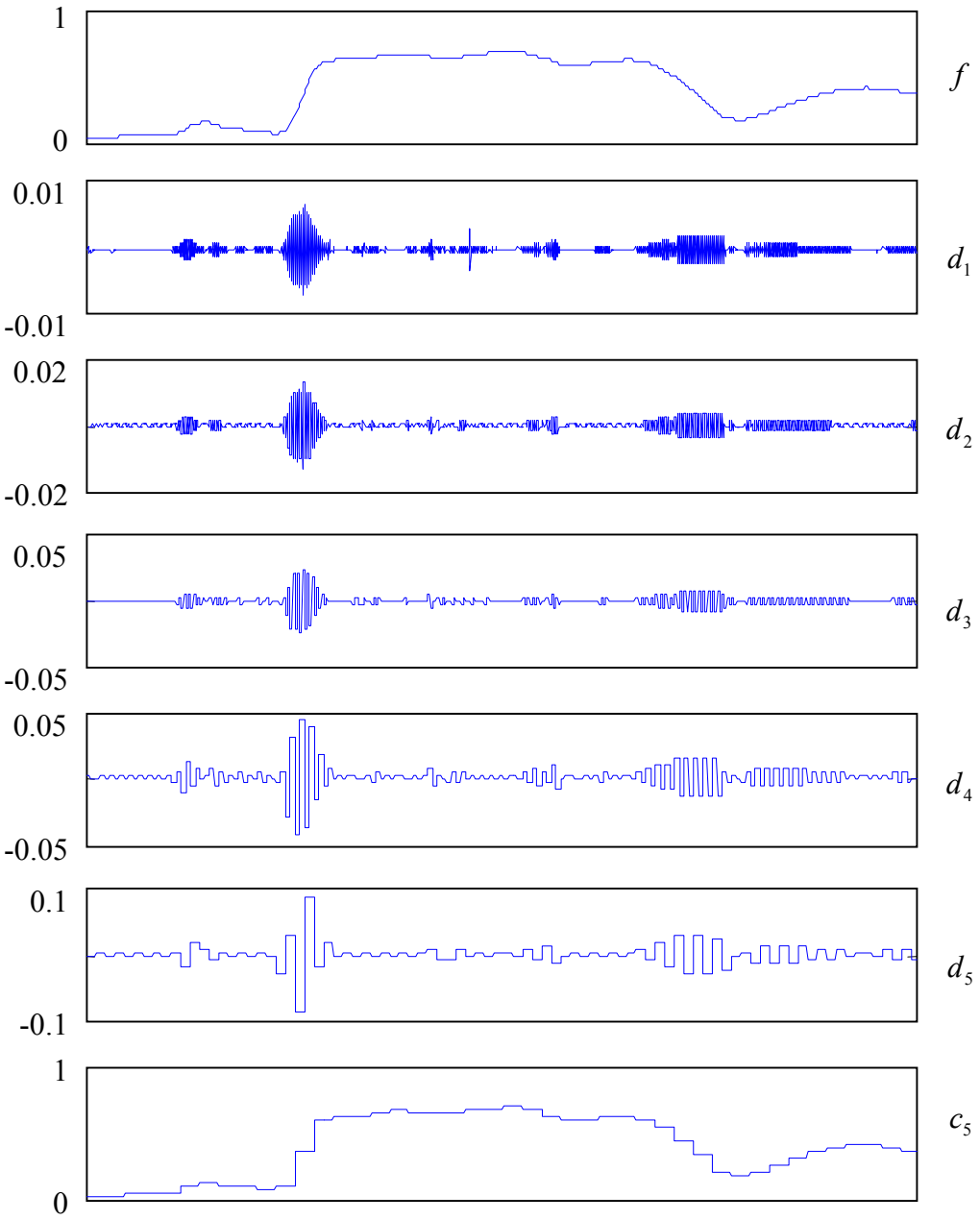

Figure 3.4. The reconstruction of wavelet coefficients in Figure 3.3.

3.2. DWT-Based Feature Extraction for Linear Unmixing of Hyperspectral Signals

Many typical wavelet bases, such as the Daubechies wavelet family, which includes the Haar wavelet as its length-2 member, are orthonormal and compactly supported [58, 59, 70, 71]. The wavelet transforms using these wavelet bases are orthonormal. Furthermore, these transforms are linear since equation 3.1 is satisfied. For the linear unmixing problem using the LMM and the LSE method, it is necessary to extract features after the wavelet analysis of hyperspectral signals. This is because of the fact that without feature extraction, only performing wavelet transform on hyperspectral signals does not help to improve linear unmixing performance. In general, there exists the following theorem:

Theorem 3.1: An orthonormal linear transform does not change the least square solution of abundance estimation using the linear mixture model.

Proof: In general, an orthonormal linear transform, of vector $\vec{a}$ is defined as,

$$\vec{b} = U\vec{a}, \tag{3.18}$$

where $\vec{b}$ is the representation of $\vec{a}$ in the transform domain, and U is an orthonormal matrix which has the following property:

$$U^{-1} = U^T, \text{ or } U^T U = I, \tag{3.19}$$

where superscript T is referred to as a matrix transpose operation, and I is an identity matrix. Applying the orthonormal linear transform, U, to the linear mixture model (LMM) defined in equation (1.1), we have the LMM expression in the transform domain,

$$U\vec{y} = UA\vec{x}_o + U\vec{e}_o . \tag{3.20}$$

According to the least square solution of abundance estimation defined in equation (2.7), using the transform domain expression of the LMM, equation (3.20), the least square solution of abundance estimation can be computed as,

$$\begin{aligned} \vec{x}_{LS} &= ((UA)^T UA)^{-1} (UA)^T U\vec{y} \\ &= (A^T U^T UA)^{-1} A^T U^T U\vec{y} \end{aligned} . \tag{3.21}$$

Using the orthonormal property of transform matrix, U, defined in equation (3.19), equation (3.21) can be further reduced to,

$$\vec{x}_{LS} = (A^T A)^{-1} A^T \vec{y} , \tag{3.22}$$

which is equivalent to equation (2.7). That is, an orthonormal linear transform does not change the least square solution of abundance estimation using the linear mixture model. In Chapter V, experimental results will also be provided to further verify this theorem.

The DWT using an orthonormal wavelet basis is a linear orthonormal transform. Based on Theorem 3.1, simply using the wavelet transform results of hyperspectral signals, *i.e.*, using all the wavelet transform coefficients, will not help to improve the abundance estimation. Moreover, it does not reduce the dimensionality of hyperspectral signals either, which is another motivation factor for feature extraction. Therefore, it is necessary to extract features following the wavelet analysis in order to possibly improve the performance of abundance estimation. A large variety of choices of features could be extracted from the wavelet analysis results, such as the energy of wavelet transform coefficients, the wavelet transform coefficients themselves, or any combination of the coefficients. Considering the linear unmixing problem, it is important to ensure the LMM

remains after the wavelet analysis and feature extraction. Otherwise, the linear unmixing problem will no longer exist. For example, a nonlinear wavelet energy feature can be extracted from the wavelet transform coefficients in the following manner,

$$\vec{F}_E = [D_1^E, D_2^E, \cdots, D_S^E, C_S^E]^T, \tag{3.23}$$

where $\vec{F}_E$ is referred to as a wavelet energy feature vector, the superscript T is referred to as a vector transpose, and S is the maximum scale (or coarsest scale) of discrete wavelet decomposition. D_j^E is referred to as the root mean square energy of the wavelet detail coefficient D_j, for $j = 1,2,\cdots,S$, and is computed as,

$$D_j^E = \sqrt{\frac{1}{K_j}\sum_{k=0}^{K_j-1}[D_j(k)]^2}, \tag{3.24}$$

where K_j is the number of elements in the wavelet detail coefficient D_j. C_S^E is referred to as the root mean square energy of the wavelet approximation coefficient C_S at the coarsest scale S, and can be similarly computed using equation (3.24). The use of the nonlinear wavelet energy feature, $\vec{F}_E$, has proven successful in our previous research for the target detection and classification problems [5-6, 39-41]. However, it may not be a good choice for the linear unmixing problem any more, since it is obvious that the LMM no longer exists when using the nonlinear wavelet energy feature. Again, in Chapter V, it will be experimentally shown that the use of the wavelet energy feature actually reduces the abundance estimation performance.

Therefore, linear features are preferred in the linear unmixing problem. Since wavelet transform is a linear transform, the wavelet transform coefficients at specific

scales could be directly used as linear features. That is, a linear wavelet feature, $\vec{F}_L$, could be formed using the wavelet detail coefficients D_j,

$$\vec{F}_L = [D_j(0), D_j(1), \cdots, D_j(K_j - 1)]^T, \tag{3.25}$$

or using the wavelet approximation coefficients C_j,

$$\vec{F}_L = [C_j(0), C_j(1), \cdots, C_j(K_j - 1)]^T, \tag{3.26}$$

where K_j is the number of the wavelet detail coefficients or the wavelet approximation coefficients at scale j, and the superscript T is referred to as a vector transpose. The use of linear wavelet features, $\vec{F}_L$, does not change the LMM, and thus the linear unmixing problem still exists. The reason for directly using the scalar subsets of DWT coefficients, *i.e.*, the DWT detail or approximation coefficients at specific scales, is because of the MRA property of the wavelet transform, discussed in previous section. The scalar subsets of DWT coefficients are the direct results from the MRA of hyperspectral signals and provide the direct insights into both global and fine information in hyperspectral signals at various resolutions. As a result, the use of the DWT scalar subset features, $\vec{F}_L$, also provide a potential for improving the abundance estimation of endmembers using the LMM and the LSE technique. To use the linear wavelet features, $\vec{F}_L$, in the LMM of equation (1.1), we simply substitute the mixed pixel vector, $\vec{y}$, and the endmember matrix, A, by their respective linear wavelet features. Then, abundances are estimated using the same LSE approach.

Note that the dimensionality of the linear wavelet features, $\vec{F}_L$, *i.e.*, the number of wavelet detail or approximation coefficients, D_j or C_j, decreases with the increase of wavelet decomposition scale. Theoretically, the number of elements in D_{j+1} or C_{j+1} is half of the number of elements in D_j or C_j. Thus, the use of the linear wavelet features is also associated with a dimensionality reduction of hyperspectral signals. As for the dimensionality of the wavelet energy feature, $\vec{F}_E$, its dimensionality is the maximum wavelet decomposition scale plus one, based on its definition in equation (3.23). For a hyperspectral signal of length 1400, the possible maximum wavelet decomposition scale is 10, and thus the dimensionality of feature vector $\vec{F}_E$ is 11, which is much less than the dimensionality of 1400 of the original hyperspectral signal. It is the DWT followed by the feature extraction that reduces the data volume (or dimensionality) of hyperspectral signals, and thus possibly reduces the computational cost of hyperspectral signal analysis.

The linear wavelet features, $\vec{F}_L$, defined in equations (3.25) and (3.26), are investigated in this study for improving the abundance estimation performance. At a specific wavelet decomposition scale, two sets of linear wavelet features are formed. One set is from the wavelet detail coefficients, and the other is from the wavelet approximation coefficients, according to equations (3.25) and (3.26). Thus, for a L-scale DWT, there are $2L$ sets of linear wavelet features, $\vec{F}_L$. Among these features, generally there exists one set of features that results in the best abundance estimation, which is referred to as an optimal set of linear wavelet features in this study. Determining the optimal DWT feature set leads to an optimization problem. Thus, an optimization

criterion needs to be determined firstly. In Chapter II, the mean square error (MSE) of abundance estimate is used as a criterion to evaluate the abundance estimation performance. Thus, it is proposed to be an optimization criterion to determine the optimal set of linear wavelet features. That is, the one producing the smallest MSE value is regarded as optimal set of linear wavelet features.

This chapter discussed the DWT-based feature extraction and feature selection. Chapter II discussed the possibility using DWT-based features to improve the LSE of abundances. Based on these discussions, a DWT-based linear unmixing system is designed and implemented to both experimentally verify the proposed approaches and practically investigate the feasibility of applying the proposed approaches to solve real linear unmixing problems. Details about the system design and implementation, as well as the system performance evaluation, will be provided in the next chapter.

CHAPTER IV

A DISCRETE WAVELET TRANSFORM BASED LINEAR UNMIXING SYSTEM FOR LEAST SQUARES ESTIMATION OF ABUNDANCES

A linear unmixing system based on feature extraction using the discrete wavelet transform (DWT) is designed and implemented. The system is specifically designed for the abundance estimation of a mixed pixel spectrum. That is, the system takes a mixed pixel spectrum as input. Assuming that the endmember spectra are known, the system outputs an estimate of the endmember abundances. The proposed DWT-based linear unmixing system basically consists of two modules. One is the pre-processing module. This module implements the DWT and the feature extraction of the input mixed pixel spectrum and the known endmember spectra. The following module is the abundance estimation, in which a constrained least squares estimation (CLSE) technique using a quadratic programming (QP) algorithm is performed to implement the abundance estimation. A system block diagram is shown in Figure 4.1.

The proposed system is a supervised linear unmixing system. The system needs to be trained to determine an optimal set of DWT-based features. The system training is performed on a set of training data, which consist of the mixed pixel spectra with known abundances and the endmember spectra. The system outputs, *i.e.*, the estimated abundances, are compared with the true abundances, and the estimation errors are computed. Among various sets of DWT-based features, the one set producing the

smallest estimation error is regarded as an optimal set of DWT-based features. For this study, since the wavelet decomposition detail or approximation coefficients at each specific scale are utilized as features, as defined in equations (3.25) and (3.26), an optimal feature set also means to record two optimal parameters. One is the optimal wavelet decomposition scale, and the other is the detail or the approximation features. Note that the optimum is based on the given training data. For different sets of training data, the optimal results may not be different. However this is the essential idea of the supervised system. Thus, it is critical for a supervised system to obtain good training data, which can realistically represent the investigated situation.

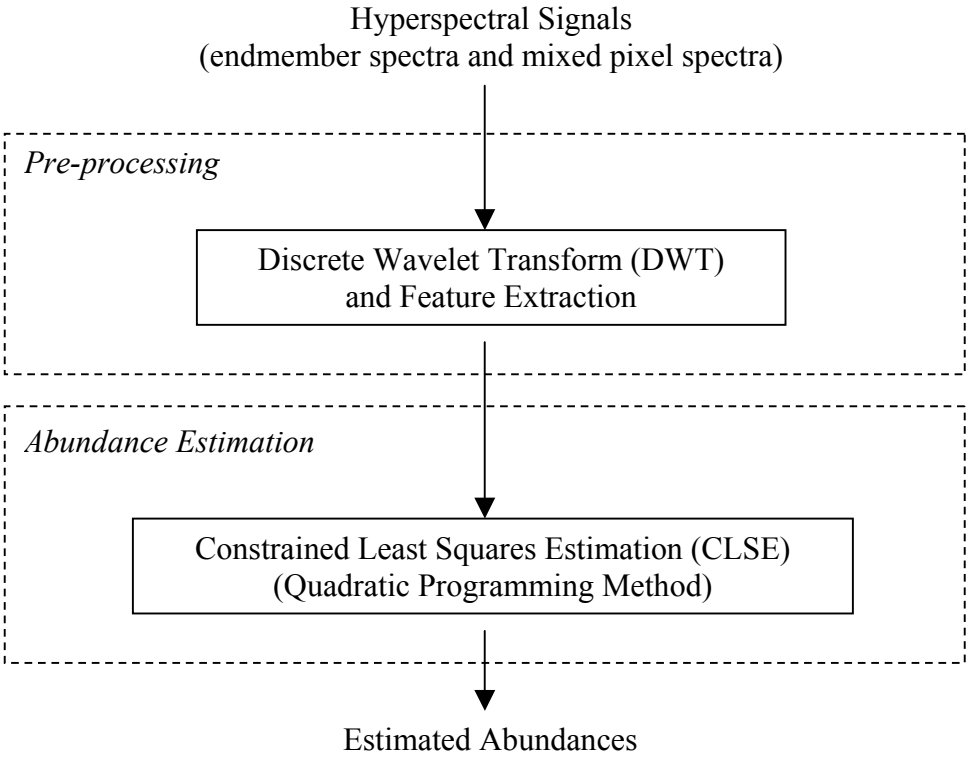

Figure 4.1. A DWT-based linear unmixing system block diagram.

The optimal feature determined in the system training phase will usually be tested to investigate how well it can work in the practical applications. Thus, a system testing procedure is performed. Another set of testing data are prepared, which are mutually exclusive from the training data. That is, the testing data do not include any information in the training data. This could avoid the bias introduced by the use of any information from the training data. In the system testing phase, the two optimal parameters recorded in the training phase, corresponding to the optimal DWT-based feature set for given training data, will be used to directly obtain the DWT-based features for testing data. Then abundance estimation is performed using the DWT-based features. Generally, the estimated results are quantitatively and/or qualitatively evaluated to investigate the estimation system performance. For this study, three quantitative evaluation metrics are used for the system performance evaluation: (i) the root mean square error of abundance estimation; (ii) the confidence of abundance estimation; and (iii) the abundance distribution diagram. These quantitative evaluation metrics, as well as the two modules of the system will be introduced in the following three sections in detail.

4.1. Comparison of Various Pre-processing Methods

The pre-processing is the first and core step of the proposed DWT-based linear unmixing system, because the proposed DWT and feature extraction approaches are implemented in this step. In Chapters II and III, we argued that this pre-processing step not only can reduce the computational cost in the next step of abundance estimation, but also more importantly provides the potential for improving the performance of abundance estimation. To show the possible improvement of abundance estimation provided by

using the DWT-based features, the abundances are also estimated using the original hyperspectral signals without any pre-processing. The results are compared with the abundance estimates using the DWT-based features.

Note that the DWT followed by the feature extraction is not the only pre-processing method. Two other pre-processing methods are also investigated in this study for the purpose of comparison. One is based on the discrete cosine transform (DCT), and the other is based on the principal components analysis (PCA) method, or the Karhunen-Loève transform (KLT). Both the DCT and the PCA are linear orthonormal transform methods, which are extensively used in signal and image processing applications such as data compression and dimensionality reduction [1, 2, 52, 63, 72].

In general, the DCT of a finite energy sequence with N samples, $f(n)$, is defined as [63, 72],

$$Y(k) = \alpha_k \sum_{n=0}^{N-1} f(n) \cos\left(\frac{(2n+1)k\pi}{2N}\right), \quad 0 \le k \le N-1 \tag{4.1}$$

where $Y(k)$ are DCT coefficients, and α_k are constant numbers defined as,

$$\alpha_k = \begin{cases} \sqrt{1/N}, & k = 0 \\ \sqrt{2/N}, & 1 \le k \le N-1 \end{cases}. \tag{4.2}$$

The DCT is a real and orthonormal linear transform, and has excellent energy compaction for many applications. For example, the two-dimensional DCT has a property of excellent energy compaction for natural images, and thus has been adopted in the Joint Photographic Experts Group (JPEG) international standard for the still image compression [73]. Figure 4.2 shows an example of the DCT of a typical vegetation

hyperspectral reflectance curve. The curve is the same as the one used in the DWT example of Figure 3.3. Figure 4.2(a) shows the original hyperspectral signal and Figure 4.2(b) shows its DCT coefficients. Specifically, Figure 4.2(c) shows the first 50 DCT coefficients. It can be seen the first 7 coefficients have relatively higher amplitude than others, and the first 30-40 coefficients include nearly all the energy of the signal. These observations clearly show the excellent energy compaction property of the DCT for hyperspectral signals, which also implies that excluding higher frequency DCT coefficients will not lose most information in the signal, while achieving the purpose of data dimensionality reduction.

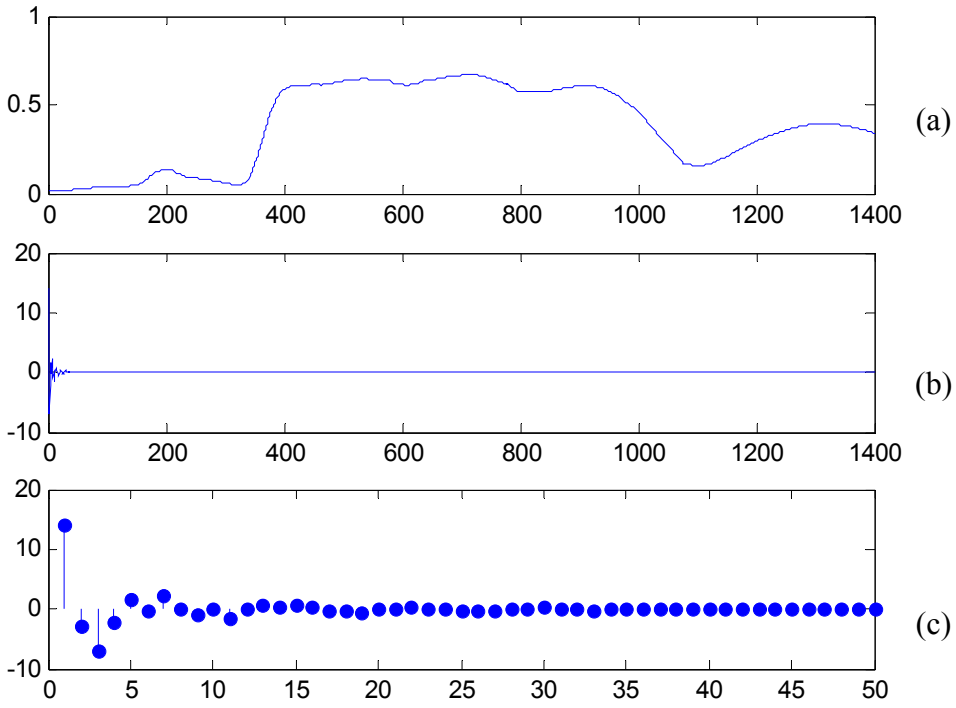

Figure 4.2. An example of the DCT of a hyperspectral signal:
(a) original hyperspectral signal; (b) DCT coefficients;
and (c) The first 50 samples of the DCT coefficients.

The KLT of a finite energy sequence with N samples, $f(n)$, is defined as [63, 72],

$$Z(k) = \sum_{n=0}^{N-1} f(n)\phi_k(n), \quad 0 \le k \le N-1 \tag{4.3}$$

where $Z(k)$ are KLT coefficients, and $\phi_k(n)$ are the eigenvector sequences of the correlation matrix, R_f, of signal $f(n)$, *i.e.*, there exists the following relationship,

$$R_f \vec{\phi}_k = \gamma_k \vec{\phi}_k, \quad 0 \le k \le N-1 \tag{4.4}$$

where $\vec{\phi}_k = [\phi_k(0), \phi_k(1), \cdots, \phi_k(N-1)]^T$ are vector forms of eigenvector sequences $\phi_k(n)$, and γ_k are the corresponding eigenvalues of correlation matrix, R_f.

If we arrange $\phi_k(n)$ in a decreasing order of magnitude of γ_k and compute the KLT coefficients using equation (4.3), then $Z(k)$ are also referred to as principal components of signal $f(n)$. Thus, the KLT of signal is also called the PCA of signal. The PCA has many good properties. First of all, the principal components, or KLT coefficients, are uncorrelated, since the correlation matrix, R_Z, of principal components $Z(k)$ has the following form,

$$R_Z = \begin{bmatrix} \gamma_0 & 0 & \cdots & 0 \\ 0 & \gamma_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \gamma_{N-1} \end{bmatrix}, \tag{4.5}$$

where $\gamma_0 > \gamma_1 > \cdots > \gamma_{N-1}$ are eigenvalues of correlation matrix, R_f, of signal $f(n)$. Note that off-diagonal elements of R_Z are all zero, *i.e.*, the principal components, or KLT

coefficients, are uncorrelated. Secondly, note that according to the definition of correlation matrix, main-diagonal elements of R_Z, *i.e.*, eigenvalues γ_k of correlation matrix R_f, are energy of principal components, or KLT coefficients. Since $\gamma_0 > \gamma_1 > \cdots > \gamma_{N-1}$ and typically $\gamma_0 >> \gamma_{N-1}$, the PCA also have a property of excellent energy compaction, *i.e.*, most of energy is packed into the first several principal components. Thirdly, due to the excellent energy compaction property, *i.e.*, $\gamma_0 >> \gamma_{N-1}$, excluding PCA coefficients associated with smaller eigenvalues will not lose most information in the signal, but achieve the purpose of reducing the data dimensionality. Figure 4.3 show an example of PCA of a typical vegetation hyperspectral reflectance curve. The curve is the same as one used in the DWT and DCT examples of Figures 3.3 and 4.2. Figure 4.3(a) shows the original hyperspectral signal and Figure 4.3(b) shows its PCA coefficients. Specifically, Figure 4.3(c) shows the first 50 PCA coefficients. It can be seen the first 5-10 coefficients include nearly all the energy of the signal. The first 2 coefficients have relatively higher amplitude than others, and especially the first coefficient has much higher amplitude than the rest. These observations indicate that the PCA shows a better energy compaction performance than the DCT for hyperpsectral signals.

Both the DCT and the PCA are linear orthonormal transforms. Thus, based on *Theorem 3.1*, simply applying all DCT or PCA coefficients to the abundance estimation will not help improving the estimation performance. Thus features have to be extracted from the DCT or PCA coefficients. Similar to the proposed DWT-based feature extraction, the DCT and PCA coefficients are directly used as features in order to retain

the validity of the LMM after feature extraction. Conventaionlly, the DCT- and PCA-based features are formed using the first several DCT and PCA coefficients, since most of the energy, or most of the information, of the original signal concentrates in the first few coefficients. Taking the PCA as an example, the first feature set would be $[Z(0)]^T$, the second one would be $[Z(0), Z(1)]^T$, and so on. Note that this conventional method works well when the aim of feature extraction is signal representation, such as the case of signal compression. However, when the aim of feature extraction is signal classification, such as the case of endmember abundance estimation, differences between signals take on importance, and simply using the first few large-amplitude coefficients may not be adequate. Therefore, an alternative approach is proposed for selecting a subset of transform coefficients. This alternative approach utilizes a sliding window of size L to select coefficient subsets of size L. Specifically, taking the PCA as an example, the first feature set would be $[Z(0), Z(1), \cdots, Z(L)]^T$, the second one would be $[Z(L+1), Z(L+2), \cdots, Z(2L)]^T$, and so on. That is, the shifting stepsize of the sliding window is also defiend as L in this study.

With all these PCA- and DCT-based feature sets obtained from the use of conventional and alternative appraoches, a feature optimization (or selection) problem exists, as in the situation of the DWT-based feature extraction. The same optimization criterion, as used for the optimal DWT-based feature selection, is used for the selection of the optimal DCT-based or PCA-based feature set. That is, among all sets of these features, the one producing the smallest abundance estimation error is regarded as the optimal feature set.

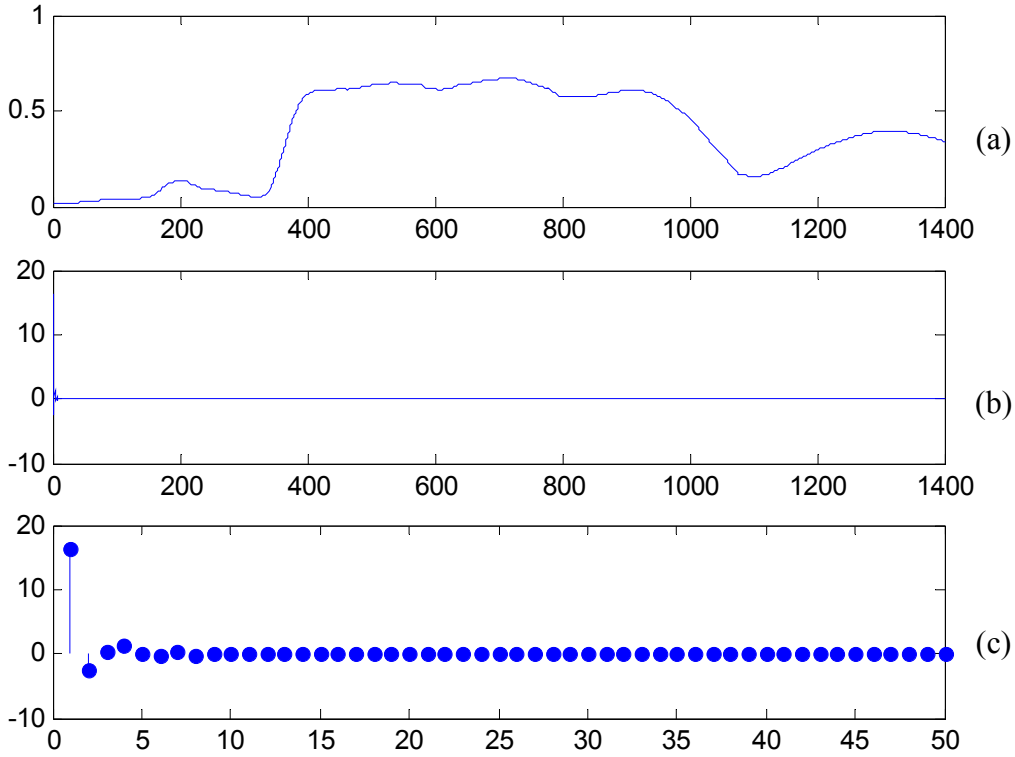

Figure 4.3. An example of the PCA of a hyperspectral signal:
(a) original hyperspectral signal; (b) PCA coefficients;
and (c) The first 50 samples of the PCA coefficients.

In summary, three pre-processing methods are investigated and compared in this study: i) DWT-based feature extraction; ii) DCT-based feature extraction; iii) PCA-based feature extraction. In addition, the direct use of original hyperspectral signals, without any pre-processing operation, is investigated and compared as well. Note that for the DCT- and PCA-based pre-processing, both conventional and alternative approaches are investigated. For the DWT-based pre-processing, we focus on using one type of mother wavelet for the purpose of the simplicity and effectiveness, though there exist infinite number of mother wavelets. The mother wavelet used in the study is the Haar mother

wavelet, as introduced and discussed in Chapter III. The major reason for using the Haar mother wavelet is because it has been shown to be very successful in our previous research of target detection and classification applications [5-6, 39-41]. For example, in [5, 6], several commonly used orthogonal and biorthogonal mother wavelets, such as Daubechies, Coiflets, Symlets and B-splines biorthogonal wavelets [71], were investigated for the subpixel target detection. It was found that on average the Haar mother wavelet performed as well as or better than other investigated mother wavelets.

4.2. Constrained Least Squares Estimation

After the pre-processing, features extracted from endmember spectra and mixed pixel spectra are used to implement the least squares estimation (LSE) of abundances. In the practical implementation of the abundance estimation using LSE, to make the estimated abundances physically meaningful, two constraints are applied to LSE, forming the constrained LSE (CLSE). One constraint is nonnegativity, *i.e.*, physically abundances should not be negative numbers. The other constraint is sum-to-one, *i.e.*, it is assumed that the mixed pixel completely consists of the endmembers used for abundance estimation, and thus the sum of abundances is one. These two constraints can be mathematically expressed as,

$$x_i \geq 0, \quad i = 1,2,\cdots,M \tag{4.6}$$

$$\sum_{i=1}^{M} x_i = 1 \tag{4.7}$$

where M is the number of endmembers and x_i (for $i = 1,2,\cdots,M$) is the estimated abundance corresponding to the i^{th} endmember.

Typically, the CLSE of endmember abundances is implemented using the quadratic programming (QP) method [74], which is an iterative technique for solving constrained optimization problems. Based on the LMM defined in equation (1.1) and the LSE derivations defined in equations (2.1) to (2.7), the CLSE using the QP method may be used to obtain an optimal estimation of abundances by minimizing the following quadratic objective function,

$$O(\vec{x}) = -2\vec{x}^T A^T \vec{y} + \vec{x}^T A^T A\vec{x}\,, \text{ subject to } \vec{b}^T\vec{x} = 1 \text{ and } x_i \geq 0\,, \tag{4.8}$$

where A is the endmember spectrum matrix defined in equation (1.5), $\vec{y}$ is the mixed pixel spectrum vector defined in equation (1.2), $\vec{b} = \vec{1}$ is a 1 vector, and $\vec{x} = [x_1, x_2, \cdots, x_M]^T$ is an estimated abundance vector. The constraint $\vec{b}^T\vec{x} = 1$ is equivalent to the constraint in equation (4.7). In general, the quadratic objective function, $O(\vec{x})$, is reduced after each iteration, and an optimal CLSE of endmember abundances can be obtained within a finite number of iterations.

Note that, according to the LSE derivations defined in equations (2.1) to (2.7), minimizing the quadratic objective function $O(\vec{x})$ is essentially minimizing the total error energy of abundance estimation defined in equation (2.2). Also note that $R = A^T A$ is an auto-correlation matrix of endmember spectra, and $\vec{d} = A^T\vec{y}$ is a cross-correlation vector of endmember spectra and mixed-pixel spectrum. Thus, equation (4.8) can be expressed using the correlation matrix,

$$O(\vec{x}) = -2\vec{x}^T\vec{d} + \vec{x}^T R\vec{x}\,. \tag{4.9}$$

Compared to the CLSE, the LSE without any constraint is referred to as the unconstrained LSE (ULSE) in this study. Note that in the DWT-based linear unmixing system shown in Figure 4.1, if we substitute the CLSE by the ULSE, then the abundance estimation results are the ULSE of abundances, which can be directly calculated using equation (2.7). For this study, experiments using both CLSE and ULSE are implemented. The results and discussion are provided in the next chapter.

4.3. System Performance Evaluation

The output of the CLSE is an optimal abundance estimate, $\vec{x}_{CLS}$, under constraints of nonnegativity and sum-to-one,

$$\vec{x}_{CLS} = [x_{CLS1}, x_{CLS2}, \cdots, x_{CLSM}]^T, \tag{4.10}$$

where M is the number of endmembers and superscript T represents a vector transpose. The abundance estimation error exists between $\vec{x}_{CLS}$ and the true abundance, $\vec{x}_o$, defined in equation (1.3). By evaluating the estimation error, the performance of the proposed DWT-based linear unmixing system is investigated. In this study, three quantitative evaluation metrics for abundance estimation errors are proposed, which are valid for both CLSE and ULSE cases. For the convenience of mathematical description, a general expression of the abundance estimation result for both cases is used,

$$\vec{x}_{LS} = [x_{LS1}, x_{LS2}, \cdots, x_{LSM}]^T, \tag{4.11}$$

where M is the number of endmembers and superscript T represents a vector transpose.

The first quantitative evaluation metric is the root-mean-square error (RMSE) of abundance estimation. For a given mixed pixel spectrum, each of endmembers

constituting the mixed pixel has its own abundance, x_{LSi} (for $i = 1,2,\cdots,M$), and thus its own abundance estimation error, $(x_{LSi} - x_{oi})$, where x_{LSi} and x_{oi} are defined in equations (4.11) and (1.3), respectively. For simplicity and effectiveness, the average value of abundance estimation error corresponding to each of the endmembers is utilized to represent the abundance estimation error of the mixed pixel. As a result, the error energy, Θ, of the abundance estimation for a given mixed pixel is defined as,

$$\Theta = \frac{1}{M}\sum_{i=1}^{M}(x_{LSi} - x_{oi})^2 , \tag{4.12}$$

where M is the number of endmembers. Suppose that Θ_k (for $k = 1,2,\cdots,K$), represents the error energy of the abundance estimation corresponding to the k^{th} mixed pixel in a set of K mixed pixel spectra. Then, the RMSE, Ω, of abundance estimations for all the mixed pixel spectra in the set is defined as,

$$\Omega = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\Theta_k} . \tag{4.13}$$

In general, Ω indicates an average deviation of the abundance estimate from the true abundance. For this study, Ω is utilized to evaluate the system performance from a global perspective, *i.e.*, indicating the estimation deviation of the system on the average. It is also utilized as a criterion to determine an optimum feature during the system training phase. Note that the square value of the RMSE, Ω^2, is essentially equivalent to the mean square error (MSE), Γ_x, defined in equation (2.10). The only difference in the concept is that in the definition of Ω a sample mean corresponding to K samples in the set of mixed pixel spectra is computed. However, in the definition of Γ_x, a population

mean corresponding to infinite samples of mixed spectra is defined. That is, Ω^2 is a practical approximation of Γ_x for limited number of samples of mixed pixel spectra. When the number of samples goes to infinity, Ω^2 tends to Γ_x.

The second quantitative evaluation metric is the confidence of abundance estimation, which is defined as a probability of correct estimation given an error interval of abundance estimate. Given a mixed pixel spectrum, if its abundance estimations and the corresponding true abundances are x_{LSi} and x_{oi} (for $i = 1,2,\cdots,M$), defined in equations (4.11) and (1.3) respectively, then an average absolute error, η, between the estimated abundances and the true abundances is defined as,

$$\eta = \frac{1}{M}\sum_{i=1}^{M} | x_{LSi} - x_{oi} | , \tag{4.14}$$

where M is the number of endmembers. Given an error interval, ε, if $\eta \le \varepsilon$, then it is said that the abundance of the mixed pixel is correctly estimated. Otherwise, it is incorrectly estimated. For the set of K mixed pixel spectra, if there are $P \le K$ mixed pixel spectra satisfying $\eta \le \varepsilon$, *i.e.*, the abundances of P mixed pixels are correctly estimated, then the probability, $\rho(\varepsilon)$, of correct estimation given an error interval, ε, of abundance estimate can be computed as,

$$\rho(\varepsilon) = \frac{P}{K}. \tag{4.15}$$

Note that $\rho(\varepsilon)$ defines a system confidence of abundance estimation. That is, given an estimation error interval, $\rho(\varepsilon)$ is a measure of how confidently the system produces an abundance estimate which has an error less than the given estimation error interval.

Also note that according to the definition of the confidence and error interval, there are $0 \le \rho(\varepsilon) \le 1$ and $0 \le \varepsilon \le 1$. Also note that when the error interval, ε, increases from 0 to 1, the confidence, $\rho(\varepsilon)$, increases from 0 to 1. That is, the confidence is a monotonically increasing function of the error interval. This is similar to a cumulative density function (CDF) of a random variable, where the random variable is ε and the CDF is $\rho(\varepsilon)$. A confidence curve can be formed when the error interval varies form 0 to 1. A larger confidence value associated with a smaller error interval indicates a more accurate abundance estimate.

The third quantitative evaluation metric of abundance estimation is the abundance distribution diagram. The diagram shows both the true abundances and the estimated abundances in two-dimensional plane, where the horizontal coordinate represents the number of given mixed spectra and the vertical coordinate represents the abundance associated with an endmember. Thus, for a M -endmember linear unmixing problem, M abundance distribution diagrams can be formed. In general, the abundance distribution diagram provides a visual and direct evaluation of how close the abundance estimation is to the truth of abundances.

In summary, this chapter introduces the design and implementation of a DWT-based linear unmixing system, as well as three quantitative metrics for the evaluation of the system performance. To practically investigate the performance of the proposed system, several sets of experiments are performed on practical hyperspectral signals. Details about the experiment design, results and discussions are presented in the next chapter.

CHAPTER V

EXPERIMENTS, RESULTS AND DISCUSSION

Generally, experiments of the linear unmixing can be divided into two broad categories. One category is the two-endmember linear unmixing, where it is assumed that the mixed pixel consists of two endmembers. That is, it is assumed that the region of interest (ROI) is covered only by two distinct ground-cover materials. For example, in the agriculture applications, the ROI may consist of the certain crop and the bare soil. In the target detection applications, it could be assumed that the ROI simply consists of the 'target' and 'non-target'. The other category is the multi-endmember linear unmixing, where it is assumed that the mixed pixel consists of more than two endmembers. This is a more typical case. For example, in the agriculture applications, the ROI may consist of the certain crop, the bare soil and typical weeds. In the forest monitoring applications, the ROI may consist of several different forest species. In the target detection applications, the ROI may consist of multiple targets.

For this study, the experiments take an agriculture application as an example case for the study of linear spectral unmxing. The investigated ground cover materials are soybean, grass and soil. The experiment data are hyperspectral reflectance signals of these three types of materials. In general, three sets of experiments are designed and implemented to evaluate the proposed DWT-based linear spectral unmixing system. Two of the three sets of experiments investigate the two-endmember linear unmxing problem,

and the third one investigates the multi-endmember linear unmixing problem. Specifically, the three sets of experiments are referred to as, respectively, Experiment I: soybean versus soil; Experiment II: soybean versus grass; and Experiment III: soybean, grass versus soil. Table 5.1 summarizes the information of the three sets of experiments. The details about the experiment data, design, results and discussion will be provided in the following sections.

Table 5.1. Information summary of the three sets of experiments.

Experiment Case	Endmembers
Experiment I (two-endmember case)	soybean and soil
Experiment II (two-endmember case)	soybean and grass
Experiment III (three-endmember case)	soybean, grass and soil

5.1. Preparation of Experiment Data

Hyperspectral reflectance spectra are measured using a handheld spectroradiometer from Analytical Spectral Devices (ASD), Inc. [75]. This ASD's instrument is called the FieldSpec Pro spectroradiometer, which has an ability to measure the electromagnetic radiance (and consequently derive the reflectance) in the wavelength range from 350nm to 2500nm. Typically, outputs of the instrument are hyperspectal reflectance curves of materials. Figure 5.1(a) shows a typical hyperspectral curve directly from the ASD's spectroradiometer. The curve is a hyperspectral reflectance spectrum of the soybean. Note that there are two regions where a large amount of noise exists. One region is between ~1350nm and ~1430nm, and the other is between ~1800nm and ~1950nm. These regions are typically called the water absorption bands, since most of the energy of the electromagnetic radiance in these regions is absorbed by water in the

atmosphere and in the ground cover materials. The parts of the hyperspectral signal in water absorption bands consist of the noise and do not include useful information. Therefore, samples in these regions are typically set to zeros, or interpolated using the samples immediately near the water absorption bands. For this study, the linear interpolation method is used to modify the parts of hyperspectral signals in the water absorption bands. Also note that after the second water absorption band, the signal is severely contaminated by noise. Therefore, the part of hyperspectral signals after the second water absorption band, together with the second water absorption band, is excluded for this study. That is, only the part of the hyperspectral signal before the second water absorption band, including the first water absorption bands interpolated, is used for the linear unmixing analysis in this study. Also note that singular samples appear in the first few spectral bands. This noise is introduced by the ASD's measurement instrument itself. Thus, the first few spectral bands of hyperspectral signals are excluded as well. As a result, the final hyperspectral signals utilized for the linear unmixing analysis in this study consist of 1400 samples ranging from 354nm to 1753nm, as shown in Figure 5.1(b).

The ground-cover materials investigated in the experiments include soybean (Glycine max), large crabgrass (Digitaria sanguinalis) and soil. The soil type is the Dundee silt loam, consisting of 26% sand, 56% silt and 18% clay. Using the ASD's spectroradiometer, a set of 60 hyperspectral reflectance spectra from the three ground-cover materials (20 spectra for each) were collected at the southern weed science research farm of the USDA Agricultural Research Service near Stoneville, Mississippi, USA, in

June, July and August, 2000. For each of the three ground-cover materials, half of the 20 spectra are utilized for the system training, and the other half are utilized for the system testing. That is, the testing data do not include any information from the training data, which ensure a fair testing. Figure 5.2 shows the 60 hyperspectral signals which are utilized for the linear unmixing analysis in the dissertation. These signals are modified versions of the raw hyperspectral signals directly from the ASD's spectroradiometer, by interpolating the first water absorption band and excluding the noise-contaminated parts of signals.

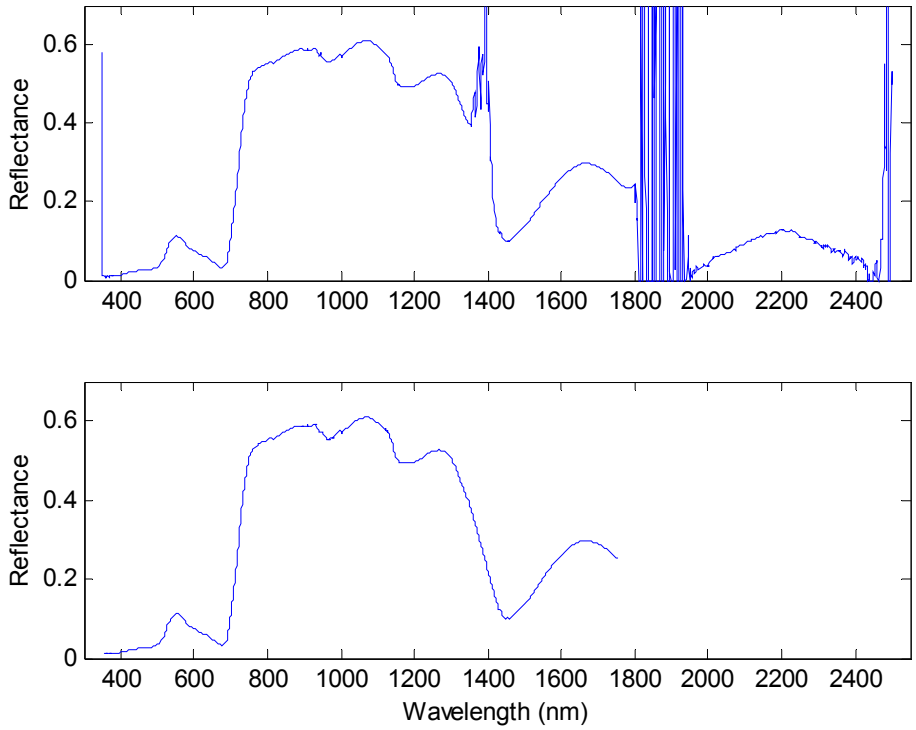

Figure 5.1. An example of ASD's hyperspectral reflectance signal:
(a) raw output signal from ASD's spectroradiometer;
(b) modified version of the signal by interpolating
the water absorption band and excluding the
noise-contaminated parts of the signal.

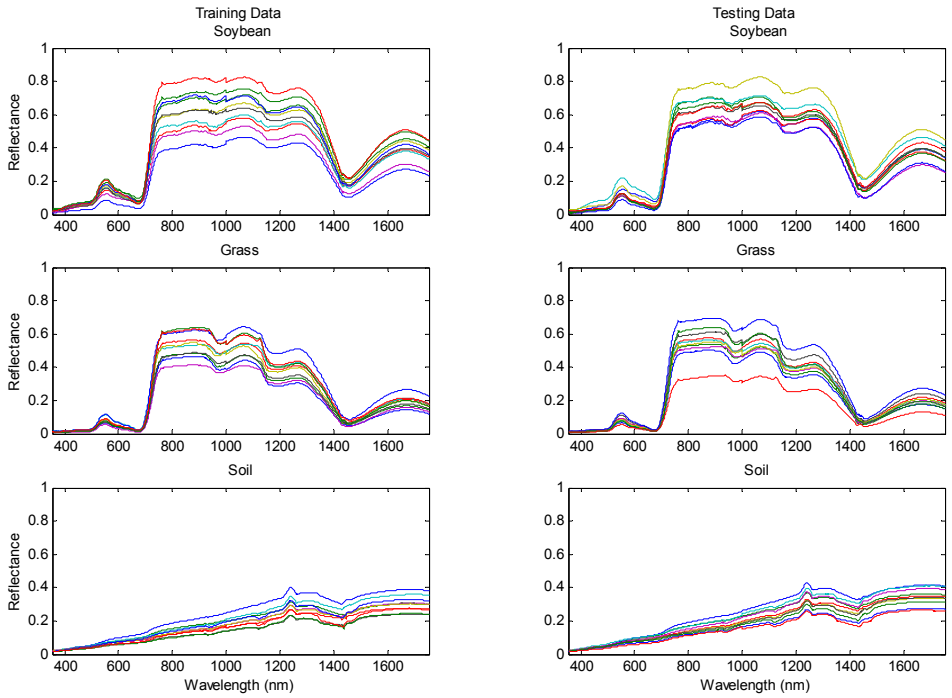

Figure 5.2. The ASD's hyperspectral reflectance signals of soybean, Grass and soil utilized for the linear unmixing analysis in the dissertation.

Note that the hyperspectral reflectance spectra from the handheld ASD's spectroradiometer can be regarded as pure pixel spectra, because the handheld measurements are made such that only one type of material is in the sensor's field of view. Moreover, the handheld measurements also reduce the atmospheric influence to the lowest limit, unlike the measurement using the spaceborne or airborne sensors where a large amount of atmospheric influence exists. Using these pure pixel spectra, mixed pixel spectra with known abundances can be synthesized. The main advantage of using the synthesized data is that the true abundances are known, which makes it possible to

implement a quantitative analysis of abundance estimation performance using the DWT-based linear unmixing system.

The mixed pixel spectra are synthesized in the following way. For the two-endmember case, given two endmember spectra, $\vec{a}_{r1}$ and $\vec{a}_{r2}$, and their abundances, x_{o1} and x_{o2}, based on the LMM defined in equation (1.1), a mixed pixel spectrum, $\vec{y}$, can be synthesized as,

$$\vec{y} = x_{o1}\vec{a}_{r1} + x_{o2}\vec{a}_{r2}. \tag{5.1}$$

Note that unlike the definition of the LMM in equation (1.1), there is no error term in equation (5.1). This is because the data synthesis assumes that $\vec{a}_{r1}$ and $\vec{a}_{r2}$ are real endmembers forming the mixed pixel $\vec{y}$, *i.e.*, the error is zero. Also note that according to the physical constraints on the abundances, as defined in equations (4.6) and (4.7), there exists the following relationship between abundances x_{o1} and x_{o2},

$$x_{o1} + x_{o2} = 1. \tag{5.2}$$

When the values of x_{o1} and x_{o2} vary from 0.0 to 1.0 with an increment of 0.1, a set of 11 mixed pixel spectra with 11 sets of different abundances are synthesized, excluding the possibly repeated combinations due to the relationship of equation (5.2). Similar relationships exist for the three-endmember case,

$$\vec{y} = x_{o1}\vec{a}_{r1} + x_{o2}\vec{a}_{r2} + x_{o3}\vec{a}_{r3}, \tag{5.3}$$

$$x_{o1} + x_{o2} + x_{o3} = 1, \tag{5.4}$$

where $\vec{a}_{r1}$, $\vec{a}_{r2}$ and $\vec{a}_{r3}$ represent the three real endmember spectra forming the mixed pixel $\vec{y}$, and x_{o1}, x_{o2} and x_{o3} are their corresponding abundances. Similarly, when the values of x_{o1}, x_{o2} and x_{o3} vary from 0.0 to 1.0 with an increment of 0.1, a set of 66 mixed pixel spectra with 66 sets of different abundances are synthesized, excluding the possibly repeated combinations due to the relationship of equation (5.4).

The training data set includes 10 collected pure pixel spectra for each of the three materials: soybean, soil and grass. That is, there are 10 sets of $\vec{a}_{r1}$, $\vec{a}_{r2}$ and $\vec{a}_{r3}$. For the two-endmember case, since 11 mixed pixel spectra with different abundances can be produced for each set of pure pixel spectra, a total of 110 mixed pixel spectra can be synthesized for all 10 sets of pure pixel spectra. Similarly, for the three-endmber case, since 66 mixed pixel spectra with different abundances can be produced for each set of pure pixel spectra, a total of 660 mixed pixel spectra can be synthesized for all 10 sets of pure pixel spectra. The testing data set includes the other 10 pure pixel spectra for soybean, soil and grass, respectively. Using the same synthesis method, for system testing, another 110 mixed pixel spectra are produced for the two-endmeber case and another 660 mixed pixel spectra are produced for the three-endmeber case.

For the implementation of the abundance estimation using the LSE method, only one endmember spectrum for each endmember is needed. That is, in equation (1.1) the endmember matrix, A, consists of one entry of endmember spectrum for each endmember. Thus, for this study the endmember spectra are formed using the average of all collected pure pixel spectra, *i.e.*,

$$\vec{a}_i = \frac{1}{K}\sum_{k=1}^{K}\vec{a}_{ri,k} \ , \tag{5.5}$$

where $\vec{a}_i$ (for $i = 1,2,\cdots,M$), as defined in equation (1.6), is the i^{th} endmember spectrum for the i^{th} ground-cover material (or endmember), M is the number of endmembers, $\vec{a}_{ri,k}$ (for $i = 1,2,\cdots,M$ and $k = 1,2,\cdots,K$) is the k^{th} pure pixel spectrum (and also real endmember spectrum forming a mixed pixel) collected from the i^{th} endmember, and K is the number of pure pixel spectra collected for the i^{th} endmember. Specifically, for the training or testing data set in this study, there are a total of 10 pure pixel spectra collected for each of endmembers, thus K is equal to 10 for the case. For the two-endmember and three-endmember cases, M are equal to 2 and 3, respectively. For this study, using the equation (5.5), the endmember spectra are formed only from the training data set. That is, the system testing uses the same endmember spectra formed in the system training. Figure 5.3 shows the formed endmember spectra of soybean, grass and soil, which are used for all the three sets of experiments in this study.

In the theoretical analysis of the abundance estimation error in Chapter II, it is argued that the error term in equation (1.1) can be generally described as stemming from the difference between the real endmember spectra constituting the mixed pixel spectrum and the library endmember spectra utilized in the LSE of abundances. This can be clearly observed from the proposed mixed pixel synthesis procedure and endmember formation procedure. The mixed pixel spectra are synthesized using the collected pure pixel spectra, $\vec{a}_{ri,k}$, and the endmember spectra, $\vec{a}_i$, are the average of all collected pure pixel spectra

from the training data set. Thus, a difference exists between $\vec{a}_{ri,k}$ and $\vec{a}_i$. This difference simulates the general difference resulting in the error term in equation (1.1).

In summary, the use of the handheld ASD's hyperspectral signals makes the ground truth of the mixed pixel and endmember spectra available. The use of the synthesized hyperspectral signals makes the true abundances available. As a result, the performance of abundance estimation using the LSE method can be quantitatively evaluated, and the proposed DWT-based feature extraction approaches for improving the abundance estimation can be quantitatively analyzed.

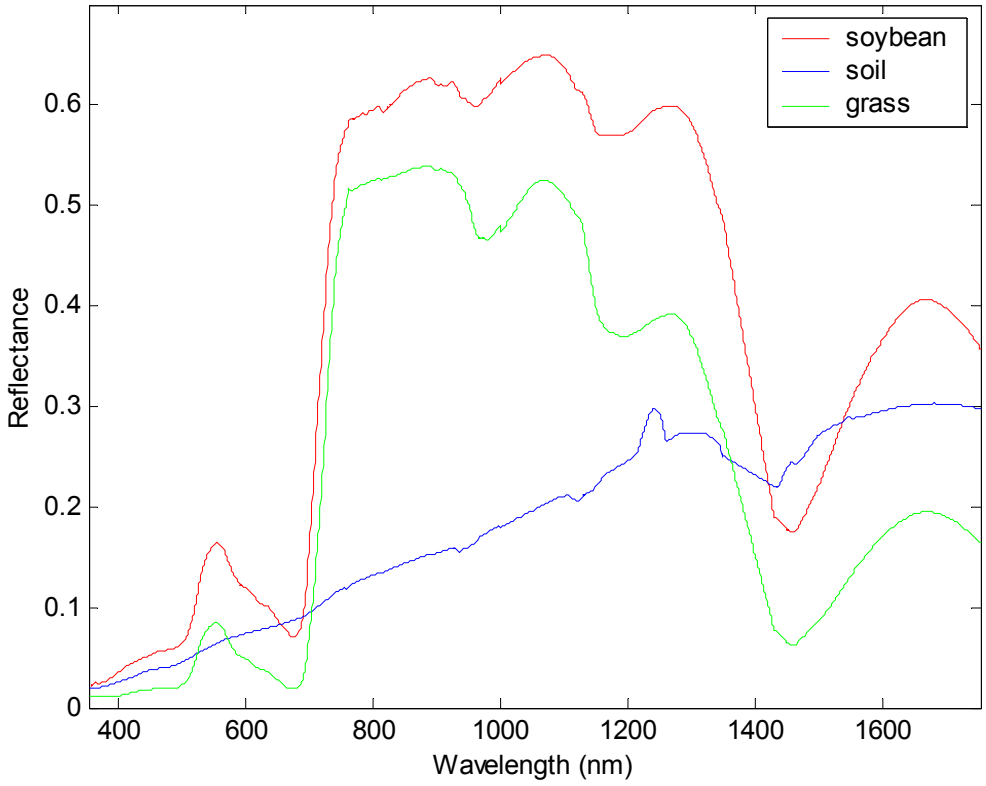

Figure 5.3. Endmember spectra of soybean, grass and soil utilized for the linear unmixing analysis in the dissertation.

5.2. Abundance Estimation Results for Experiment I: Soybean versus Soil

Experiment I investigates a two-endmember linear unmixing case. The two endmembers are soybean and soil. It is investigated as an example of a generic application where the ROI consists of the vegetation and the non-vegetation.

5.2.1. System Training Results and Feature Selection for Experiment I

Based on the proposed DWT-based linear unmixing system, the DWT using the Haar mother wavelet is first applied to the two endmember spectra and the 110 synthesized mixed pixel spectra using the training data set. The DWT is performed until the 9^{th} decomposition scale. Thus 9 sets of wavelet detail coefficient features and 9 sets of wavelet approximation coefficient features are extracted according to equations (3.25) and (3.26). For each of the 18 sets of DWT-based features, both the ULSE and the CLSE of abundances are completed. The RMSE of the abundance estimation is computed and utilized as the criterion of selecting the optimal DWT-based feature set. Table 5.2 summarizes the RMSE results based on the training data set, where "ORG" represents the use of the original hyperspectral signal without any pre-processing, and "All" represents the use of all PCA or DCT coefficients as features for abundance estimation.

Results in Table 5.2 show that when using the original hyperspectral signals, conventional PCA- and DCT-based features for the abundance estimation, the ULSE method produces a smaller RMSE, *i.e.*, a better abundance estimation, than the CLSE method. However, as discussed in Chapter IV, in order to make the abundance estimation results physically meaningful, in practical abundance estimation applications two constraints defined in equation (4.6) and (4.7) need to be taken into account. Therefore,

for this study the focus is on the CLSE of abundances, and the optimal DWT-based feature set is determined according to the CLSE results.

Using the CLSE method for the abundance estimation, the wavelet detail features at the 7th decomposition scale, "D7", produces the smallest RMSE value, 0.0924, among the 18 sets of DWT-based features. It is interesting to note that this RMSE value is even smaller than all RMSE values when using the ULSE method. That is, it is possible that the CLSE method can produce a better abundance estimation than the ULSE method when using the DWT-based features for the abundance estimation.

Table 5.2. RMSE of abundances estimation based on training data set for feature selection in Experiment I (soybean vs. soil).

DWT			PCA			DCT		
Coeffs	ULSE	CLSE	# Coeffs	ULSE	CLSE	# Coeffs	ULSE	CLSE
D1	0.1167	0.1154	1	0.9008	0.1333	1	0.8507	0.1645
D2	0.1094	0.1047	2	0.1069	0.1304	2	0.1169	0.1610
D3	0.1177	0.1053	3	0.1075	0.1303	3	0.1038	0.1338
D4	0.1111	0.1058	4	0.1076	0.1303	4	0.1068	0.1335
D5	0.1246	0.1061	5	0.1076	0.1303	5	0.1063	0.1324
D6	0.1256	0.1141	6	0.1075	0.1303	6	0.1068	0.1324
D7	0.1046	0.0924	7	0.1075	0.1303	7	0.1073	0.1314
D8	0.1615	0.1125	8	0.1075	0.1303	8	0.1073	0.1314
D9	0.1300	0.1087	9	0.1075	0.1303	9	0.1075	0.1315
A1	0.1074	0.1303	10	0.1075	0.1303	10	0.1076	0.1314
A2	0.1074	0.1303	11	0.1075	0.1303	11	0.1071	0.1306
A3	0.1074	0.1303	12	0.1075	0.1303	12	0.1070	0.1305
A4	0.1073	0.1304	13	0.1075	0.1303	13	0.1070	0.1305
A5	0.1073	0.1305	14	0.1075	0.1303	14	0.1070	0.1305
A6	0.1072	0.1308	15	0.1075	0.1303	15	0.1073	0.1305
A7	0.1076	0.1314	16	0.1075	0.1303	16	0.1074	0.1305
A8	0.1037	0.1364	17	0.1075	0.1303	17	0.1074	0.1305
A9	0.0998	0.1405	18	0.1074	0.1303	18	0.1075	0.1304
ORG	0.1074	0.1303	All	0.1074	0.1303	All	0.1074	0.1303

Since the DWT-based feature set of "D7" produced the smallest RMSE value, it is determined as an optimal set of DWT-based features for the abundance estimation, which will also be used in the system testing phase. As discussed in Chapter IV, two optimal parameters are recorded for the optimal set of DWT-based features. One optimal parameter is the optimal decomposition scale, which is the "7" for this case. The other optimal parameter is the detail or the approximation features, which is the detail ("D") for this case. During the system testing phase, the two recorded optimal parameters can be used to directly obtain the optimal DWT-based feature set.

Also note that using the CLSE method the wavelet detail features generally result in a smaller RMSE of abundance estimation than the wavelet approximation features. This observation indicates that the multiresolution analysis of wavelet transform extracts the detailed information from the endmember spectra, using which it is possible to better distinguish two endmembers, soybean and soil. As a result, the abundance estimation error is reduced. In contrast, the approximation information does not help to improve the abundance estimation at all, since using the wavelet approximation features brings the same or worse RMSE values, compared with using the original signals.

Training results in Table 5.2 also show that using the conventional PCA-based features and DCT-based features do not provide any help for improving the abundance estimation, since they all produce the same or worse RMSE, compared with using the original signals. Note that, however, for the PCA method, using the first three principal components (or coefficients) as feature leads to the same abundance estimation performance as using the original signals or all PCA coefficients. Compared with using

original signals, the PCA-based feature with 3 elements greatly reduces the dimensionality (or volume) of data, and thus could greatly reduce the computational cost of abundance estimation. This is a merit of using PCA-based pre-processing, even though it does not help improve the performance of abundance estimation. The same conclusion is valid for the DCT-based pre-processing, where using the first 18 DCT coefficients as the feature set leads to nearly the same RMSE of abundance estimation as using the original signals or all DCT coefficients. The PCA-based feature with the first 3 coefficients and the DCT-based feature with the first 18 coefficients are regarded as the "optimal" features for PCA and DCT pre-processing, respectively. Note that this "optimum" is in the sense of reducing the computational cost, rather than improving the abundance estimation performance. However, these "optimal" PCA- and DCT-based features will also be used in the system testing phase, for the purpose of comparing with the optimal DWT-based features.

5.2.2. System Testing Results and Performance Evaluation for Experiment I

The optimal features extracted using the three types of pre-processing methods, DWT, PCA and DCT, are further investigated in the system testing phase. The CLSE results of abundances from using the three types of features and the original hyperpsectral signals are quantitatively evaluated using the proposed three metrics, the RMSE of abundance estimation, the confidence of abundance estimation, and the abundance distribution diagram. For a comparison, the ULSE results of abundances are obtained and evaluated as well.

Figure 5.4 shows the RMSE results based on the testing data set. When using the original hyperspectral signals, the PCA-based features or the DCT-based features, the RMSE of abundance estimation is ~0.1, which indicates the average deviation of abundance estimation is ~0.1. That is, for example, if the truth of abundance is 0.4, then on average the estimated abundance could be any value between ~0.4±0.1. When using the DWT-based feature, the RMSE of abundance estimation, or the average deviation of abundance estimation, is ~0.05. That is, for the same example, the estimated abundance could lie between ~0.4±0.05 on average. This improvement indicates that on average the deviation of abundance estimate is reduced by ~50%, which is a promising improvement in the context of abundance estimation.

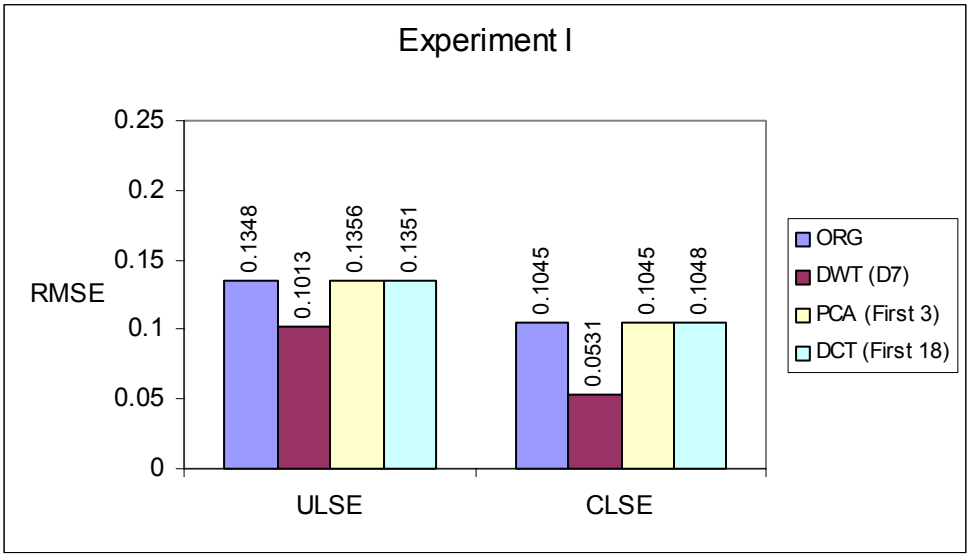

Figure 5.4. RMSE of abundance estimation based on testing data set in Experiment I (soybean vs. soil).

Figure 5.5 shows confidence curves of the CLSE of abundances, where the error interval ranges from 0 to 0.3. When the error interval is greater than 0.3, all confidence values are 1, thus they are no longer shown in the figure for the purpose of conciseness. Note that confidence curves for using the original hyperspectral signal, the PCA-based features and the DCT-based features almost overlap one another. This indicates that compared with using the original hyperspectral signals, using the PCA- and DCT-based features does not provide any improvement of abundance estimation other than reducing the computational cost. However, using the DWT-based feature set does provide a higher confidence than using the original hyperspectral signal, when given error intervals range from 0.05 to 0.2. For example, when the given error interval is 0.1, using the DWT-based feature results in a ~95% confidence, however using the other two types of features or the original hyperspectral signals only results in a ~60% confidence. That is, if the true abundance of an endmember is 0.4, then the confidence that the estimated abundance could lie between ~0.4±0.1 is ~95% when using the DWT-based feature, and ~60% when using the original hyperspectral signal, PCA- or DCT-based features. To reach ~95% confidence, the error interval has to be increased to 0.2 when using the original hyperspectral signal, PCA- or DCT-based features.

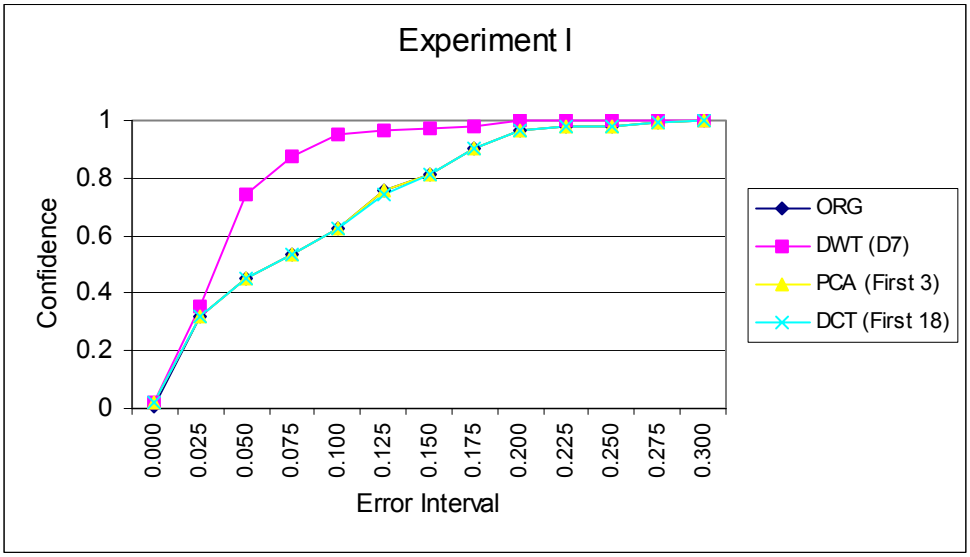

Figure 5.5. Confidence of abundance estimation based on testing data set in Experiment I (soybean vs. soil).

Figures 5.6(a) and (b) show the abundance distribution diagrams of soybean and soil, respectively. Note that the diagrams reflect the relationship of abundances for two endmembers defined in equation (5.2). That is, the sum of abundances for two endmembers is 1. In Figure 5.6, the true abundances and the abundance estimation results from using the original hyperspectral signal and the DWT-based features are shown. Clearly, the use of the DWT-based features leads to a better abundance estimation than the use of the original hyperspectral signals, because the abundance distribution corresponding to the DWT-based features is much closer to the true abundance, as compared with abundance distribution corresponding to the original hyperspectral signals.

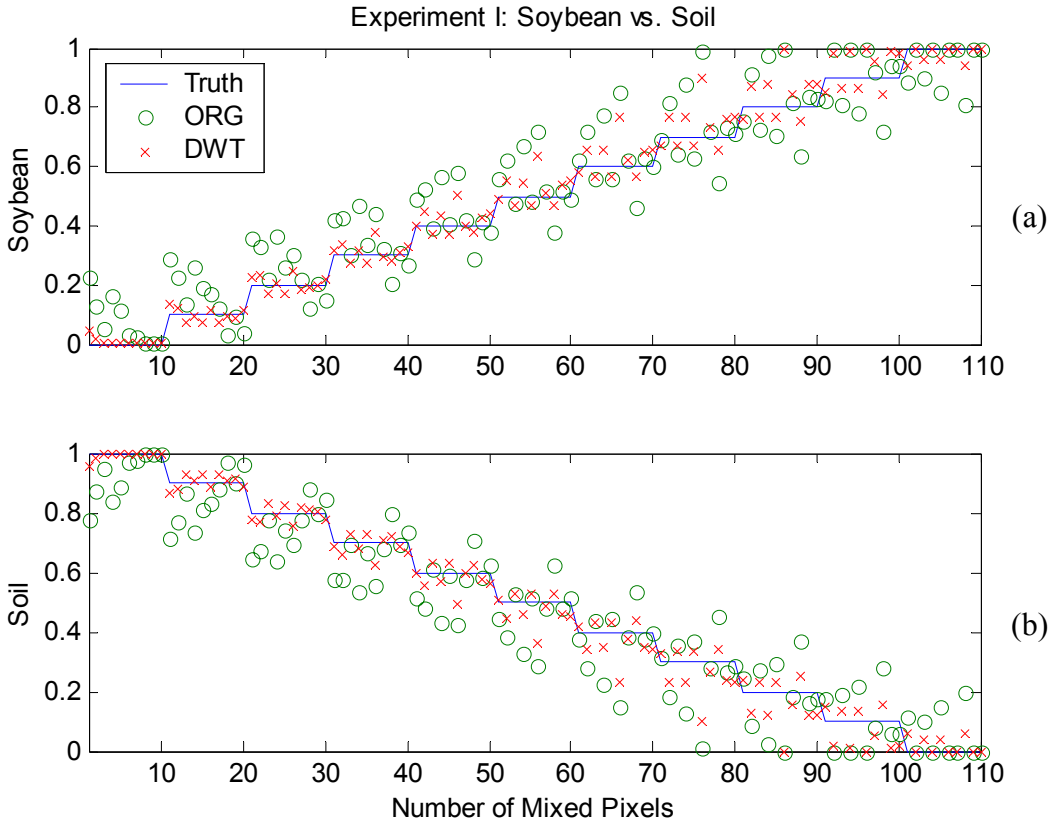

Figure 5.6. Abundance distribution diagrams based on testing data set in Experiment I (soybean vs. soil): (a) for soybean; and (b) for soil.

5.3. Abundance Estimation Results for Experiment II: Soybean versus Grass

Experiment II investigates another two-endmember linear unmixing case. The two endmembers are soybean and grass. It is investigated as an example of a generic application where the ROI consists of two similar vegetations. This is a more difficult unmixing case than Experiment I of soybean versus soil, since the two vegetations investigated have much similar hyperpsectral reflectance spectra. However, phenomena appearing in Experiment I can be similarly observed in this experiment.

5.3.1. System Training Results and Feature Selection for Experiment II

Similar to Experiment I, the system training is first accomplished to determine the optimal feature set. The DWT using the Haar mother wavelet and the feature extraction are performed on the two endmember spectra and the 110 synthesized mixed pixel spectra. Both the ULSE and the CLSE of abundances are implemented, and the RMSE of the abundance estimation is calculated. The RMSE results from using the CLSE are utilized to determine the optimal DWT-based feature set. Table 5.3 summarizes the RMSE results based on the training data set.

Table 5.3. RMSE of abundances estimation based on the training data set for feature selection in Experiment II (soybean vs. grass).

DWT			PCA			DCT		
Coeffs	ULSE	CLSE	# Coeffs	ULSE	CLSE	# Coeffs	ULSE	CLSE
D1	0.1885	0.1587	1	0.5820	0.3327	1	0.5916	0.3158
D2	0.1856	0.1608	2	0.1309	0.2960	2	0.1341	0.2953
D3	0.1909	0.1739	3	0.1302	0.2960	3	0.1292	0.3018
D4	0.1875	0.1832	4	0.1300	0.2959	4	0.1275	0.3009
D5	0.1723	0.1530	5	0.1298	0.2959	5	0.1274	0.3012
D6	0.1707	0.2820	6	0.1299	0.2959	6	0.1299	0.2969
D7	0.1416	0.2216	7	0.1297	0.2959	7	0.1302	0.2978
D8	0.1927	0.1519	8	0.1297	0.2959	8	0.1299	0.2976
D9	0.1478	0.1153	9	0.1297	0.2959	9	0.1299	0.2978
A1	0.1297	0.2959	10	0.1297	0.2959	10	0.1300	0.2963
A2	0.1297	0.2959	11	0.1297	0.2959	11	0.1296	0.2969
A3	0.1297	0.2959	12	0.1297	0.2959	12	0.1297	0.2968
A4	0.1297	0.2952	13	0.1297	0.2959	13	0.1291	0.2967
A5	0.1296	0.2954	14	0.1297	0.2959	14	0.1291	0.2967
A6	0.1297	0.2955	15	0.1297	0.2959	15	0.1292	0.2969
A7	0.1300	0.2960	16	0.1297	0.2959	16	0.1289	0.2968
A8	0.1284	0.2849	17	0.1297	0.2959	17	0.1290	0.2968
A9	0.1343	0.2917	18	0.1297	0.2959	18	0.1291	0.2964
ORG	0.1297	0.2959	All	0.1297	0.2959	All	0.1297	0.2959

Using the CLSE method for the abundance estimation, the wavelet approximation features, the PCA-based features and the DCT-based features do not provide much help on improvement of abundance estimation, but do provide a potential to reduce the computational cost. For example, in this experiment a PCA-based feature set consisting of the first 4 principal components (or coefficients) produces the smallest RMSE among all PCA-based features, and thus can be regarded as an "optimal" PCA-based feature set. A DCT-based feature set consisting of the first 2 coefficients produces the smallest RMSE among all DCT-based features, and thus can be regarded as an "optimal" DCT-based feature set. Also note that the wavelet detail features do provide an improvement in abundance estimation. Among all wavelet detail feature sets, the one at 9^{th} decomposition scale, "D9", produces the smallest RMSE value, 0.1153. Thus, the feature set of "D9" is determined as the optimal DWT-based features, and the two optimal parameters, "D" and "9", are recorded for the direct obtaining of DWT-based features during the system testing phase.

5.3.2. System Testing Results and Performance Evaluation for Experiment II

The optimal features extracted using the three types of pre-processing methods, DWT, PCA and DCT, are further investigated in the system testing phase. The CLSE results of abundances from using the three types of features and the original hyperpsectral signals are quantitatively evaluated using the proposed three metrics, the RMSE of abundance estimation, the confidence of abundance estimation, and the abundance distribution diagram. For a comparison, the ULSE results of abundances are obtained and evaluated as well.

Figure 5.7 shows the RMSE results based on the testing data set. When using the original hyperspectral signals, the PCA-based or the DCT-based features, the RMSE of abundance estimation (or the average deviation of abundance estimation) is ~0.2. However, when using the DWT-based features, the RMSE of abundance estimation is reduced to ~0.14. That is, on average the deviation of abundance estimate is reduced by ~30%. This is a promising improvement, considering that the Experiment II is a much more difficult two-endmember abundance estimation scenario than Experiment I.

Figure 5.8 shows confidence curves of the CLSE of abundances, where the error interval ranges from 0 to 0.65. When the error interval is greater than 0.65, all confidence values are 1, thus they are no longer shown in the figure for the purpose of conciseness. Again, note that confidence curves for using the original hyperspectral signals, the PCA-based features and the DCT-based features almost overlap one another. This indicates that using the PCA- and DCT-based features does not provide any improvement of abundance estimation other than reducing the computational cost, compared with using the original hyperspectral signals. However, using the DWT-based feature set does provide a higher confidence than using the original hyperspectral signal, the PCA- or DCT-based features, when given error intervals range form 0.1 to 0.4. For example, when the given error interval is 0.275, using the DWT-based feature results in a ~95% confidence, however using the other two types of features or the original signals only results in a ~80% confidence. To reach ~95% confidence, the error interval has to be increased to 0.425 when using the original hyperspectral signal, PCA- or DCT-based features.

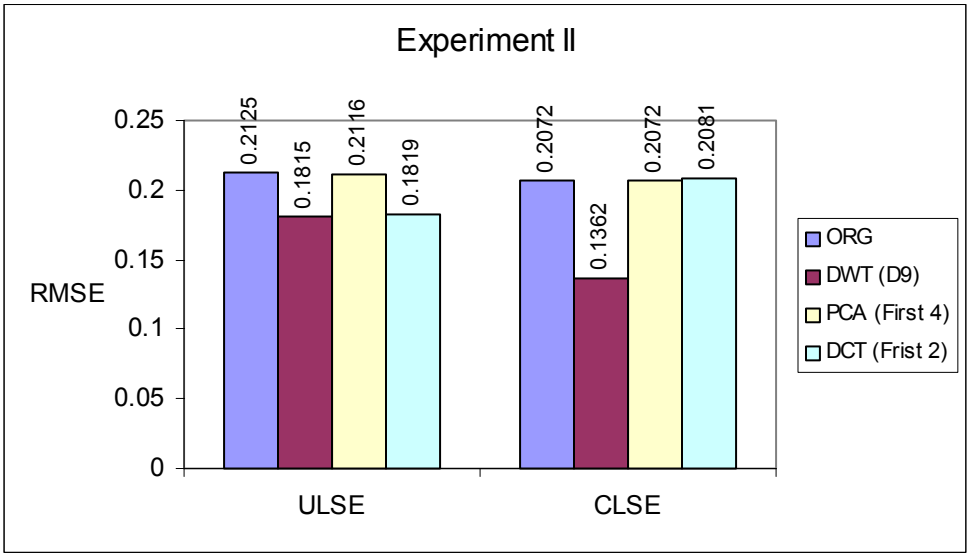

Figure 5.7. RMSE of abundance estimation based on testing data set in Experiment II (soybean vs. grass).

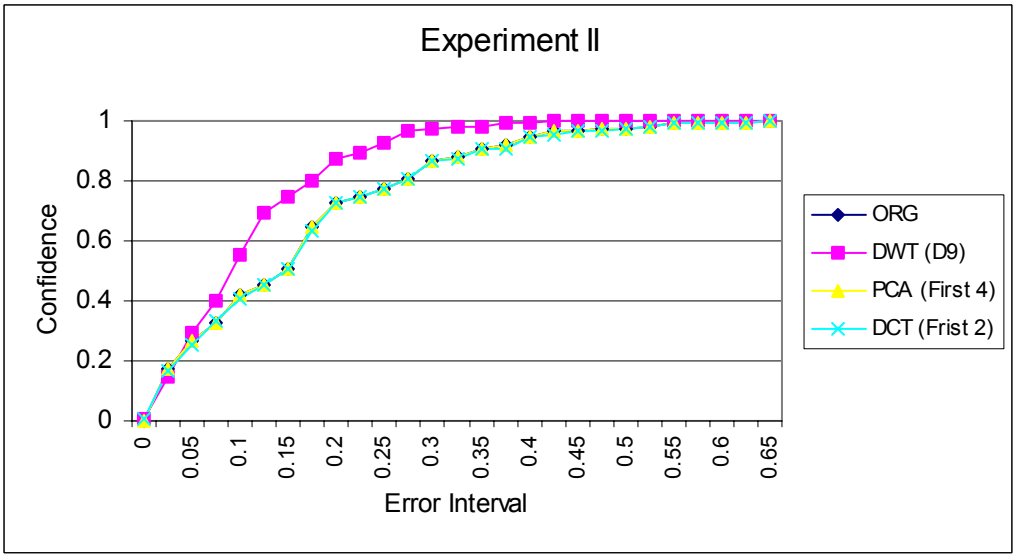

Figure 5.8. Confidence of abundance estimation based on testing data set in Experiment II (soybean vs. grass).

Figures 5.9(a) and (b) show the abundance distribution diagrams of soybean and grass, respectively. Note that the use of the DWT-based features leads to a better abundance estimation than the use of the original hyperspectral signals, because the abundance distribution corresponding to the DWT-based features is closer to the true abundance, as compared with abundance distribution corresponding to the original hyperspectral signals.

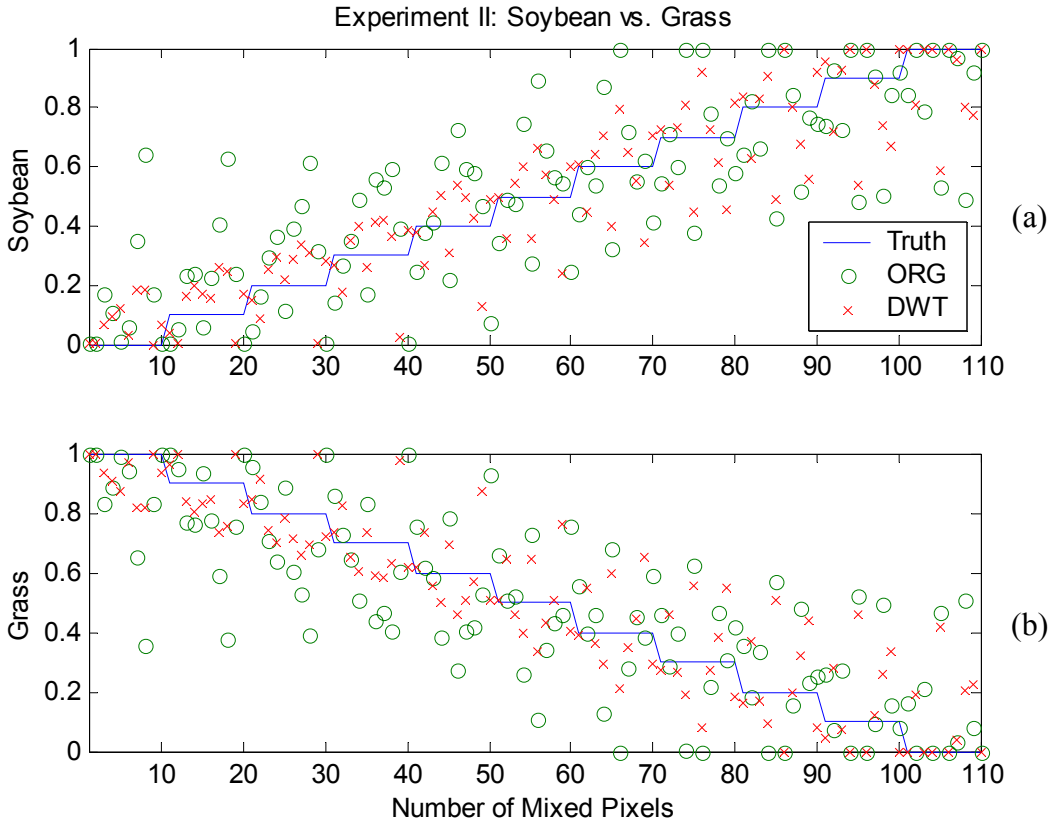

Figure 5.9. Abundance distribution diagrams based on testing data set in Experiment II (soybean vs. grass): (a) for soybean; and (b) for grass.

In general, worse abundance estimation results are obtained in Experiment II, as compared with the results of Experiment I. This is predictable because Experiment II is a more difficult two-endmember linear unmixing case than Experiment I. However, in both Experiment I and II, the use of the DWT-based features does improve the LSE of abundances of two endmembers.

5.4. Abundance Estimation Results for Experiment III: Soybean, Grass versus Soil

Experiment III is a three-endmember linear unmixing case, which is designed and implemented as an example of the mulit-endmember linear unmixing case. The three endmembers are soybean, grass, and soil. It is investigated as an example of a generic application where the ROI consists of two different types of vegetations and one type of non-vegetation.

5.4.1. System Training Results and Feature Selection for Experiment III

Similar to Experiment I and II, the system training is first accomplished to determine the optimal feature set. The DWT using the Haar mother wavelet and the feature extraction are performed on the three endmember spectra and the 660 synthesized mixed pixel spectra. Both the ULSE and the CLSE of abundances are implemented, and the RMSE of the abundance estimation is calculated. The RMSE results from using the CLSE are utilized to determine the optimal DWT-based feature set. Table 5.4 summarizes the RMSE results based on the training data set.

Table 5.4. RMSE of abundances estimation based on training data set for feature selection in Experiment III (soybean, grass and soil).

DWT			PCA			DCT		
Coeffs	ULSE	CLSE	# Coeffs	ULSE	CLSE	# Coeffs	ULSE	CLSE
D1	0.1384	0.1161	1	Inf	0.3140	1	0.7126	0.3328
D2	0.1246	0.1150	2	0.4824	0.1498	2	0.8945	0.1555
D3	0.1320	0.1188	3	0.1620	0.1491	3	0.3694	0.1524
D4	0.1274	0.1191	4	0.1444	0.1491	4	0.1765	0.1517
D5	0.1278	0.1120	5	0.1425	0.1491	5	0.1743	0.1517
D6	0.1223	0.1254	6	0.1306	0.1491	6	0.1686	0.1501
D7	0.1177	0.1119	7	0.1267	0.1491	7	0.1621	0.1502
D8	0.3072	0.1323	8	0.1266	0.1491	8	0.1509	0.1501
D9	0.1842	0.1200	9	0.1260	0.1491	9	0.1494	0.1501
A1	0.1250	0.1491	10	0.1260	0.1491	10	0.1429	0.1496
A2	0.1250	0.1491	11	0.1259	0.1491	11	0.1435	0.1495
A3	0.1250	0.1491	12	0.1259	0.1491	12	0.1437	0.1495
A4	0.1252	0.1490	13	0.1258	0.1491	13	0.1362	0.1495
A5	0.1261	0.1491	14	0.1258	0.1491	14	0.1347	0.1495
A6	0.1296	0.1494	15	0.1258	0.1491	15	0.1287	0.1495
A7	0.1408	0.1497	16	0.1258	0.1491	16	0.1270	0.1494
A8	0.1377	0.1488	17	0.1254	0.1491	17	0.1272	0.1494
A9	0.8096	0.1522	18	0.1254	0.1491	18	0.1256	0.1493
ORG	0.1250	0.1491	All	0.1250	0.1491	All	0.1250	0.1491

Using the CLSE method for the abundance estimation, the wavelet approximation features, the PCA-based features and the DCT-based features do not provide much help on improvement of abundance estimation, but do provide a potential to reduce the computational cost. For example, in this experiment a PCA-based feature set consisting of the first 3 principal components (or coefficients) produces the smallest RMSE among all PCA-based features, and thus can be regarded as an “optimal” PCA-based feature set. A DCT-based feature set consisting of the first 18 coefficients produces the smallest RMSE among all DCT-based features, and thus can be regarded as an “optimal” DCT-based feature set. Also note that the wavelet detail features do provide an improvement in

abundance estimation. Among all wavelet detail feature sets, the one at 7^{th} decomposition scale, "D7", produces the smallest RMSE value, 0.1119. Thus, the feature set of "D7" is determined as the optimal DWT-based features, and the two optimal parameters, "D" and "7", are recorded for the direct obtaining of DWT-based features during the system testing phase.

It is interesting to note that using the ULSE method for the abundance estimation, the RMSE value is infinite when using only the first coefficient of PCA as feature sets, and the RMSE values are unusually large when using the first one or two coefficients of PCA or DCT as feature sets. This phenomenon can also be observed in the results of Experiments I and II, as shown in Tables 5.2 and 5.3. Based on the theoretical analysis in Chapter II, this phenomenon stems from the fact that the correlation matrix of endmember spectra is singular or close to being singular, corresponding to the infinite RMSE value or the unusually large RMSE values respectively. The singular, or close-to-singular, correlation matrix of endmember spectra implies an unstable abundance estimation according to the discussion in Section 2.4. Thus, it is not a good choice to use only the first one or two coefficients of PCA or DCT as features for the abundance estimation. The further discussion about the stability of the least squares solution of abundance estimation is provided in Section 5.6.

5.4.2. System Testing Results and Performance Evaluation for Experiment III

The optimal features extracted using the three types of pre-processing methods, DWT, PCA and DCT, are further investigated in the system testing phase. The CLSE results of abundances from using the three types of features and the original hyperpsectral

signals are quantitatively evaluated using the proposed three metrics, the RMSE of abundance estimation, the confidence of abundance estimation, and the abundance distribution diagram. For a comparison, the ULSE results of abundances are obtained and evaluated as well.

Figure 5.10 shows the RMSE results based on the testing data set. When using the original hyperspectral signals, the PCA-based or the DCT-based features, the RMSE of abundance estimation (or the average deviation of abundance estimation) is ~0.14. However, when using the DWT-based features, the RMSE of abundance estimation is reduced to ~0.1. That is, on average the deviation of abundance estimate is reduced by ~30%. This is a promising improvement, considering the copmplicated abundance estimation scenario where there are three endmembers and two of them are vegetations with much similar hyperpsectral reflectance spectra.

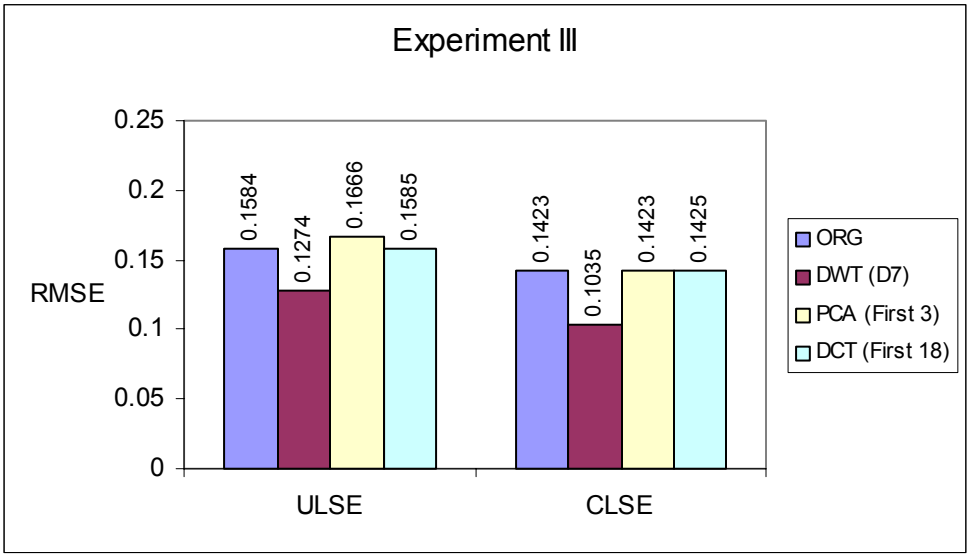

Figure 5.10. RMSE of abundance estimation based on testing data set in Experiment III (soybean, grass vs. soil).

Figure 5.11 shows confidence curves of the CLSE of abundances, where the error interval ranges from 0 to 0.45. When the error interval is greater than 0.45, all confidence values are 1, thus they are no longer shown in the figure for the purpose of conciseness. Note that confidence curves for using the original hyperspectral signals, the PCA-based features and the DCT-based features almost overlap one another. This indicates that using the PCA- and DCT-based features does not provide any improvement of abundance estimation other than reducing the computational cost, compared with using the original hyperspectral signals. However, using the DWT-based feature set does provide a higher confidence than using the original hyperspectral signal, the PCA- or DCT-based features, when given error intervals range form 0.05 to 0.2. For example, when the given error interval is 0.175, using the DWT-based feature results in a ~95% confidence, however using the other two types of features or the original signals only results in a ~80% confidence. To reach ~95% confidence, the error interval has to be increased to 0.225 when using the original hyperspectral signal, PCA- or DCT-based features.

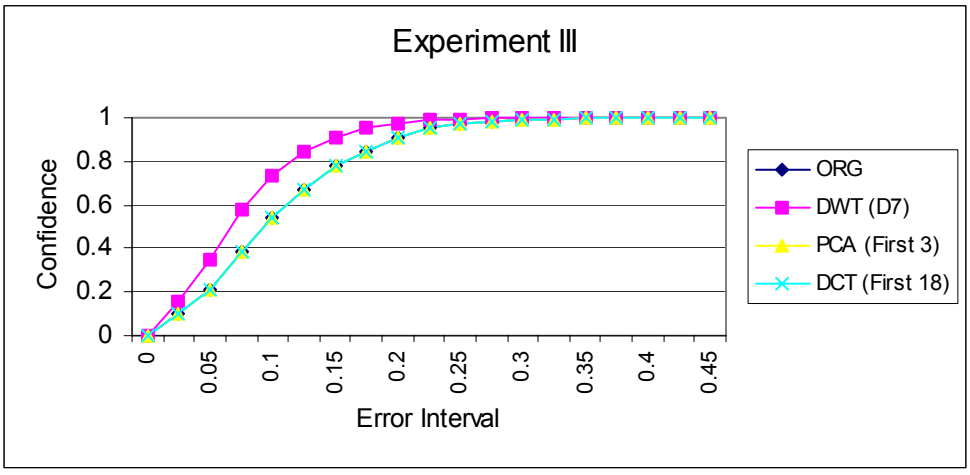

Figure 5.11. Confidence of abundance estimation based on testing data set in Experiment III (soybean, grass vs. soil).

Figures 5.12, 5.13 and 5.14 show the abundance distribution diagrams of soybean grass and soil, respectively. Recall that a total of 660 mixed pixels are tested in this experiment. For each of the three endmembers, abundance distribution corresponding to the 660 mixed pixels are shown in three diagrams, where the upper diagram shows the abundance distribution for the first 220 mixed pixels, the middle diagram shows the abundance distribution for the second 220 mixed pixels, and the lower diagram shows the abundance distribution for the final 220 mixed pixels. Note that the diagrams reflect the relationship of abundances for three endmembers defined in equation (5.4). That is, the sum of abundances for three endmembers is 1. In Figures 5.12, 5.13 and 5.14, the true abundances and the abundance estimation results from using the original hyperspectral signal and the DWT-based features are shown. Note that the use of the DWT-based features leads to a better abundance estimation than the use of the original hyperspectral signals, because the abundance distribution corresponding to the DWT-based features is closer to the true abundance, as compared with abundance distribution corresponding to the original hyperspectral signals.

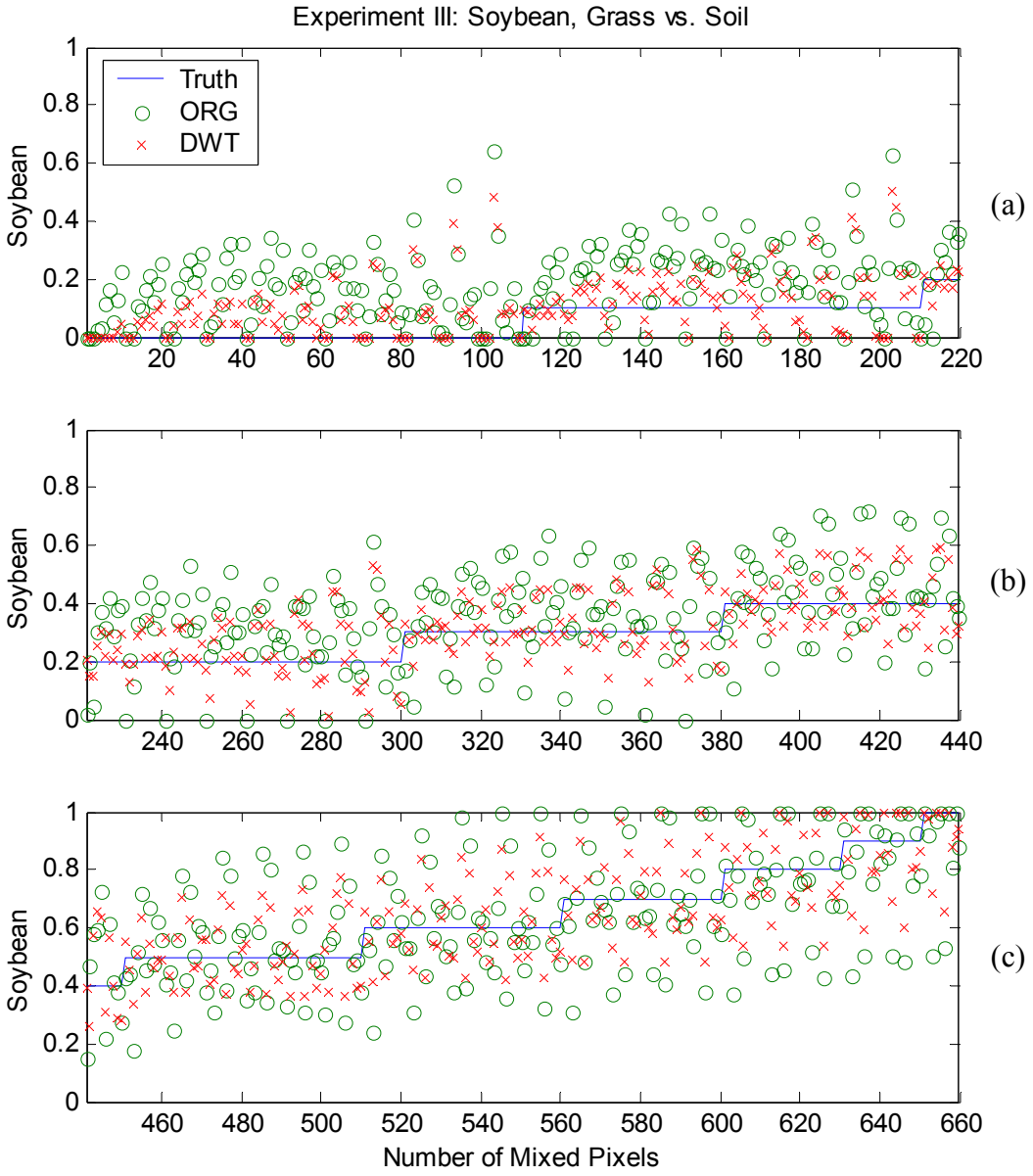

Figure 5.12. Abundance distribution diagrams for soybean based on testing data set in Experiment III (soybean, grass vs. soil): (a) the first 220 samples; (b) the second 220 samples; and (c) the final 220 samples.

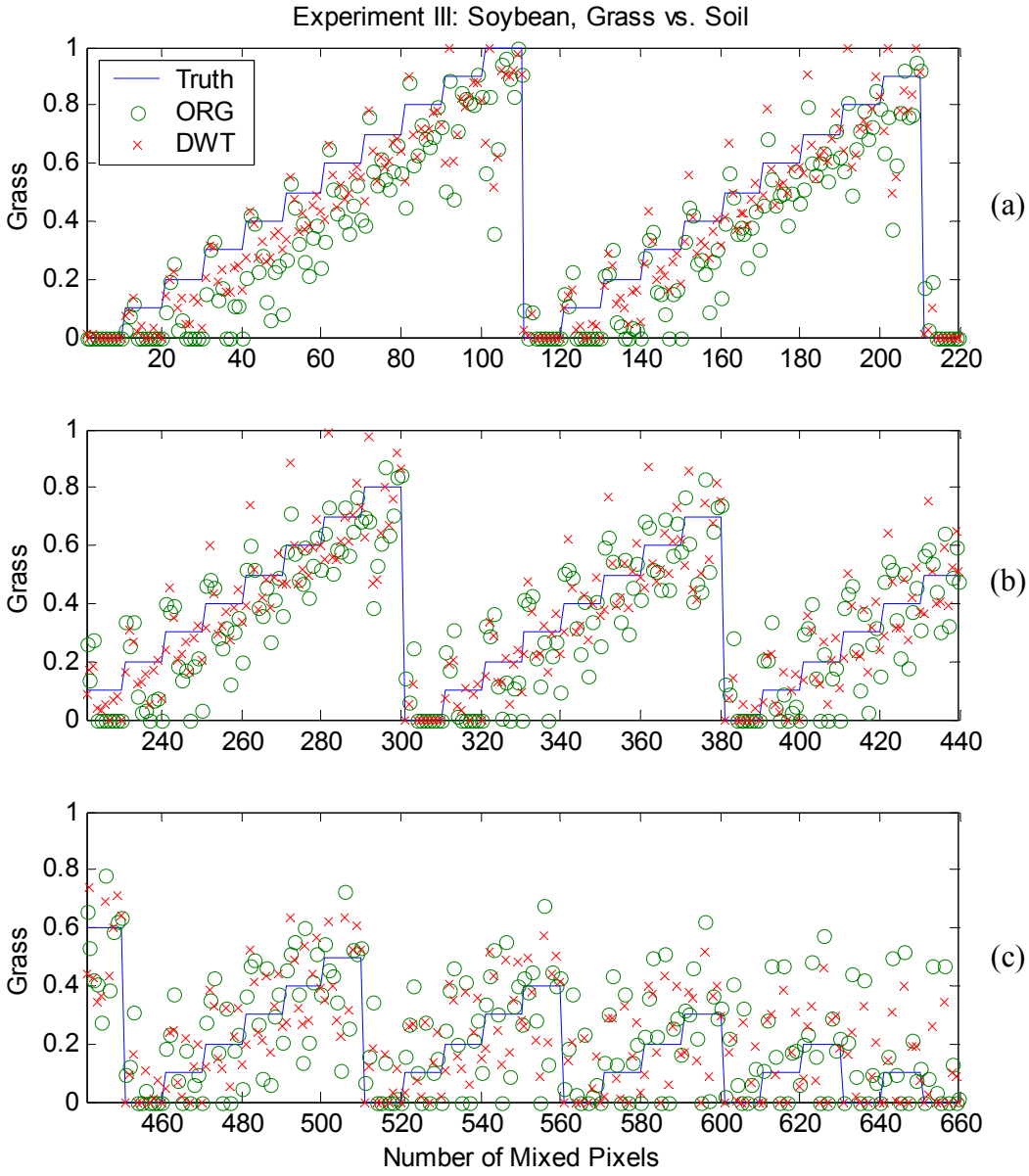

Figure 5.13. Abundance distribution diagrams for grass based on testing data set in Experiment III (soybean, grass vs. soil): (a) the first 220 samples; (b) the second 220 samples; and (c) the final 220 samples.

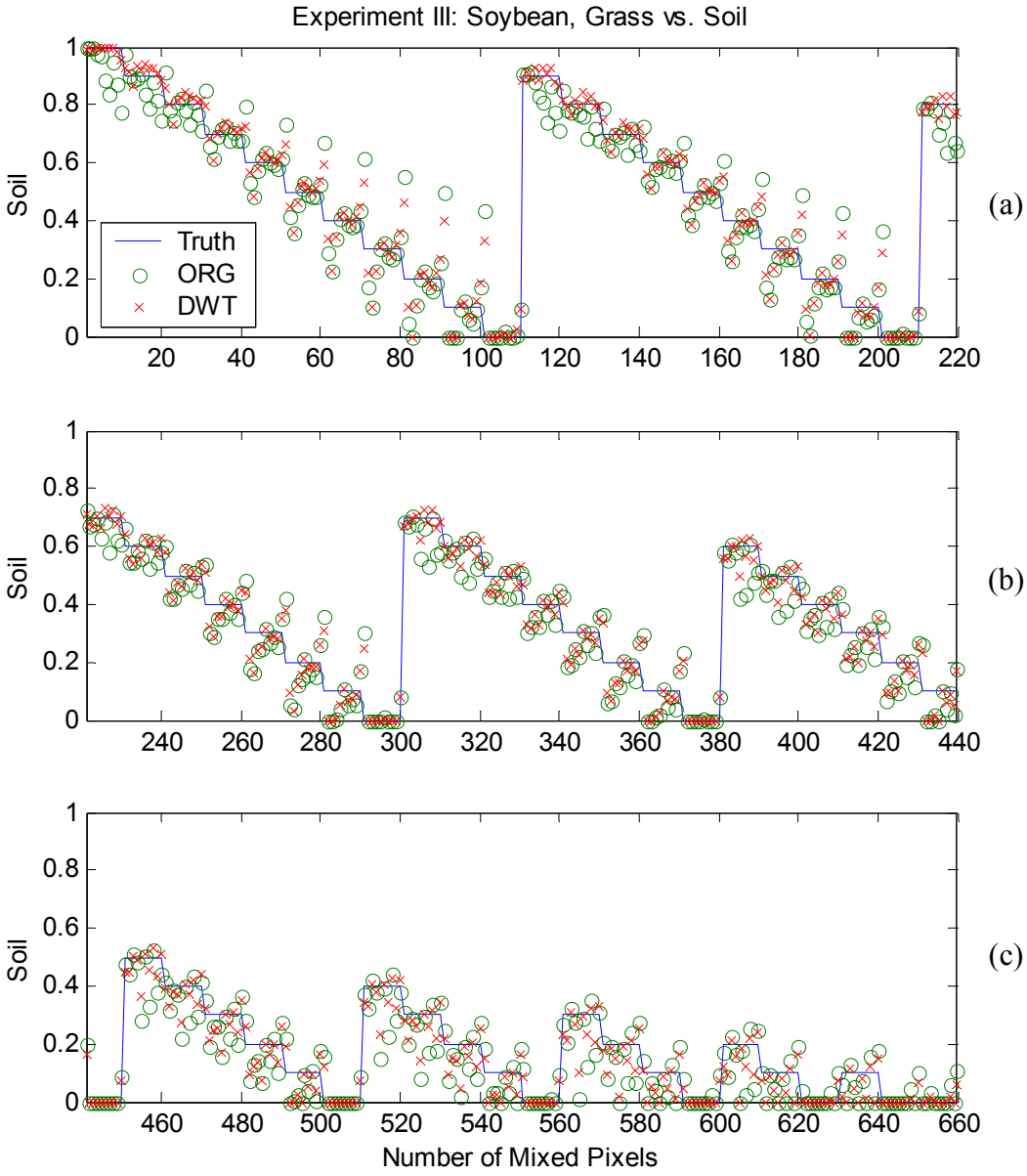

Figure 5.14. Abundance distribution diagrams for soil based on testing data set in Experiment III (soybean, grass vs. soil): (a) the first 220 samples; (b) the second 220 samples; and (c) the final 220 samples.

In summary, all three sets of experiments show that the use of the DWT-based features improves the abundance estimation using the LSE method. For the relatively simple abundance estimation scenario of Experiment I, the improvement is relatively high. For more complicated abundance estimation scenarios of Experiment II and III, the improvement is much less. In general, the average deviation of abundance estimation is reduced by 30-50%. The experiments also show that the use of the PCA- and DCT-based features do not provide any improvement of abundance estimation using the LSE method, but they do provide the potential for reducing the computational cost of abundance estimation by using the lower-dimensional features. Another advantage of using the DWT-based features is that it provides the potential for improving the stability of abundance estimation solutions, which will be discussed in Section 5.6.

5.5. Discussion about Linearity of Features

Theorem 3.1 argues that an orthonormal linear transform does not change the least square solution of abundance estimation. The DWT using the Haar mother wavelet, the PCA and the DCT are all orthonormal linear transforms. Thus, using all the transform coefficients from any of the three transforms should produce the same abundance estimates as using the original hyperspectral signals. This is proven true from the experiment results of the three sets of experiments in this study. The experiments are performed on the testing data sets of hyperspectral signals. For the simplicity, Figure 5.15 shows an abundance distribution diagram of soybean from Experiment I, as an example of all experiment results. Figure 5.15(a) shows the results of using all DWT coefficients and the original hyperspectral signals, Figure 5.15(b) shows the results of using all PCA

coefficients and the original hyperspectral signals, and Figure 5.15(c) shows the results of using all DCT coefficients and the original hyperspectral signals. Clearly, they all produce the same abundance estimations.

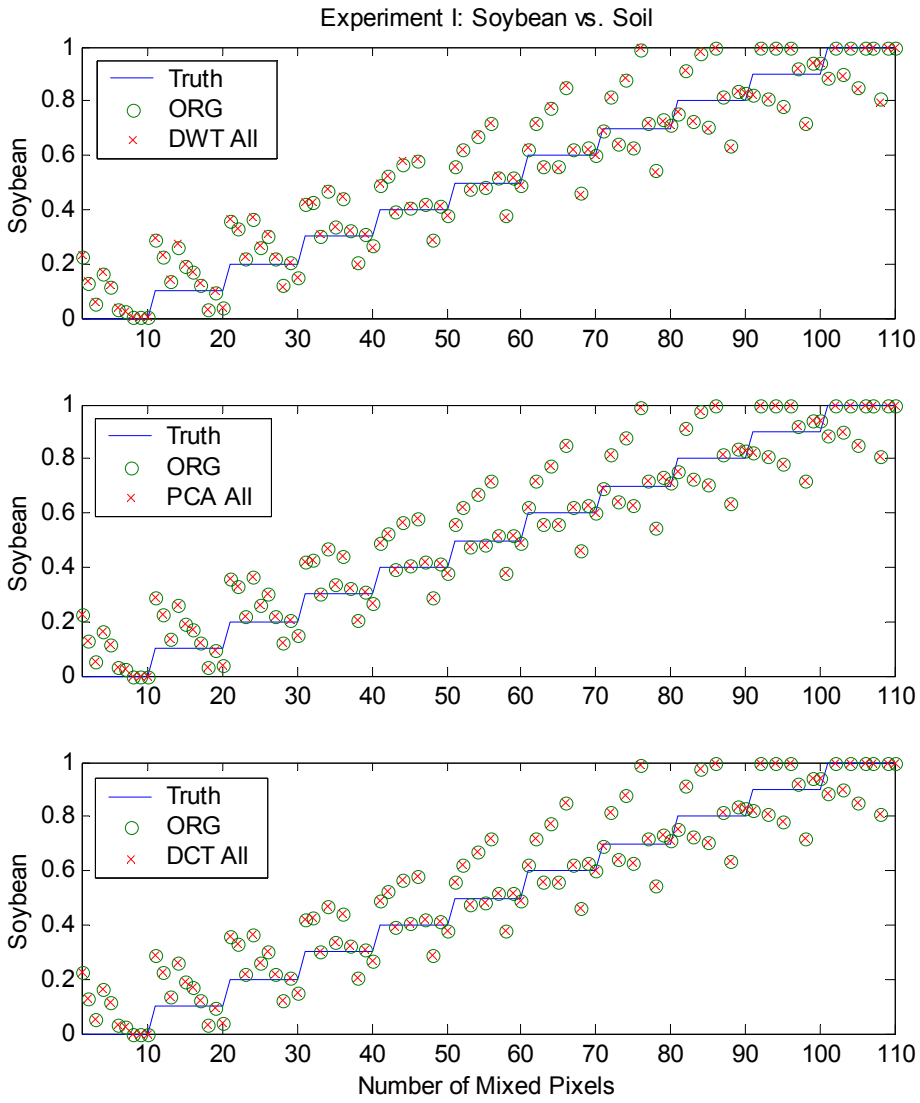

Figure 5.15. Abundance distribution diagrams for soybean based on testing data set in Experiment I (soybean vs. soil) when using all DWT, PCA, or DCT coefficients as features: (a) when using all DWT coefficients; (b) when using all PCA coefficients; (c) when using all DCT coefficients.

It is also argued in Chapter III that it is important to ensure the LMM defined in equation (1.1) remains after the wavelet analysis and feature extraction. The nonlinear wavelet energy feature, which has proven successful in our previous research for target detection and classification problem [5-6, 39-41], is no longer a good choice for the linear unmixing problem. To verify this argument, the three sets of experiments are implemented using the nonlinear DWT-based energy features. The abundance estimation results are compared with ones from using the linear DWT-based coefficient features and the original hyperspectral signals. Note that the linear DWT-based features utilized in this comparison are the DWT-based optimal feature set determined in the system training phase. The nonlinear DWT-based energy feature is calculated according to equation (3.23). All experiments are performed on the testing data sets of hyperspectral signals.

The resulting abundance estimations are evaluated using two metrics: the RMSE of abundance estimation and the confidence of abundance estimation. The RMSE results are shown in Figure 5.16, 5.18 and 5.20, for Experiment I, II and III, respectively. The confidence curves of the CLSE of abundances for the three sets of experiments are shown in Figure 5.17, 5.19 and 5.21, respectively. Clearly, it can be seen that using the nonlinear DWT-based energy features results in worse abundance estimation than using the linear DWT-based coefficient features, and even than using the original hyperspectral signals. Note that the use of the nonlinear DWT energy feature does reduce the dimensionality (or volume) of hyperspectral signals, but at the cost of losing abundance estimation accuracy. This is different from the case of using the PCA- or DCT-based features, where the dimensionality of hyperspecral signals is reduced without degrading the abundance

estimation performance, as compared to using the original hyperspectral signal. In short, it is a better choice to use the linear features when using the LMM for the abundance estimation.

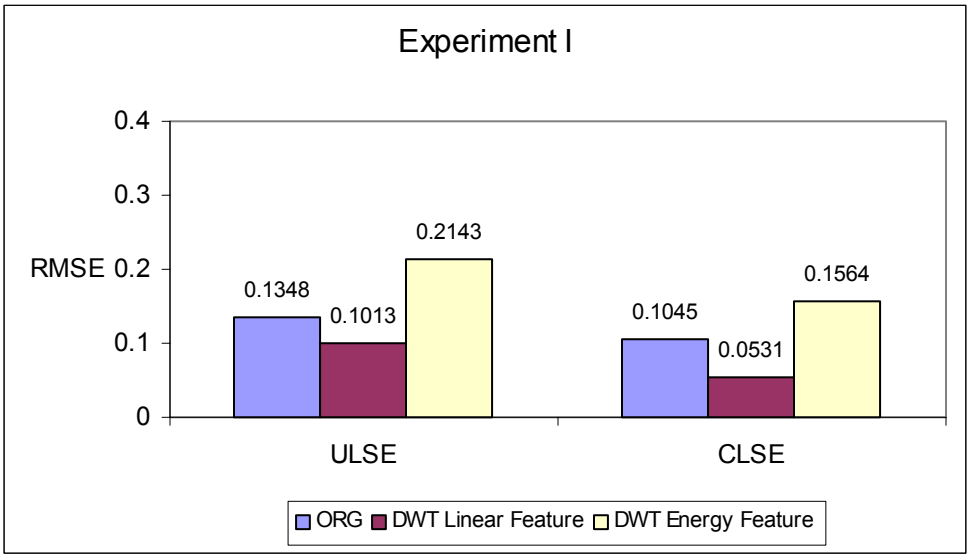

Figure 5.16. RMSE of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment I (soybean vs. soil).

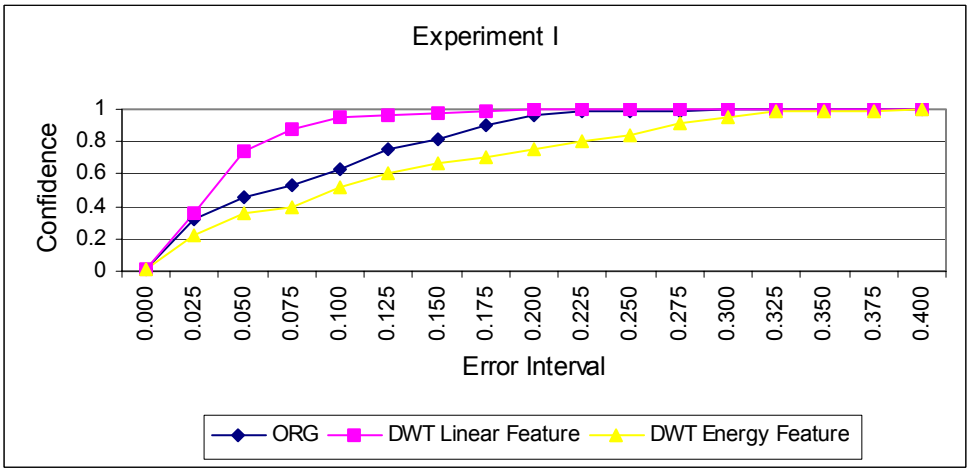

Figure 5.17. Confidence of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment I (soybean vs. soil).

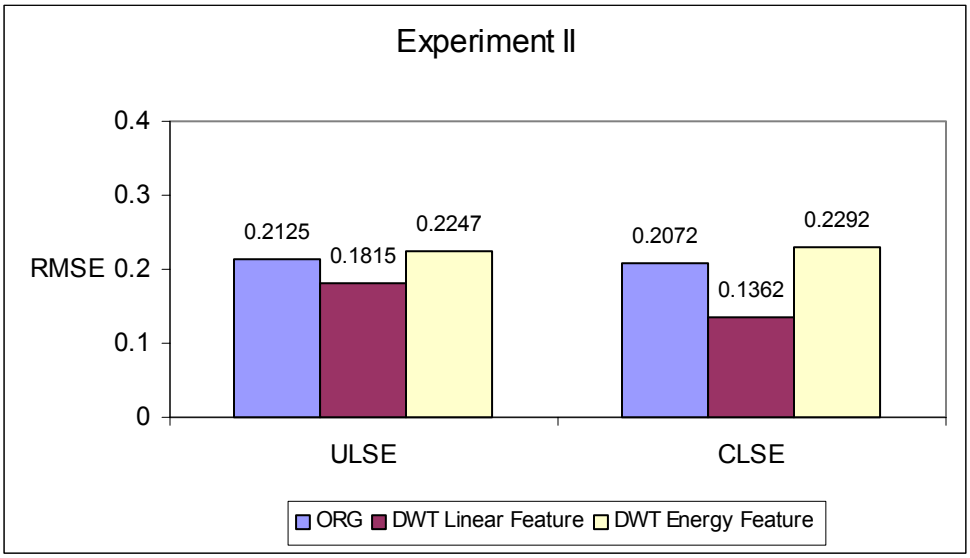

Figure 5.18. RMSE of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment II (soybean vs. grass).

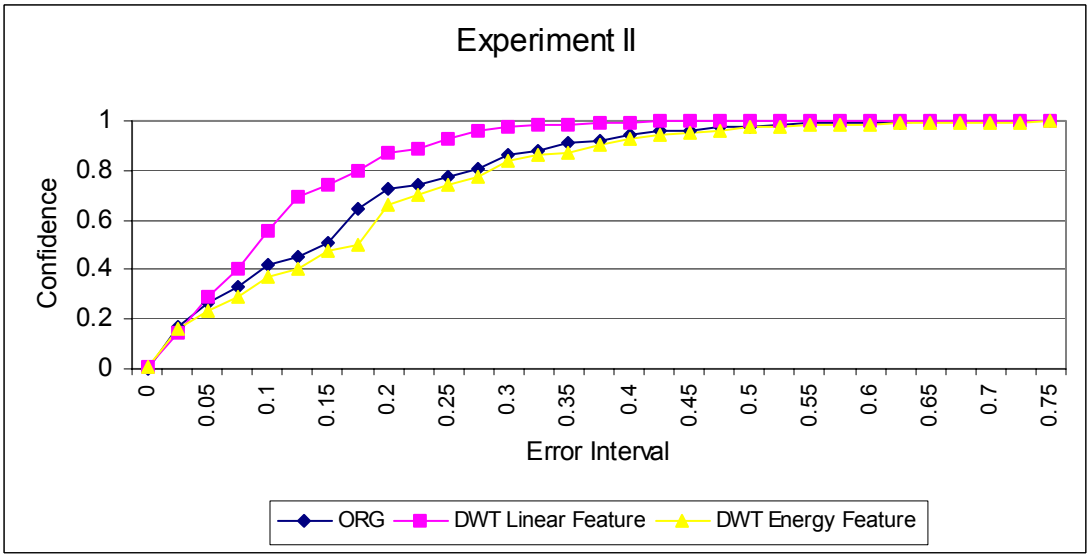

Figure 5.19. Confidence of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment II (soybean vs. grass).

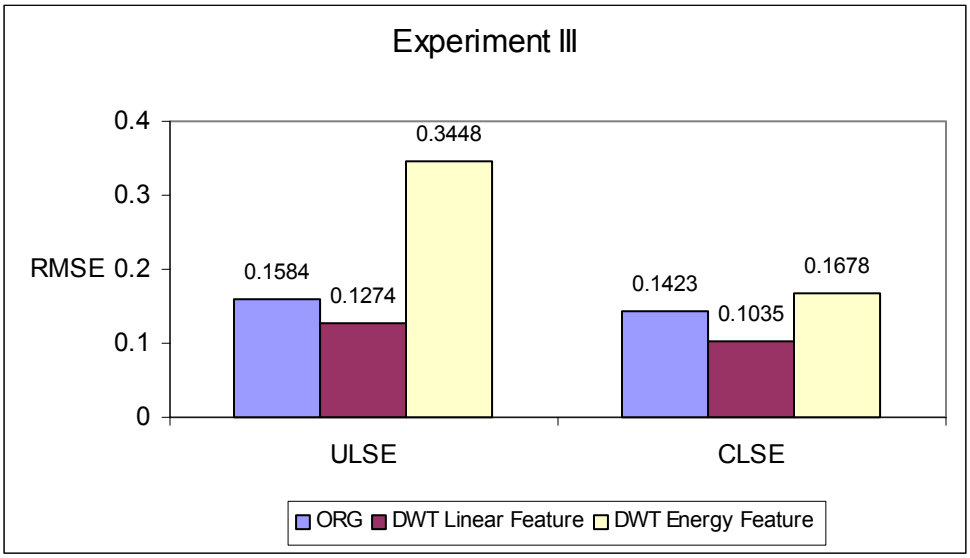

Figure 5.20. RMSE of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment III (soybean, grass vs. soil).

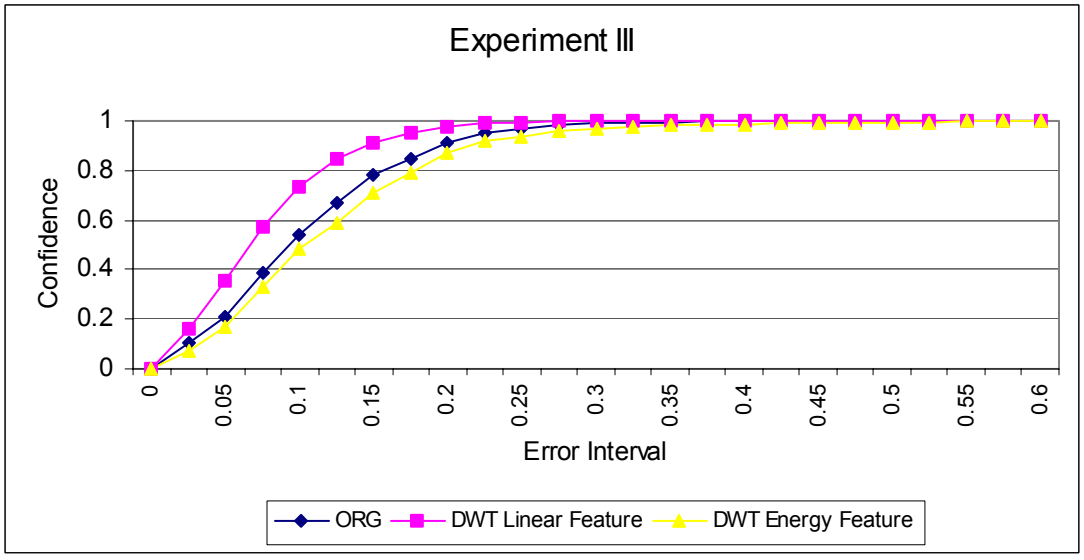

Figure 5.21. Confidence of abundance estimation based on testing data set for comparison of the linear coefficient feature and the nonlinear energy feature of DWT in Experiment III (soybean, grass vs. soil).

5.6. Discussion about Stability of Least Squares Solution

In Section 2.4 it is argued that the stability of abundance estimation is affected by the eigenvalues of the correlation matrix, R, of endmember spectra. To obtain a stable least squares solution of abundance estimation, an ill-conditioned R matrix, or an extremely small eigenvalue of R matrix, should be avoided. The condition number is utilized to measure the ill-condition of the correlation matrix R. The condition number, $CN(R)$, of the correlation matrix R is computed in equation (2.31), and it is always greater than or equal to 1. A larger condition number indicates a matrix closer to the ill-condition. It is also argued that the ill-conditioned R matrix, or the extremely small eigenvalue of R, could be avoided by using appropriate feature extraction approaches, particularly the feature extraction method based on the DWT.

To further support these arguments, the eigenvalues, λ_i, and the condition number, $CN(R)$, of the correlation matrix, R, of the endmember spectra are calculated for the three sets of experiments. The results are summarized in Tables 5.5, 5.6, and 5.7 for Experiments I, II and III, respectively. Note that the DWT-, PCA- and DCT-based features utilized in the experiments are all optimal features determined in the system training phase, as discussed in Sections 5.2-5.4. The calculations of λ_i and $CN(R)$ are all based on using these optimal features and the testing data sets of hyperspectral signals.

In general, for all the three experiments, using the DWT-based features avoids the extremely small eigenvlaues of the correlation matrix of the endmember spectra, and greatly reduces the condition number of the correlation matrix, as compared to using the original hyperspectral signal. For example, for Experiment I, the condition number,

$CN(R)$, is reduced from ~31 to ~13. For Experiment II, the condition number is reduced from ~112 to ~41. For Experiment III, the condition number is reduced from ~1518 to ~68. Among all the three experiments, Experiment III of the three-endmember case has the largest condition number, which is because there exists a very small eigenvalue, $\lambda_1 = 0.0011$, and the correlation matrix is close to an ill-condition matrix.

Table 5.5. Experimental values of parameters in the theoretical analysis of abundance estimation, based on testing data set, for Experiment I (soybean vs. soil).

λ_1	0.0384	0.0759	0.0383	0.0382
λ_2	1.1735	1.0098	1.1731	1.1741
$CN(R)$	30.5203	13.3026	30.6283	30.7062
$Tr[R^{-1}]$	26.8602	14.1644	26.9603	27.0051
σ_o^2	5.55E-06	7.65E-04	2.49E-03	4.31E-04
Γ_x	1.21E-02	5.67E-03	1.22E-02	1.21E-02
$(\Gamma_x)^{1/2}$	0.1099	0.0753	0.1106	0.1101
Γ_{xs}	7.46E-05	5.42E-03	3.36E-02	5.81E-03
$(\Gamma_{xs})^{1/2}$	0.0086	0.0736	0.1832	0.0763

Table 5.6. Experimental values of parameters in the theoretical analysis of abundance estimation, based on testing data set, for Experiment II (soybean vs. grass).

	ORG	DWT	PCA	DCT
λ_1	0.0135	0.0456	0.0135	0.0041
λ_2	1.5210	1.8641	1.5205	1.4356
$CN(R)$	112.2642	40.9038	112.6089	353.0108
$Tr[R^{-1}]$	74.4647	22.4798	74.7205	246.5956
σ_o^2	6.5701E-06	4.60E-03	2.27E-03	4.10E-03
Γ_x	0.0401	0.0296	0.0400	0.0315
$(\Gamma_x)^{1/2}$	0.2003	0.1721	0.2000	0.1775
Γ_{xs}	0.0002	0.0517	0.0847	0.5052
$(\Gamma_{xs})^{1/2}$	0.0156	0.2274	0.2910	0.7108

Table 5.7. Experimental values of parameters in the theoretical analysis of abundance estimation, based on testing data set, for Experiment III (soybean, grass vs. soil).

	ORG	DWT	PCA	DCT
λ_1	0.0011	0.0243	0.0008	0.0009
λ_2	0.0638	0.0971	0.0637	0.0635
λ_3	1.6817	1.6540	3.0000	3.0000
$CN(R)$	1.52E+03	6.80E+01	2.03E+03	1.79E+03
$Tr[R^{-1}]$	9.19E+02	5.20E+01	1.22E+03	1.08E+03
σ_o^2	3.87E-06	6.36E-04	1.75E-03	3.00E-04
Γ_x	0.0180	0.0113	0.0210	0.0180
$(\Gamma_x)^{1/2}$	0.1340	0.1063	0.1448	0.1343
Γ_{xs}	0.0012	0.0110	0.7138	0.1079
$(\Gamma_{xs})^{1/2}$	0.0344	0.1050	0.8449	0.3285

The use of the PCA- and DCT-based features does not help to reduce the condition number of the correlation matrix of the endmember spectra, as compared with the use of the original hyperspectral signals. On the contrary, the use of the PCA-based features increases the condition number for Experiment III of the three-endmember case, where a very small eigenvalue, $\lambda_1 = 0.0008$, occurs. The use of the DCT-based features dramatically increases the condition number for both Experiment II, where $\lambda_1 = 0.0041$, and Experiment III, where $\lambda_1 = 0.0001$. Note that in Sections 5.2-5.4 it is argued that the use of the PCA- and DCT-based features provides the potential to reduce the computational cost of abundance estimation, though it does not help to improve the abundance estimation performance. However, based on the experimental results about the eigenvalues and condition number of correlation matrix of the endmember spectra in this section, it has to be argued that the reduction of computational cost from using the PCA- and DCT-based features comes with a loss of stability of the abundance estimation.

On the other hand, however, the use of the DWT-based features has all the advantages: (i) the improvement of the abundance estimation; (ii) the reduction of computational cost of the abundance estimation; (iii) the improvement of the stability of abundance estimation.

5.7. Discussion about Assumptions in Mathematical Derivations of Abundance Estimation

In Section 2.1, equation (2.10) defines a mean square error (MSE), Γ_x, of abundance estimation,

$$\Gamma_x = E\left[\frac{\vec{e}_x{}^T \vec{e}_x}{M}\right]. \tag{5.6}$$

In Section 2.2, Γ_x is further derived as equation (2.14) without any assumption,

$$\Gamma_x = \frac{1}{M} Tr\left[(A^+)^T A^+ R_o\right]. \tag{5.7}$$

In Section 2.3, given *Assumptions 2.1.1* and *2.4*, Γ_x is further simplified as equation (2.25),

$$\Gamma_{xs} = \frac{1}{M} Tr\left[R^{-1}\right]\sigma_o{}^2. \tag{5.8}$$

This simplified expression shows that the MSE of abundance estimation is uniquely determined by the two parameters: correlation matrix R (or more specifically $Tr[R^{-1}]$) and variance $\sigma_o{}^2$, for given number of endmembers, M. As discussed in Section 2.3, the use of the simplified version of MSE, Γ_{xs}, provides a further insight into how the abundance estimation performance can be improved by using appropriate feature

extraction approaches to reduce the within-endmember variance and increase the between-endmember variance. Unfortunately, however, in order to use the simplified version, Γ_{xs}, *Assumptions 2.1.1* and *2.4* have to hold.

To verify the appropriateness of the assumptions, the values of $Tr[R^{-1}]$, σ_o^2, Γ_x and Γ_{xs} are quantitatively calculated for Experiments I, II and III. Γ_x is calculated using equation (2.14) or (5.7), and Γ_{xs} is calculated using equation (2.25) or (5.8). The results are also summarized in Tables 5.5, 5.6 and 5.7 for Experiments I, II and III, respectively. Note that among the uses of the three types of features and the original hyperspectral signals, only the use of the DWT-based features produces the close values of Γ_x and Γ_{xs}. This observation shows that *Assumptions 2.1.1* and *2.4* for deriving the simplified MSE expression, Γ_{xs}, are appropriate when using the DWT-based features. That is, the values of variance σ_o^2 listed in Tables 5.5-5.7, corresponding to the use of original hyperspectral signals, the PCA- or DCT-based features, are not accurate. In other words, it is meaningless to compare those values of variance σ_o^2.

Recall that equation (2.30) builds a relationship between $Tr[R^{-1}]$ and the eigenvalues λ_i of correlation matrix R,

$$Tr[R^{-1}] = \sum_{i=1}^{M} \frac{1}{\lambda_i}. \tag{5.9}$$

From the experiment results in Tables 5.5-5.7, clearly it can be seen how the value of $Tr[R^{-1}]$ is affected by the eigenvalues λ_i. Taking the results of Experiment III as an

example, the two eigenvalues, λ_2 and λ_3, are comparable for the use of the three types of features and the original hyperspectral signals. However, as for a third eigenvalue, λ_1, the DCT-based feature has a very small value, $\lambda_1 = 0.0001$, thus correspondingly a very large value of $Tr[R^{-1}]$ is produced.

Since the assumptions for deriving the simplified expression, Γ_{xs}, are valid when using the DWT-based feature, it is interesting to compare experiment results in Tables 5.5-5.7 for the DWT case among the three experiment cases. First of all, note that the values of $(\Gamma_{xs})^{1/2}$, or $(\Gamma_x)^{1/2}$, are ~0.07, ~0.2 and ~0.11 for Experiments I, II and III, respectively. They are close to the RMSE values obtained from the experiments in Sections 5.2, 5.3 and 5.4, where the RMSE values using the ULSE method are ~0.1, ~0.18, ~0.12 for the three experiments, respectively. This observation matches the discussion about the relationship between the RMSE evaluation metric and the square root value of Γ_x in Section 4.3. That is, the RMSE is an approximation of the square root value of Γ_x for limited number of samples.

Secondly, note that for Experiment I, II and III, the values of $Tr[R^{-1}]$ are ~14, ~22 and ~52, and the values of σ_o^2 are ~0.0008, ~0.004 and ~0.0006, respectively. Recall that in Chapter II it is argued that $Tr[R^{-1}]$ reflects a between-endmember variance to a certain extent, and a small value of $Tr[R^{-1}]$ indicates a large between-endmember variance. σ_o^2 is a within-endmember variance. A better abundance estimation (or a smaller MSE of abundance estimation) results from a smaller within-endmember

variance and a larger between-endmember variance. Based on these arguments, note that for Experiment I and III cases, the values of $Tr[R^{-1}]$ are ~14 and ~52, and the values of σ_o^2 are ~0.0008 and ~0.0006, respectively. This means that Experiment I case has a similar within-endmember variance to Experiment III case, but a larger between-endmember variance than Experiment III case. Thus, Experiment I case has a smaller MSE of abundance estimation than Experiment III case, even taking into account the difference of factor $\frac{1}{M}$ in equation (5.8), *i.e.*, $M = 2$ for the two-endmember case and $M = 3$ for the three-endmember case.

In summary, the experimental analyses in this section further show the advantages of using the wavelet-based pre-processing and feature extraction for improving the LSE of abundances.

5.8. Discussion about Alternative DCT and PCA Features

In Section 4.1, it is argued that the conventional method for DCT- and PCA-based feature extraction is simply the use of the first few transform coefficients. The conventional method works well when goal is signal representation, such as the case of signal compression. However, when the goal is signal classification, such as the case of endmember abundance estimation, differences between signals take on importance, and simply using the first few large-amplitude coefficients may not be adequate. Therefore, an alternative approach is proposed for selecting a subset of transform coefficients. This alternative approach utilizes a sliding window of size L to select coefficient subsets of size L.

In this section, the endmember abundance estimation based on the alternative DCT- and PCA-based features is implemented for the same three sets of experiments, as a comparison with the use of the conventional DCT- and PCA-based features. Four different window sizes are investigated in this study. For the alternative DCT-based features, the four window sizes are 5, 10, 50, and 100. For the alternative PCA-based feaures, note that in Experiment I the absolute differences of PCA coefficients for soybean and soil are less than 10^{-12} after the 20^{th} PCA coefficient. Thus, only the first 20 PCA coefficients are investigated for the alternative feature extraction method in Experiment I. For the same reason, the first 20 and 30 PCA coefficients are investigated in Experiments II and III, repectively. Based on these observations, the four window sizes investigated for the alternative PCA-based features are 2, 3, 4, and 5. Optimal feature sets are firstly selected through the system training, and then are quantitatively evaluated during the system testing.

Based on the training data sets, the CLSE of endmember abundances are implemented using the alternative PCA- and DCT-based features, and the RMSE of abundance estimation is calculated. The RMSE value is utilized as the criterion of selecting the optimal feature sets. Two optimal parameters are recorded for the optimal feature sets. One is the window size, and the other is the window shift.

Figures 5.22 shows the training results from the use of the alternative DCT-based features in Experiment I. The smallest RMSE is obtained when the window size is 5 and the window shift is 13. That is, the optimal feature set consists of the DCT coefficients from the 66^{th} to the 70^{th} coefficients, which are a set of high-frequency DCT coefficients.

Note that the use of this set of optimal DCT-based features results in a better CLSE of endmember abundances than the use of the conventional DCT-based features or the use of the original hyperspectral signals. That is, the use of certain sets of high-frequency (or detailed) DCT coefficients improves the endmember abundance estimation. This is similar to the use of the detailed DWT coefficients. These results show that features based on optimal signal approximations are not necessarily good features for improving the endmember abundance estimation. Many other features, such as features based on high-frequency DCT coefficient subsets and DWT detail coefficient subsets, have abilites to improve the endmember abundance estimation.

Figures 5.23 shows the training results from the use of the alternative PCA-based features in Experiment I. The smallest RMSE is obtained when the window size is 3 and the window shift is 0. Note that this is the same result as the use of conventional PCA-based features. These results show that features formed by simply using the subsets of PCA coefficients does not help for improving the endmember abundance estimation, except for reducing the dimensionality of hyperspectral signals and thus the computational expenses.

Similar training results are obtained in Experiments II and III. Figures 5.24 and 5.26 show the training results from the use of the alternative DCT-based features in Experiments II and III, respectively. Figures 5.25 and 5.27 show the training results from the use of the alternative PCA-based features in Experiments II and III, respectively. The optimal window parameters are recorded in Table 5.8.

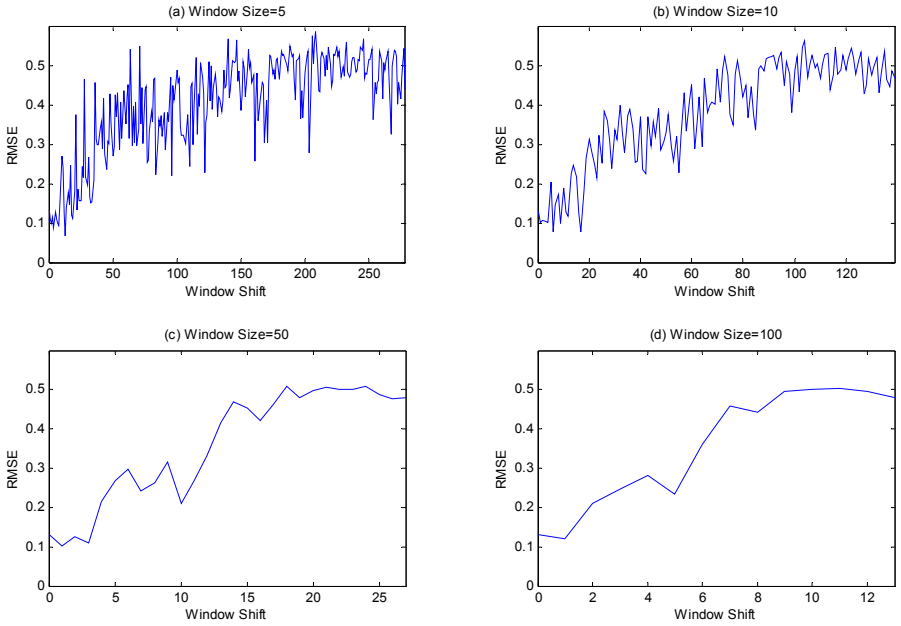

Figure 5.22. RMSE of abundance estimation based on training data set when using the sliding window method for DCT-based feature extraction in Experiment I.

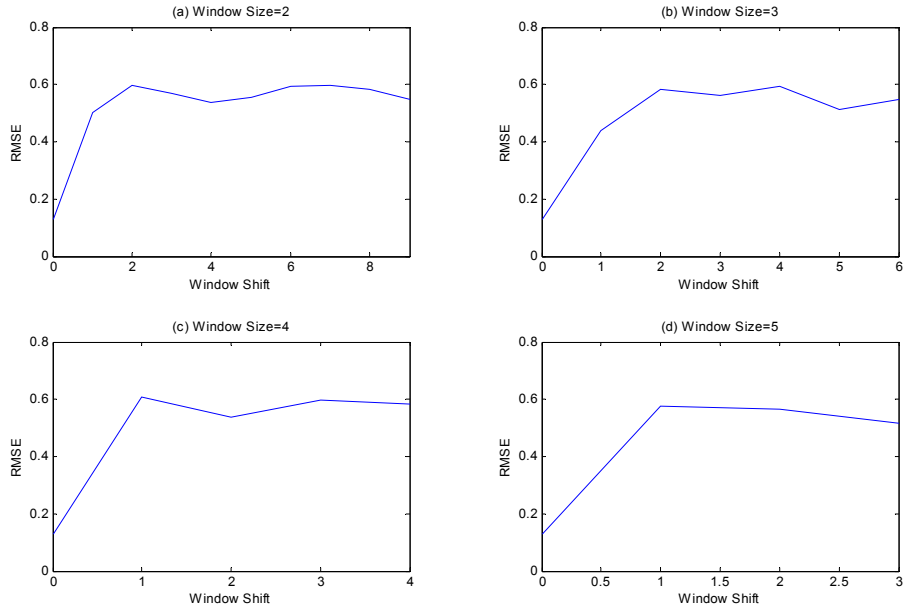

Figure 5.23. RMSE of abundance estimation based on training data set when using the sliding window method for PCA-based feature extraction in Experiment I.

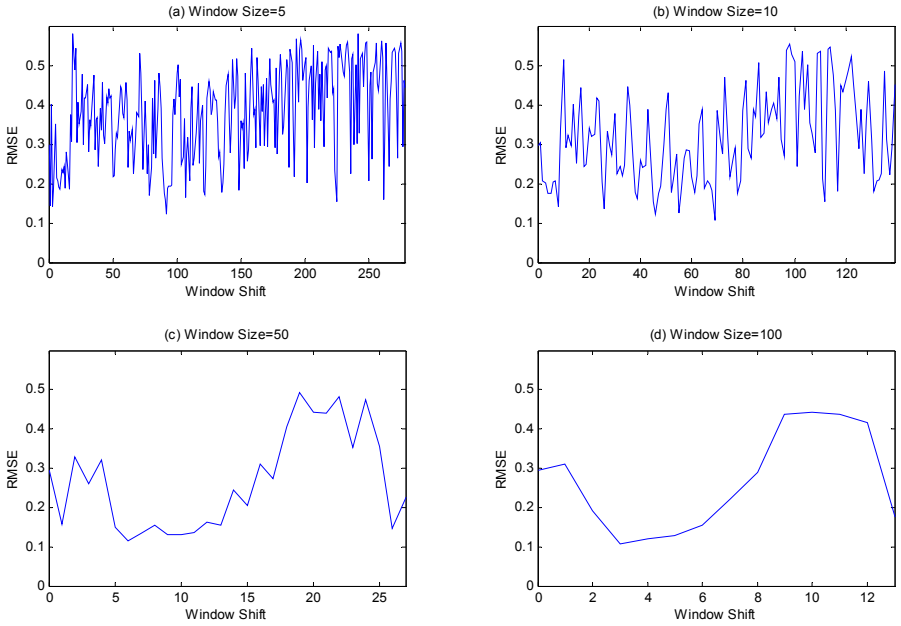

Figure 5.24. RMSE of abundance estimation based on training data set when using the sliding window method for DCT-based feature extraction in Experiment II.

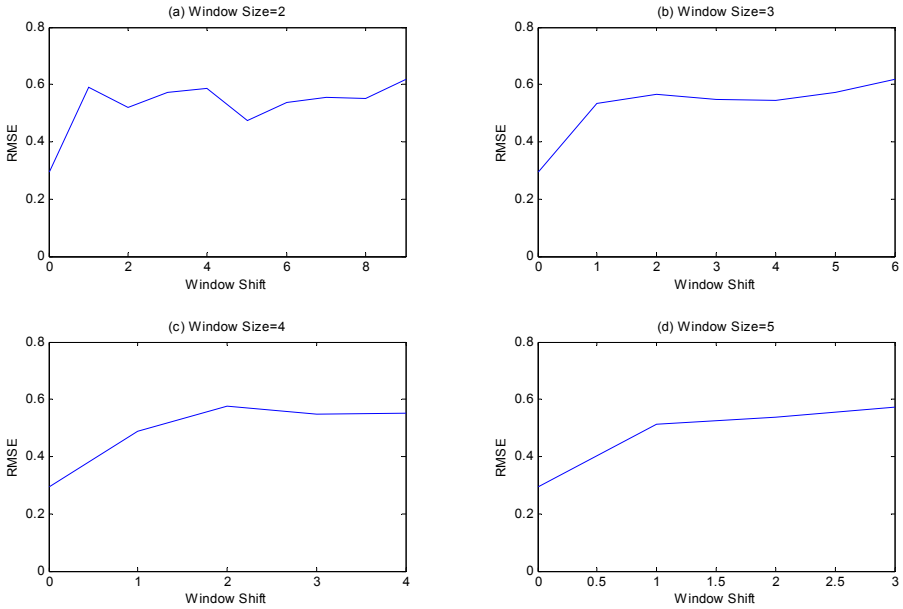

Figure 5.25. RMSE of abundance estimation based on training data set when using the sliding window method for PCA-based feature extraction in Experiment II.

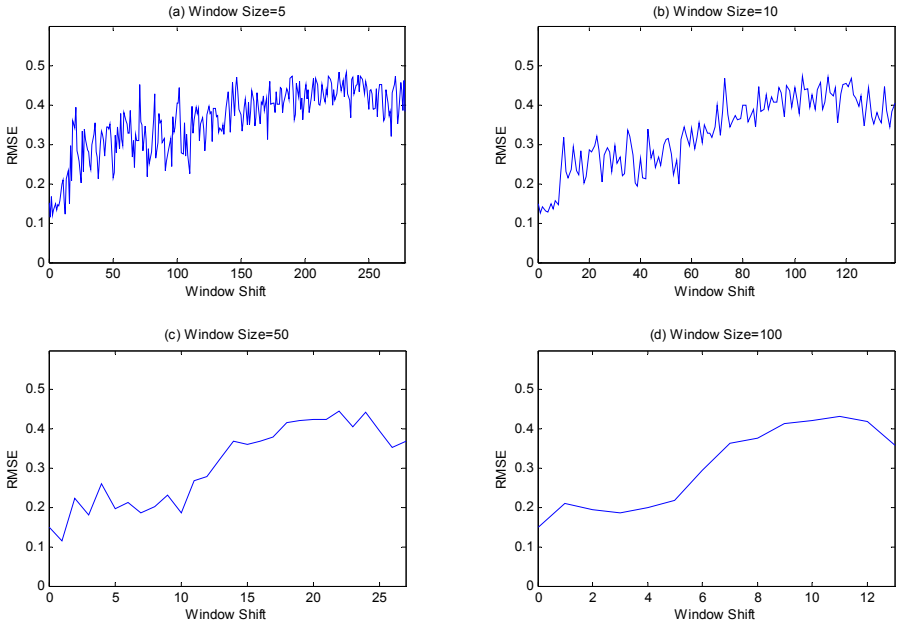

Figure 5.26. RMSE of abundance estimation based on training data set when using the sliding window method for DCT-based feature extraction in Experiment III.

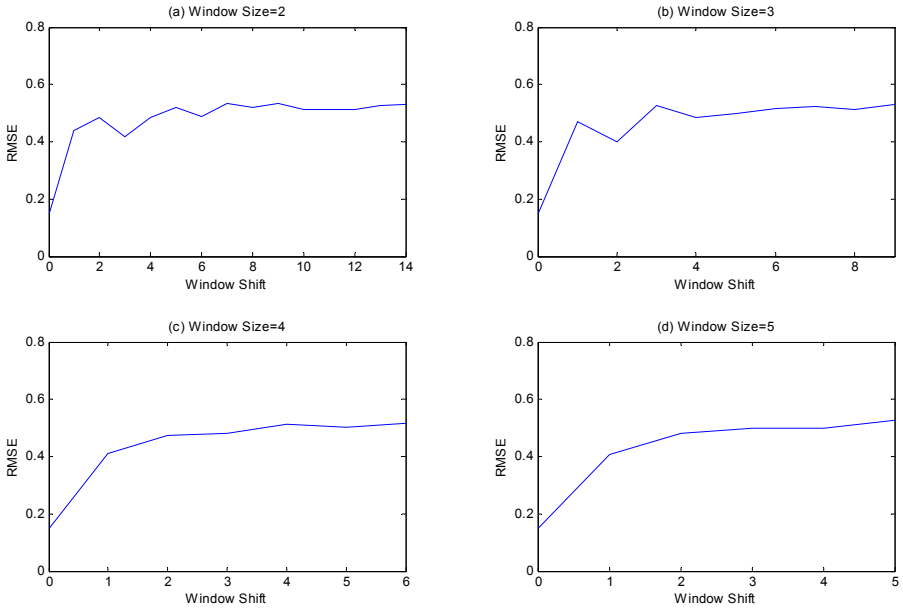

Figure 5.27. RMSE of abundance estimation based on training data set when using the sliding window method for PCA-based feature extraction in Experiment III.

Table 5.8. Optimal window parameters when using the sliding window method for DCT- and PCA-based feature extraction in Experiment I, II and III.

		Window Size	Window Shift
Experiment I	DCT	5	13
	PCA	3	0
Experiment II	DCT	100	3
	PCA	4	0
Experiment III	DCT	50	1
	PCA	3	0

Based on the testing data sets, the CLSE of endmember abundances are implemented using these optimal feature sets. The CLSE results are quantitatively evaluated by using two metrics: the RMSE of abundance estimation and the confidence of abundance estimation. For DCT-based features, the RMSE results are shown in Figure 5.28, 5.30 and 5.32, for Experiment I, II and III, respectively. The confidence curves of the CLSE of abundances for the three sets of experiments are shown in Figure 5.29, 5.31 and 5.33, respectively. Clearly, it can be seen that using the alternative DCT-based features results in better abundance estimation than using the conventional DCT-based features, which are based on optimal signal approximation, and the original hyperspectral signals. Also note that while the use of the alternative DCT-based features improves the endmember abundance estimation, the DWT-based features still perform better than the alternative DCT-based features.

As for PCA-based features, since the same optimal feature sets are obtained from using the conventional and alternative feature extraction methods, the testing results are the same for both cases and available in Sections 5.2, 5.3 and 5.4.

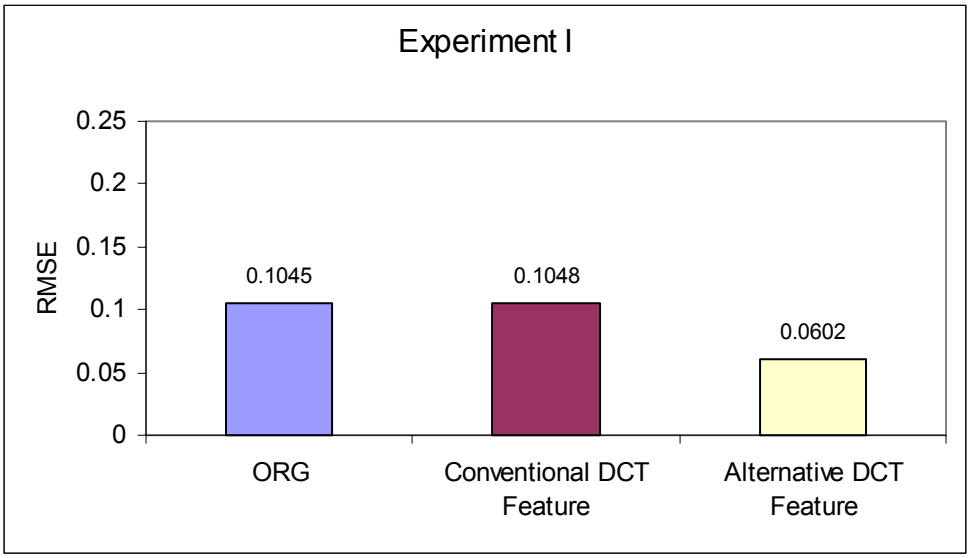

Figure 5.28. RMSE of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment I (soybean vs. soil).

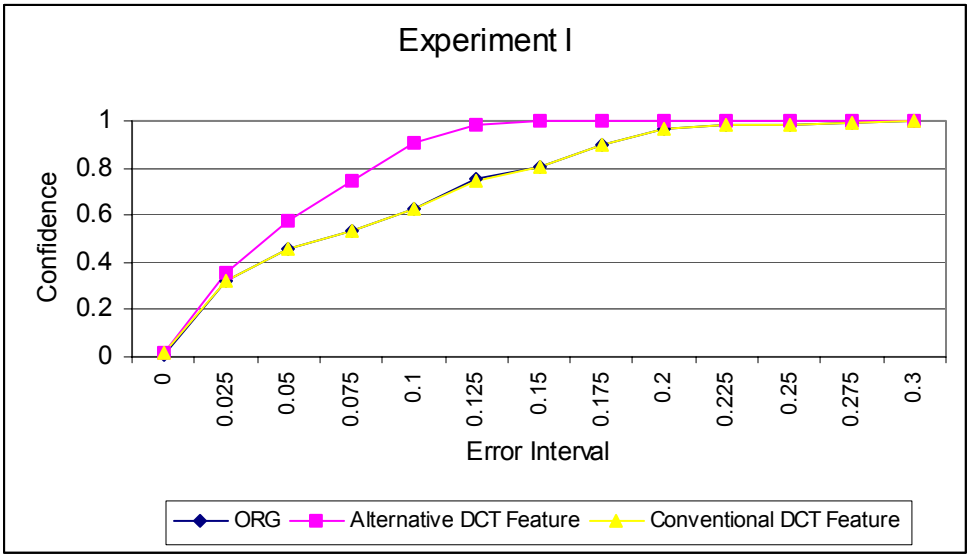

Figure 5.29. Confidence of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment I (soybean vs. soil).

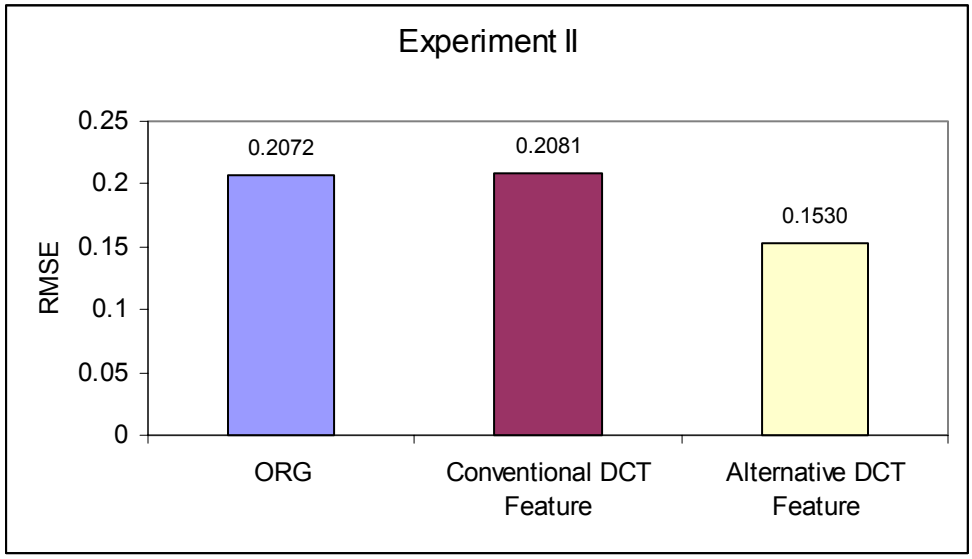

Figure 5.30. RMSE of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment II (soybean vs. grass).

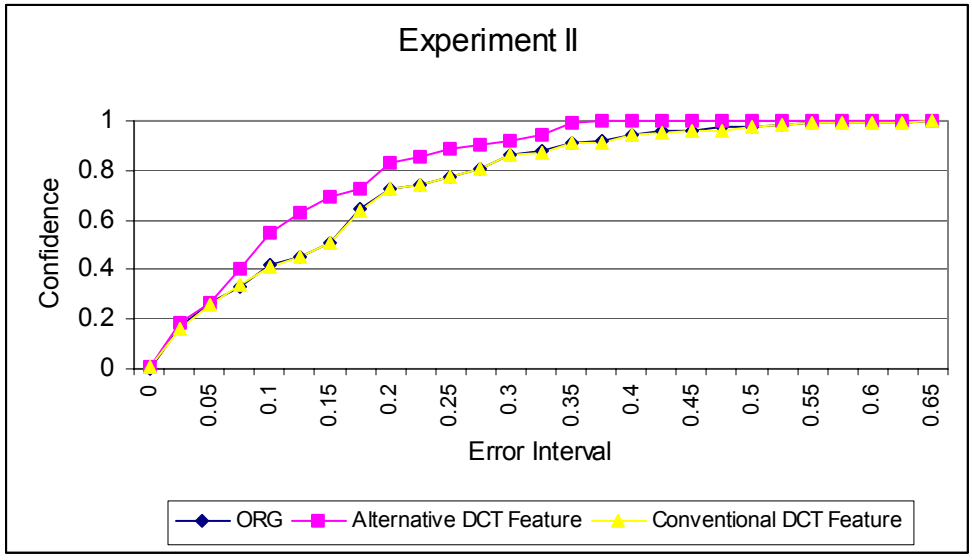

Figure 5.31. Confidence of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment II (soybean vs. grass).

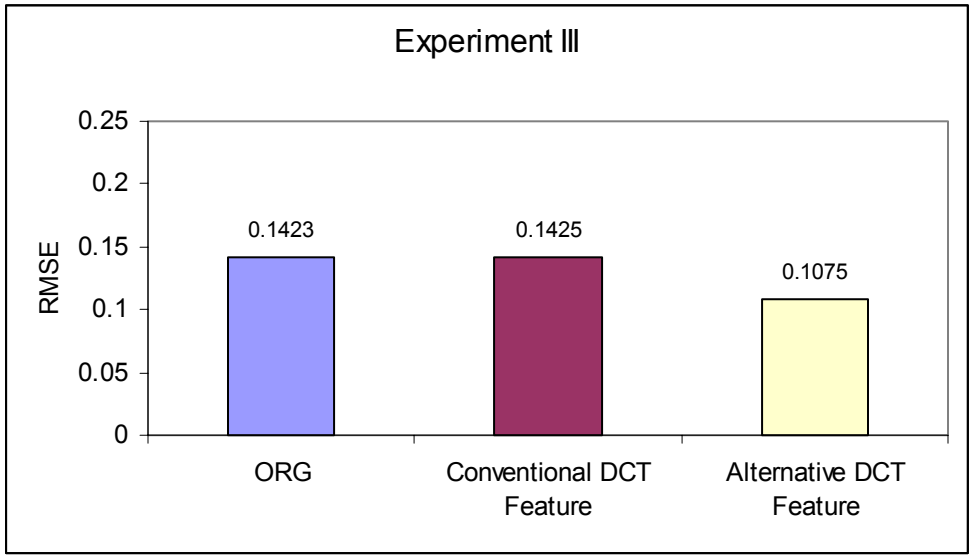

Figure 5.32. RMSE of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment III (soybean, grass vs. soil).

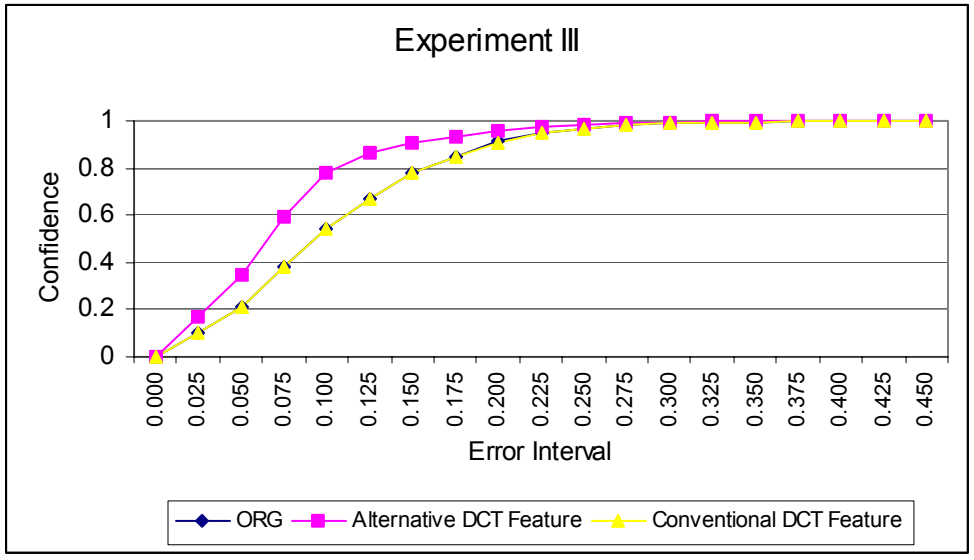

Figure 5.33. Confidence of abundance estimation based on testing data set for comparison of the conventional and alternative DCT features in Experiment III (soybean, grass vs. soil).

CHAPTER VI

CONCLUSIONS

This dissertation provides a complete investigation on how the use of the appropriate features can benefit the linear spectral unmixing using hyperspectral signals. Both experimental verification and theoretical analysis are reported in the dissertation. The feature extraction is proposed as a pre-processing step before the linear spectral unmixing. It is proven that this pre-processing step is a key to improving the performance of the linear spectral unmixing, specificially the least squares estimation (LSE) of endmember abundances.

The use of appropriate features extracted from the hyperspectral endmember signals provides a potential to increase the separability among the endmember spectra, which is a fundamental reason for improving the LSE of endmember abundances. The separability is typically measured by the within-endmember variance and between-endmember variance. Specifically, the endmember separability is increased when the within-endmember variance is reduced and/or the between-endmember variance is increased. In the target detection and classification applications, it has been proven that the use of the appropriate features, such as the discrete wavelet transform (DWT) based features, extracted from the hyperpsectral signals, can reduce the within-class variance and increase the between-class variance [5-6, 39-41]. As a result, the class separability is increased and the classification performance is improved. Motivated by the successful

applications of the feature extraction to the target detection and classification using the hyperspectral signals, the dissertation investigates the potential of the feature extraction technique in the linear spectral unmixing applications using the hyperspectral signals. The results show that the linear spectral unmixing using hyperspectral signals is another successful application of the feature extraction technique.

Feature extraction based on the DWT is the primary focus in this dissertation, since it has been successfully applied to target detection and classification using hyperspectral signals. Based on the DWT-based feature extraction, a linear unmixing system is designed specially for the abundance estimation of endmembers. The system utilizes the DWT for the feature extraction, which is also referred to as a pre-processing step before the linear spectral unmixing. In this study, a DWT-based feature set consists of the DWT detail or approximation coefficients at a specific DWT scale. Based on the DWT-based features, at the linear unmixing step the system utilizes the constrained LSE (CLSE) for the abundance estimation of endmembers. The use of the CLSE method makes the abundance estimation results physically meaningful. The abundance estimation results are quantitatively evaluated to show whether or not and to how much extent the pre-processing based on the feature extraction improves the performance of abundance estimation. The quantitative evaluation metrics used in the dissertation are: (i) the root mean square error (RMSE) of the abundance estimation; (ii) the confidence of the abundance estimation; and (iii) the abundance distribution diagram.

Based on the proposed DWT-based linear unmixing system, three sets of experiments are designed and implemented to quantitatively evaluate the proposed DWT-

based features. Experiment I investigates a two-endmember linear unmixing problem and the endmembers consist of a vegetation (soybean) and a non-vegetation (soil). Experiment II investigates another two-endmember linear unmixing problem and the endmembers consist of two vegetations (soybean and grass). Experiment III investigates a multi-endmember linear unmixing problem and the endmembers include two vegetations (soybean and grass) and a non-vegetation (soil).

Results from the three sets of experiments show that generally the use of the proposed DWT-based features reduces the average deviation of abundance estimation by 30-50%, as compared with the use of the original hyperspectral signals without the pre-processing. Moreover, with a 95% confidence of abundance estimation, the average estimation deviation from the true abundance is ~0.1, ~0.275, and ~0.175 for Experiments I, II and III, respectively, when using the DWT-based features. However, when using the original hyperspectral signals without the pre-processing, with the same confidence the average deviation is ~0.2, ~0.425, and ~0.225 for Experiments I, II and III, respectively. The abundance distribution diagrams for the three sets of experiments also show that the abundance estimation results from the use of the DWT-based features have a much closer distribution to the true abundances than the abundance estimation results from the use of the original hyperspectral signals without pre-processing. These results show a great potential of using the DWT-based feature extraction and pre-processing for improving the LSE of endmember abundances.

Realizing these promising experiment results, the dissertation further investigates the underlying fundamental reasons leading to these results. Based on the linear mixture

model (LMM) and the LSE method, a series of theoretical analyses are derived to reveal the fundamental reasons why the use of the appropriate features, such as the DWT-based features, can improve the LSE of endmember abundances. Under some reasonable assumptions, the dissertation builds a mathematical relationship among the mean square error (MSE) of the abundance estimation, the between-endmember variance, and within-endmember variance. In the mathematical relationship, the between-endmember variance is indicated by the correlation matrix of endmmeber spectra, and the within-endmember variance is a weighted sum of each endmember variance. That is, the MSE of the abundance estimation is uniquely determined by the correlation matrix and variances of endmember spectra. The MSE of the abundance estimation can be decreased through reducing the correlation of endmember spectra, *i.e.*, consequently increasing the between-endmember variance, and reducing each endmember variance. In other words, the abundance estimation performance can be improved through increasing the endmember separability.

The stability of the least square solution of the abundance estimation is also theoretically analyzed using the concept of the condition number. A smaller condition number of the correlation matrix of endmember spectra indicates a more stable least square solution of the abundance estimation. The experiment results from the three sets of experiments show that the condition numbers are greatly reduced when using the DWT-based features extracted from the original hyperspectral signals of endmembers. That is, the use of the DWT-based features not only improves the abundance estimation of

endmembers, but also leads to a more stable LSE of endmember abundances, as compared to the use of the original hyperspectral signals.

Features based on the principal components analysis (PCA) and the discrete cosine transform (DCT) are also investigated for a comparison with the DWT-based features. Firstly, PCA- and DCT-based features are formed utilizing the first few transform coefficients. This is a conventional feature extraction method that is based on the optimal signal representation. The same three sets of experiments are performed on the conventional PCA- and DCT-based features. The experiment results show that the use of the conventional PCA- and DCT-based features does not help to improve the abundance estimation at all, other than reducing the dimensionality of hyperspectral signals and thus reducing the computational cost of the abundance estimation. However, the dimensionality reduction resulting from the use of the conventional PCA- and DCT-based features does come with a loss of stability of the abundance estimation. The comparison results show that the use of the DWT-based features has all the advantages: (i) the improvement of the abundance estimation; (ii) the dimensionality reduction of hyperspectral signals and the reduction of computational cost of the abundance estimation; and (iii) the improvement of the stability of abundance estimation.

An alternative method is also investigated for PCA- and DCT-based feature extraction. Rather than using only first few coefficients as features, the alternative method utilizes a sliding window of size L to extract other subset coefficient features of size L. Experimental results show that the use of the subset features of high-frequency DCT coefficients can improve the endmember abundance estimation. In general, when

dealing with hyperpsectral signals, traditionally the dimensionality reduction has been based on methods that provide superior energy compaction, such as PCA and DCT. The reduction of dimensionality has stemmed from the use of only the first few transform coefficients. The approach works well when the aim is signal representation, such as the case of signal compression. However, this approach may be misguided when the aim is signal classification, which is also the aim of spectral unmxing. In this case, differences between signals take on importance, and simply using the first few large-amplitude transform coefficients may not be adequate. Therefore, there is a great need for the remote sensing community to investigate feature extraction methods that are based on signal classification for linear unmixing problems. For the same reason, the DWT-based approach is proposed in this dissertation for linear unmixing of hyperspectral signals.

Four aspects of future work are recommended: (i) to investigate other advanced feature extraction techniques; (ii) to investigate other mother wavelets, as well as the multi-channel DWT and the wavelet packets [58, 59]; (iii) to investigate the applications of the proposed DWT-based linear unmixing system to the two-dimensional hyperspectral images; and (iv) to investigate the weighted least squares technique [21, 22] based on the feature extraction for the endmember abundance estimation.

The pre-processing based on the feature extraction plays a key role in the proposed linear unmixing system using the hyperpsectral signals. In the dissertation, features based on the DWT, PCA and DCT are extracted and investigated. For each of them, only one feature extraction method is proposed. It is shown that the proposed DWT-based feature extraction method is successful in improving the abundance

estimation performance, but the proposed PCA- and DCT-based feature extraction methods do not provide good features for improving the abundance estimation. Based on the theoretical analysis in the dissertation, any feature set that increases the endmember separability can be a good feature set for improving the abundance estimation performance. Thus, it will be interesting to investigate other advanced feature extraction methods based on not only the DWT, PCA, DCT but also other techniques in the future.

For the DWT-based feature extraction, only one mother wavelet, the Haar mother wavelet, is investigated in the dissertation. Since there exist many other mother wavelets, it will be interesting to investigate the DWT-based features using other mother wavelets in the future. Moreover, in this study only the dyadic (or two-channel) DWT implementation is investigated. Since the multi-channel DWT could provide a more complete multiresolution analysis of signals, it could be worthwhile to investigate feature extraction based on the multi-channel DWT in the future. Note that the DWT implements an iterative decomposition of only signal approximations. Wavelet packet (WP) technique [58, 59] provides an iterative decomposition of both signal approximations and signal details. Feature sets could be obtained from optimal WP trees. Thus, it will also be interesting to investigate the WP for hyperspectral feature extraction in the future.

The proposed DWT-based linear unmixing system is evaluated on the one-dimensional hyperspectral signals in the dissertation. However, the proposed system is generic and can be utilized to implement the linear unmixing analysis based on two-dimensional hyperspectral images. Working on the one-dimensional hyperspectral signals, only the spectral information is utilized for the feature extraction. However, both

spectral and spatial information can be utilized for the feature extraction when working on the two-dimensional hyperspectral images. Thus, it will be interesting to investigate the applications of the proposed DWT-based linear unmixing system to the two-dimensional hyperspectral images in the future.

The LSE based on the feature extraction is investigated for the endmember abundance estimation in the dissertation, where each feature is treated equally when the least squares criterion is applied. However, it is possible to treat each feature differently according to its importance to the LSE. That is, based on *a priori* information, larger weights are assigned to the features that could have larger effects in reducing the LSE error. As a result, the LSE error could be further reduced, as compared to treating each feature equally. This is a process of the feature optimization. According to the LMM of equation (1.1), since weighted features of endmember and mixed-pixel spectra lead to weighted errors, this is also a process of the weighted LSE (WLSE). Therefore, it will be interesting to investigate the WLSE based on the feature extraction for the endmember abundance estimation in the future.

APPENDIX A

PROOF OF EXISTENCE OF $(A^T A)^{-1}$ IN EQUATION (2.7)

Theorem: The inverse of the auto-correlation matrix $R = A^T A$ exists if and only if the column vectors in matrix A are linearly independent.

Proof: If the column vectors in matrix A are linearly independent, then for all non-zero vectors, $\vec{b}$, there exists $A\vec{b} \neq \vec{0}$, and thus,

$$(A\vec{b})^T (A\vec{b}) = \left\| A\vec{b} \right\|^2 > 0 . \tag{A.1}$$

where $\|\cdot\|$ is referred to as a vector norm. Note that,

$$\vec{b}^T R\vec{b} = \vec{b}^T A^T A\vec{b} = (A\vec{b})^T (A\vec{b}) , \tag{A.2}$$

that is, for all non-zero vectors, $\vec{b}$, there exists $\vec{b}^T R\vec{b} > 0$. Therefore, the real symmetric matrix R is positive definite, according to the necessary and sufficient condition for a real symmetric matrix R to be positive definite [19]. For a positive definite matrix R, there exists,

$$Det[R] > 0 , \tag{A.3}$$

where $Det[\cdot]$ represents an operation of matrix determinant. Since the determinant of matrix R is non-zero, the inverse of matrix R exists.

If the column vectors in matrix A are linearly dependent, then there exists a non-zero vectors, $\vec{b}_o$, such that $A\vec{b}_o = \vec{0}$. Thus, there exists,

$$A^T A\vec{b}_o = R\vec{b}_o = \vec{0} , \tag{A.4}$$

for a non-zero vectors, $\vec{b}_o$. That is, the column vectors in matrix R are linearly dependent. Therefore, matrix R is not full rank, *i.e.*, matrix R is not invertible.

APPENDIX B

DERIVATION OF EQUATION (2.14)

Rewriting equations (2.12) and (2.13) as,

$$\vec{e}_x = (A^T A)^{-1} A^T \vec{e}_o , \tag{B.1}$$

$$\Gamma_x = \frac{1}{M} E\left[Tr[\vec{e}_x{}^T \vec{e}_x]\right], \tag{B.2}$$

and substituting equation (B.1) into equation (B.2), we have,

$$\begin{aligned} \Gamma_x &= \frac{1}{M} E\left[Tr[\vec{e}_x{}^T \vec{e}_x]\right] \\ &= \frac{1}{M} E\left[Tr[(A^+ \vec{e}_o)^T (A^+ \vec{e}_o)]\right], \\ &= \frac{1}{M} E\left[Tr[\vec{e}_o{}^T (A^+)^T A^+ \vec{e}_o]\right] \end{aligned} \tag{B.3}$$

where $A^+ = (A^T A)^{-1} A^T$ is defined as a pseudo-inverse of matrix A. Utilizing the property of the matrix trace operation: given any two matrices, U and V, there exists,

$$Tr[UV] = Tr[VU], \tag{B.4}$$

equation (B.3) can be further derived, as equation (2.14) as,

$$\begin{aligned} \Gamma_x &= \frac{1}{M} E\left[Tr[(A^+)^T A^+ \vec{e}_o \vec{e}_o{}^T]\right] \\ &= \frac{1}{M} Tr\left[(A^+)^T A^+ E[\vec{e}_o \vec{e}_o{}^T]\right] \\ &= \frac{1}{M} Tr\left[(A^+)^T A^+ R_o\right] \end{aligned} . \tag{B.5}$$

where $R_o = E[\vec{e}_o \vec{e}_o{}^T]$ is a correlation matrix of random measurement error vector $\vec{e}_o$.

APPENDIX C

DERIVATION OF EQUATIONS (2.22) AND (2.27)

Starting with equation (2.19),

$$\vec{e}_o = \Delta_A \vec{x}_o, \tag{C.1}$$

where vector $\vec{x}_o = [x_{o1}, x_{o2}, \cdots, x_{oM}]^T$ represents true abundances of endmember spectra, as defined in equation (1.3), and matrix $\Delta_A = [\vec{\delta}_1, \vec{\delta}_2, \cdots, \vec{\delta}_M]$ represents the differences between the library endmember spectra and the true endmember spectra constituting the mixed-pixels, then the correlation matrix of $\vec{e}_o$ can be computed as,

$$\begin{aligned} E[\vec{e}_o \vec{e}_o{}^T] &= E[\Delta\vec{x}_o (\Delta\vec{x}_o)^T] \\ &= E\left[\left(\sum_{i=1}^{M} x_{oi}\vec{\delta}_i\right)\left(\sum_{k=1}^{M} x_{ok}\vec{\delta}_k\right)^T\right], \\ &= \sum_{i=1}^{M}\sum_{k=1}^{M} x_{oi}x_{ok} E[\vec{\delta}_i \vec{\delta}_k{}^T] \end{aligned} \tag{C.2}$$

where the i^{th} column vector, $\vec{\delta}_i = [\delta_{i1}, \delta_{i2}, \cdots, \delta_{iN}]^T$ (for $i = 1,2,\cdots,M$), in matrix Δ_A represents the difference between the i^{th} library and true endmember spectra, where M is the number of endmembers and N is the number of spectral bands.

Based on *Assumptions 2.3.1*, *2.3.2* and *2.3.3*, we have,

$$E[\vec{\delta}_i \vec{\delta}_k{}^T] = 0, \text{ for } i \neq k. \tag{C.3}$$

Therefore, in equation (C.2), only terms for $i = k$ left, that is,

$$E[\vec{e}_o \vec{e}_o{}^T] = \sum_{i=1}^{M} x_{oi}{}^2 E[\vec{\delta}_i \vec{\delta}_i{}^T]. \tag{C.4}$$

Based on *Assumptions 2.3.3*, we have equation (2.21), *i.e.*,

$$E[\vec{\delta}_i \vec{\delta}_i^{\,T}] = \begin{bmatrix} \sigma_{i1}^{\,2} & 0 & \cdots & 0 \\ 0 & \sigma_{i2}^{\,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{iN}^{\,2} \end{bmatrix}. \tag{C.5}$$

Substituting equation (C.5) into equation (C.4), we have,

$$E[\vec{e}_o \vec{e}_o^{\,T}] = \begin{bmatrix} \sum_{i=1}^{M} x_{oi}^{\,2} \sigma_{i1}^{\,2} & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^{M} x_{oi}^{\,2} \sigma_{i2}^{\,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sum_{i=1}^{M} x_{oi}^{\,2} \sigma_{iN}^{\,2} \end{bmatrix}. \tag{C.6}$$

According to equation (2.16), we have,

$$E[\vec{e}_o \vec{e}_o^{\,T}] = \begin{bmatrix} \sigma_{o1}^{\,2} & 0 & \cdots & 0 \\ 0 & \sigma_{o2}^{\,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{oN}^{\,2} \end{bmatrix}. \tag{C.7}$$

Comparing equations (C.6) and (C.7), we can obtain equation (2.22),

$$\sigma_{oj}^{\,2} = \sum_{i=1}^{M} \sigma_{ij}^{\,2} x_{oi}^{\,2}, \quad (j = 1,2,\cdots,N), \tag{C.8}$$

where M is the number of endmembers and N is number of spectral bands.

Rewriting equation (2.26) as,

$$[\vec{\delta}_i \vec{\delta}_i^{\,T}] = \sigma_i^{\,2} I, \tag{C.9}$$

and substituting equation (C.9) into equation (C.4), we have,

$$E[\vec{e}_o \vec{e}_o^{\,T}] = \sum_{i=1}^{M} x_{oi}^{\,2} \sigma_i^{\,2} I. \tag{C.10}$$

According to equation (2.24), we have,

$$E[\vec{e}_o \vec{e}_o{}^T] = \sigma_o{}^2 I \,, \tag{C.11}$$

Comparing equations (C.10) and (C.11), we can obtain equation (2.27),

$$\sigma_o{}^2 = \sum_{i=1}^{M} \sigma_i{}^2 x_{oi}{}^2 \,. \tag{C.12}$$

where M is the number of endmembers.

APPENDIX D

DERIVATION OF EQUATION (2.25)

Rewriting equations (2.14) and (2.24) as,

$$\Gamma_x = \frac{1}{M} Tr\left[(A^+)^T A^+ R_o\right], \tag{D.1}$$

$$R_o = E[\vec{e}_o \vec{e}_o{}^T] = \sigma_o{}^2 I\,, \tag{D.2}$$

where $A^+ = (A^T A)^{-1} A^T$ is defined as a pseudo-inverse of matrix A, and substituting equation (D.1) into equation (D.2), we have,

$$\Gamma_{xs} = \frac{1}{M} Tr\left[(A^+)^T A^+ \sigma_o{}^2 I\right] = \frac{1}{M} \sigma_o{}^2 Tr\left[(A^+)^T A^+\right]\,, \tag{D.3}$$

where Γ_{xs} is referred to as a simplified version of Γ_x. Note that,

$$\begin{aligned} (A^+)^T A^+ &= [(A^T A)^{-1} A^T]^T (A^T A)^{-1} A^T \\ &= A[(A^T A)^{-1}]^T (A^T A)^{-1} A^T \\ &= A[(A^T A)^T]^{-1} (A^T A)^{-1} A^T \\ &= A(A^T A)^{-1} (A^T A)^{-1} A^T \end{aligned}\,. \tag{D.4}$$

Substituting equation (D.4) into equation (D.3), and utilizing the property of the matrix trace operation defined in equation (B.4), we can obtain equation (2.25),

$$\begin{aligned} \Gamma_x &= \frac{1}{M} \sigma_o{}^2 Tr\left[(A^+)^T A^+\right] \\ &= \frac{1}{M} \sigma_o{}^2 Tr\left[A(A^T A)^{-1} (A^T A)^{-1} A^T\right] \\ &= \frac{1}{M} \sigma_o{}^2 Tr\left[(A^T A)^{-1} (A^T A)^{-1} A^T A\right]\,, \\ &= \frac{1}{M} \sigma_o{}^2 Tr\left[(A^T A)^{-1}\right] \\ &= \frac{1}{M} Tr\left[R^{-1}\right] \sigma_o{}^2 \end{aligned} \tag{D.5}$$

where $R = A^T A$ is an auto-correlation matrix of endmember spectra, as defined in equation (2.6).

APPENDIX E

DERIVATION OF EQUATION (2.30)

Staring with equation (2.29),

$$R = Q\Lambda Q^T , \tag{E.1}$$

we have

$$Tr[R^{-1}] = Tr[(Q\Lambda Q^T)^{-1}] = Tr[(Q^T)^{-1}\Lambda^{-1}Q^{-1}], \tag{E.2}$$

where Q is a matrix with each column being an eigenvector of R, and Λ is a diagonal matrix with diagonal elements being eigenvalues of R. Utilizing the property of the matrix trace operation defined in equation (B.4), equation (E.2) can be further derived as,

$$Tr[R^{-1}] = Tr[\Lambda^{-1}Q^{-1}(Q^T)^{-1}]. \tag{E.3}$$

Since Q is an orthogonal matrix, there exists $Q^T = Q^{-1}$ or $Q^TQ = QQ^T = I$. Then we can have,

$$(Q^TQ)^{-1} = Q^{-1}(Q^T)^{-1} = I^{-1} = I, \tag{E.4}$$

where I is an identity matrix. Substituting equation (E.4) into equation (E.3), we can obtain equation (2.30),

$$Tr[R^{-1}] = Tr[\Lambda^{-1}] = \sum_{i=1}^{M}\frac{1}{\lambda_i}, \tag{E.5}$$

where M is the number of endmembers and λ_i are M distinct eigenvalues of R.

APPENDIX F

DERIVATION OF EQUATIONS (2.39) AND (2.40)

Staring with the normal equation (2.38),

$$R\vec{x}_{LS} = \vec{d}\,, \tag{F.1}$$

and supposing that there is a disturbance, $\vec{\delta}_{\vec{d}}$, existing in the cross-correlation vector, $\vec{d}$, which results in an error, $\vec{\delta}_{\vec{x}_{LS}}$, of the least squares solution, $\vec{x}_{LS}$, then we have,

$$R(\vec{x}_{LS} + \vec{\delta}_{\vec{x}_{LS}}) = \vec{d} + \vec{\delta}_{\vec{d}}\,. \tag{F.2}$$

Note that $R\vec{x}_{LS} = \vec{d}$, then equation (F.2) can be further derived as,

$$R\vec{\delta}_{\vec{x}_{LS}} = \vec{\delta}_{\vec{d}}\,. \tag{F.3}$$

Since the inverse of R exists, as proven in Appendix A, solving equation (F.3) for $\vec{\delta}_{\vec{x}_{LS}}$, we obtain,

$$\vec{\delta}_{\vec{x}_{LS}} = R^{-1}\vec{\delta}_{\vec{d}}\,. \tag{F.4}$$

According to equation (2.33), we have,

$$\left\|R\vec{z}\right\| \le \left\|R\right\|\left\|\vec{z}\right\|. \tag{F.5}$$

Utilizing equation (F.5), from equation (F.4), we can derive,

$$\left\|\vec{\delta}_{\vec{x}_{LS}}\right\| = \left\|R^{-1}\vec{\delta}_{\vec{d}}\right\| \le \left\|R^{-1}\right\|\left\|\vec{\delta}_{\vec{d}}\right\|, \tag{F.6}$$

and from equation (F.1), we can derive,

$$\left\|\vec{d}\right\| = \left\|R\vec{x}_{LS}\right\| \le \left\|R\right\|\left\|\vec{x}_{LS}\right\|. \tag{F.7}$$

Multiplying equations (F.6) by (F.7) and rearranging terms, we obtain equation (2.39),

$$\frac{\left\|\vec{\delta}_{\vec{x}_{LS}}\right\|}{\left\|\vec{x}_{LS}\right\|} \le \left\|R\right\|\left\|R^{-1}\right\|\frac{\left\|\vec{\delta}_{\vec{d}}\right\|}{\left\|\vec{d}\right\|} = CN(R)\frac{\left\|\vec{\delta}_{\vec{d}}\right\|}{\left\|\vec{d}\right\|}. \tag{F.8}$$

Suppose that there is a disturbance, δ_R, existing in the auto-correlation matrix, R, which results in an error, $\vec{\delta}_{\vec{x}_{LS}}$, of the least squares solution, $\vec{x}_{LS}$, then we have,

$$(R + \delta_R)(\vec{x}_{LS} + \vec{\delta}_{\vec{x}_{LS}}) = \vec{d}\,. \tag{F.9}$$

Note that $R\vec{x}_{LS} = \vec{d}$, then equation (F.9) can be further derived as,

$$R\vec{\delta}_{\vec{x}_{LS}} = -\delta_R \vec{x}_{LS} - \delta_R \vec{\delta}_{\vec{x}_{LS}}\,. \tag{F.10}$$

Ignoring the second-order term, $\delta_R \vec{\delta}_{\vec{x}_{LS}}$, and solving equation (F.10) for $\vec{\delta}_{\vec{x}_{LS}}$, we obtain,

$$\vec{\delta}_{\vec{x}_{LS}} = -R^{-1}\delta_R \vec{x}_{LS}\,. \tag{F.11}$$

Again, utilizing equation (F.5), from equation (F.11), we can derive,

$$\left\|\vec{\delta}_{\vec{x}_{LS}}\right\| = \left\|- R^{-1}\delta_R \vec{x}_{LS}\right\| = \left\|R^{-1}\delta_R \vec{x}_{LS}\right\| \le \left\|R^{-1}\right\|\left\|\delta_R\right\|\left\|\vec{x}_{LS}\right\|. \tag{F.12}$$

Rearranging terms in equation (F.12), we obtain equation (2.40),

$$\frac{\left\|\vec{\delta}_{\vec{x}_{LS}}\right\|}{\left\|\vec{x}_{LS}\right\|} \le \left\|R\right\|\left\|R^{-1}\right\|\frac{\left\|\delta_R\right\|}{\left\|R\right\|} = CN(R)\frac{\left\|\delta_R\right\|}{\left\|R\right\|}. \tag{F.13}$$

In equations (F.8) and (F.13), $CN(R) = \left\|R\right\|\left\|R^{-1}\right\|$, as defined in equation (2.31), is referred to as the condition number of matrix R.

BIBLIOGRAPHY

[1] R.A. Schowengerdt, *Remote sensing: Models and methods for image processing*, 2nd ed., San Diego: Academic Press, 1997.

[2] J.A. Richards and X. Jia, *Remote sensing digital image analysis: An introduction*, 3rd ed., New York: Springer, 1999.

[3] J. Li, "Fast algorithms for wavelet-based analysis of hyperspectral signatures," the Master's thesis, University of Nevada Las Vegas, 1999.

[4] L.M. Bruce and J. Li, "Wavelets for computationally efficient hyperspectral derivative analysis," *IEEE Trans. Geosci. Remote Sensing*, vol.39, no.7, pp.1540-1546, July 2001.

[5] L.M. Bruce, C. Morgan and S. Larsen, "Automated detection of subpixel targets with continuous and discrete wavelet transforms," *IEEE Trans. Geosci. Remote Sensing*, vol.39, no.10, pp.2217-2226, October 2001.

[6] L.M. Bruce, J. Li and Y. Huang, "Automated detection of subpixel hyperspectral targets with adaptive multichannel discrete wavelet transform," *IEEE Trans. Geosci. Remote Sensing*, vol.40, no.4, pp.977-980, April 2002.

[7] J.F. Mustard, L. Li and G. He, "Nonlinear spectral mixture modeling of lunar multispectral data: Implications for lateral transport," *J. Geophys. Res.*, vol.103, no.E8, pp.19419-19425, August 1998.

[8] J.F. Mustard and C.M. Pieters, "Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra," *J. Geophys. Res.*, vol.94, no.B10, pp.13619-13634, October 1989.

[9] C.C. Borel and S.A. Gerstl, "Nonlinear spectral mixing models for vegetative and soil surface," *Remote Sensing Environ.*, vol.47, pp.403-416, 1994.

[10] B. Hapke, "Bidirectional reflectance spectroscopy 1. Theory," *J. Geophys. Res.*, vol.86, no.B4, pp.3039-3054, April 1981.

[11] C.C. Borel, S.A. Gerstl and B.J. Powers, "The radiosity method is optical remote sensing of structured 3-d surfaces," *Remote Sensing Environ.*, vol.36, pp.13-44, 1991.

[12] J.B. Adams, M.O. Smith and P.E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site," *J. Geophys. Res.*, vol.91, no.B8, pp.8098-8112, July 1986.

[13] A.M. Cross, J.J. Settle, N.A. Drake and R.T. Paivinen, "Subpixel measurement of tropical forest cover using AVHRR data," *Int. J. Remote Sensing*, vol.12, no.5, pp.1119-1129, 1991.

[14] P. Gong, J.R. Miller and M. Spanner, "Forest canopy closure from classification and spectral unmixing of scene components - multisensor evaluation of an open canopy," *IEEE Trans. Geosci. Remote Sensing*, vol.32, no.5, pp.1067-1080, September 1994.

[15] C.A. Hlavka and M.A. Spanner, "Unmixing AVHRR imagery to assess clearcuts and forest regrowth in Oregon," *IEEE Trans. Geosci. Remote Sensing*, vol.33, pp.788-795, May 1995.

[16] M.O. Smith, S.L. Ustin, J.B. Adams and A.R. Gillespie, "Vegetation in deserts: I. A regional measure of abundance from multispectral images," *Remote Sensing Environ.*, vol.31, pp.1-26, 1990.

[17] N.A. Quarmby, J.R. Townshend, J.J. Settle, K.H. White, M. Milnes, T.L. Hindle and N. Silleos, "Linear mixture modelling applied to AVHRR data for crop area estimation," *Int. J. Remote Sensing*, vol.13, no.3, pp.415-425, 1992.

[18] F.J. Garcia-Haro, M.A. Gilabert and J. Melia, "Linear spectral mixture modeling to estimate vegetation amount from optical spectral data," *Int. J. Remote Sensing*, vol.17, no.17, pp.3373-3400, 1996.

[19] G. Strang, *Linear algebra and its applications*, 2nd ed., New York: Academic Press, 1980.

[20] D.C. Montgomery and E.A. Peck, *Introduction to linear regression analysis*, New York: John Wiley & Sons, 1982.

[21] D.G. Manolakis, V.K. Ingle and S.M. Kogon, *Statistical and adaptive signal processing: Spectral estimation, signal modeling, adaptive filtering and array processing*, New York: McGraw-Hill, 2000.

[22] Y.E. Shimabukuru and J.A. Smith, "The least squares mixing models to generate fraction images derived from remote sensing multispectral data," *IEEE Trans. Geosci. Remote Sensing*, vol.29, no.1, pp.16-20, January 1991.

[23] J.J. Settle and N.A. Drake, "Linear mixing and the estimation of ground cover proportion," *Int. J. Remote Sensing*, vol.14, no.6, pp.1159-1177, 1993.

[24] Y.H. Hu, H.B. Lee and F.L. Scarpace, "Optimal linear spectral unmixing," *IEEE Trans. Geosci. Remote Sensing*, vol.37, no.1, pp.639-644, January 1999.

[25] M. Brown, H.G. Lewis and S.R. Gunn, "Linear spectral mixture models and support vector machines for remote sensing," *IEEE Trans. Geosci. Remote Sensing*, vol.38, no.5, pp.2346-2360, September 2000.

[26] S. Haykin, *Adaptive Filter Theory*, 2nd ed., New Jersey: Prentice Hall, 1991.

[27] J.C. Harsanyi and C.I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sensing*, vol.32, no.4, pp.779-785, July 1994.

[28] J. T.M. Tu, C.H. Chen and C.I. Chang, "A posteriori least squares orthogonal subspace projection approach to desired signature extraction and detection," *IEEE Trans. Geosci. Remote Sensing*, vol.35, no.1, pp.127-139, January 1997.

[29] C.I. Chang, X.L. Zhao, M. Althouse and J.J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol.36, no.3, pp.898-912, May 1998.

[30] P.J. Rousseeuw and A.M. Leroy, *Robust regression and outlier detection*, New York: John Wiley & Sons, 1987.

[31] P.L. Rosin, "Robust pixel unmixing," *IEEE Trans. Geosci. Remote Sensing*, vol.39, no.9, pp.1978-1983, September 2001.

[32] J.T. Kent and K.V. Mardia, "Spectral classification using fuzzy membership models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.10, pp.659-671, 1986.

[33] P. Bosdogianni, M. Petrou and J. Kittler, "Mixture models with higher order moments," *IEEE Trans. Geosci. Remote Sensing*, vol.35, no.2, pp.341-353, March 1997.

[34] X. Jia and J.A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sensing*, vol.37, no.1, pp.538-542, January 1999.

[35] L.O. Jimenez and D.A. Landgrebe, “Hyperspectral data analysis and supervised feature reduction via projection pursuit,” *IEEE Trans. Geosci. Remote Sensing*, vol.37, no.6, pp.2653-2667, November 1999.

[36] Y. Huang, L.M. Bruce, J. Li, C. Leon and D. Shaw, “Brushlet transforms for hyperspectral feature extraction in automated detection of nutsedge presence in soybean,” *Proc. IEEE IGARSS*, vol.1, pp.527-529, 2001.

[37] S. Mallat, “A theory for multi-resolution signal decomposition: The wavelet representation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol.11, pp.674-693, 1989.

[38] F.G. Meyer and R.R. Coifman, “Brushlets: a tool for directional image analysis and image compression,” *Applied and Computational Harmonic Analysis*, vol.4, pp.147-187, 1997.

[39] J. Li, L.M. Bruce, J. Byrd and J. Barnett, “Automated detection of Pueraria montana (kudzu) through Haar analysis of hyperspectral reflectance data,” *Proc. IEEE IGARSS*, vol.5, pp.2247-2249, 2001.

[40] Y. Huang, L.M. Bruce, J. Byrd and B. Mask, “Using wavelet transforms of hyperspectral reflectance curves for automated monitoring of Imperata cylindrica (cogongrass),” *Proc. IEEE IGARSS*, vol.5, pp.2244-2246, 2001.

[41] Y. Huang, L.M. Bruce, T. Koger and D. Shaw, “Analysis of the effects of cover crop residue on hyperspectral reflectance discrimination of soybean and weeds via Haar transform,” *Proc. IEEE IGARSS*, vol.3, pp.1276-1278, 2001.

[42] Z. Zeng and I.G. Cumming, “SAR image data compression using a tree-structured wavelet transform,” *IEEE Trans. Geosci. Remote Sensing*, vol.39, no.3, pp.546-552, March 2001.

[43] P.L. Dragotti, G. Poggi and A. Ragozini, “Compression of multispectral images by three-dimensional SPIHT algorithm,” *IEEE Trans. Geosci. Remote Sensing*, vol.38, no.1, pp.416-428, January 2000.

[44] M. Petrou, P. Hou, S. Kamata and C. Underwood, "Region-based image coding with multiple algorithms," *IEEE Trans. Geosci. Remote Sensing*, vol.39, no.3, pp.562-570, March 2001.

[45] B. Garguet-Duport, J. Girel, J.M. Chassery and G. Pautou, “The use of multiresolution analysis and wavelets transform for merging of SPOT panchromatic and multispectral image data,” *Photogramm. Eng. Remote Sensing*, vol.62, no.9, pp.1057-1066, 1996.

[46] J. Zhou, D.L. Civco and J.A. Silander, “A wavelet transform method to merge Landsat TM and SPOT panchromatic data,” *Int. J. Remote Sensing*, vol.19, no.4, pp.743-757, 1998.

[47] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala and R. “Multiresolution-based image fusion with additive wavelet decomposition,” *IEEE Trans. Geosci. Remote Sensing*, vol.37, no.3, pp.1204-1211, May 1999.

[48] Y. Du, B. Guindon and J. Cihlar, "Haze detection and removal in high resolution satellite image with wavelet analysis," *IEEE Trans. Geosci. Remote Sensing*, vol.40, no.1, pp.210-217, January 2002.

[49] K.K. Simhadri, S.S. Iyengar, R.J. Holyer, M. Lybanon and J.M. Zachary, “Wavelet-based feature extraction from oceanographic images,” *IEEE Trans. Geosci. Remote Sensing*, vol.36, no.3, pp.767-778, May 1998.

[50] S. Fukuda and H. Hirosawa, “A wavelet-based texture feature set applied to classification of multifrequency polarimetric SAR images,” *IEEE Trans. Geosci. Remote Sensing*, vol.37, no.5, pp.2282-2286, September 1999.

[51] A. Niedermeier, E. Romaneeben and S. Lehner, “Detection of coastlines in SAR images using wavelet methods,” *IEEE Trans. Geosci. Remote Sensing*, vol.38, no.5, pp.2270-2281, September 2000.

[52] A. Webb, *Statistical Pattern Recognition*, London: Arnold, 1999.

[53] J. Hanley and B. McNeil, “The meaning and use of the area under a receiver operating characteristics (ROC) curves,” *Diagnostic Radiology*, vol.143, no.1, pp.29-36, 1982.

[54] M.O. Smith, P.E. Johnson and J.B. Adams, “Quantitative determination of minearal types and abundances from reflectance spectra using principal component analysis,” *Proc. 15th Lunar and Planetary Sci. Conf., Part 2, J. Geophys. Res.*, vol.90, suppl., pp.C797-C804, February 1985.

[55] A.R. Huete, “Separation of soil-plant spectral mixtures by factor analysis,” *Remote Sensing Environ.*, vol.19, pp.237-251, 1986.

[56] T. Bajjouk, J. Populus and B. Guillaumont, “Quantification of subpixel cover fractions using principal component analysis and a linear programming method: Application to the coastal zone of roscoff (France),” *Remote Sensing Environ.*, vol.64, pp.153-165, 1998.

[57] N. Keshava and J.F. Mustard, "Spectral unmixing," IEEE Siganl Processing Magazine, vol.19, no.1, pp.44-57, January 2002.

[58] C.S. Burrus, R.A. Gopinath and H.Guo, *Introduction to wavelets and wavelet transform: A primer*, New Jersey: Prentice Hall, 1998.

[59] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1999.

[60] W. Sweldens, *The lifting scheme: A construction of second generation wavelets*, Technical Report TR-1995-6, Math Dept., Univ. of South Carolina, May 1995.

[61] W. Sweldens, "The lifting scheme: A custom-design construction of biothogonal wavelets," *Appl. Computational Harmonic Anal.*, vol.3, no.2, pp.186-200, 1996.

[62] C.K. Chui, *An Introduction to Wavelets*, San Diego: Academic Press, 1992.

[63] K.R. Castleman, *Digital image processing*, Prentice Hall, 1996.

[64] M. Vetterli and C. Herley, "Wavelet and filer banks: theory and design," *IEEE Trans. Signal Processing*, vol.40, no.9, pp.2207-2232, 1992.

[65] G. Strang and T. Nguyen, *Wavelets and filter banks*, Wellesley-Cambridge Press, 1996.

[66] P. P. Vaidyanathan, "Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having the perfect-reconstruction property," *IEEE Trans. ASSP*, vol.35, no.4, pp.476-492, 1987.

[67] P. P. Vaidyanathan, *Multirate systems and filter banks*, Prentice-Hall, 1992.

[68] R. Turcajova and J. Kautsky, "Shift products and factorization of wavelet matrices," *Numercial Algorithms*, vol.8, pp.27-54, 1994.

[69] R.C. Gonzalez and R.E. Woods, *Digital image processing*, 2nd ed., New Jersey: Prentice Hall, 2002.

[70] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. on Pure and Appl. Math.*, vol.41, pp.909-996, 1988.

[71] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, PA, 1992.

[72] A.K. Jain, *Fundamentals of digital image processing*, Prentice Hall, 1989.

[73] G.K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol.34, pp.30-44, April 1991.

[74] R. Fletcher, *Practical methods of optimization*, 2nd Ed., New York: John Wiley & Sons, 2000.

[75] http://www.asdi.com/asdi_t2_pr_sp.html