Correspondence

Correction to "Wavelet-Based Feature Extraction for Improved Endmember Abundance Estimation in Linear Unmixing of Hyperspectral Signals"

In the above paper [1], the second author (Lori Mann Bruce) was omitted from the byline. The complete byline should read as "Jiang Li, *Member, IEEE*, and Lori Mann Bruce, *Senior Member, IEEE*."

References

[1] J. Li, "Wavelet-based feature extraction for improved endmember abundance estimation in linear unmixing of hyperspectral signals," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, pp. 644–649, Mar. 2004.

Lori Mann Bruce (S'91–M'96–SM'01) received the B.S.E. degree from the University of Alabama, Huntsville, in 1991, the M.S. degree from the Georgia Institute of Technology, Atlanta, in 1992, and the Ph.D. degree from the University of Alabama, in 1996, all in electrical engineering.

She has served as a member of the technical staff at the U.S. Army Strategic Defense Command, Washington, DC, and from 1996 to 2000, she was an Assistant Professor in the Electrical and Computer Engineering Department, Howard R. Hughes College of Engineering, University of Nevada, Las Vegas. Currently, she is an Assistant Professor in the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, where she is also affiliated with the Remote Sensing Technology Center. Her research interests include applying advanced digital signal processing techniques such as discrete wavelet transforms to automated pattern recognition in hyperspectral remote sensing and digital mammography.

Dr. Bruce is a member of Eta Kappa Nu, Phi Kappa Phi, and Tau Beta Pi.

Digital Object Identifier 10.1109/TGRS.2004.829663

Wavelet-Based Feature Extraction for Improved Endmember Abundance Estimation in Linear Unmixing of Hyperspectral Signals

Jiang Li, *Member, IEEE*

***Abstract*—This paper shows that the use of appropriate features, such as discrete wavelet transform (DWT)-based features, can improve the least squares estimation of endmember abundances using remotely sensed hyperspectral signals. On average, the abundance estimation deviation is reduced by 30% to 50% when using the DWT-based features, as compared to the use of original hyperspectral signals or conventional principal component analysis (PCA)-based features. Theoretical analyses further reveal that the increase of endmember separability is a fundamental reason leading to this improvement. In addition, the robustness of the DWT-based features is verified experimentally. Finally, the idea is generalized as a point that the remote sensing community needs to investigate feature extraction (or dimensionality reduction) methods that are based on signal classification, such as the DWT approach, for linear unmixing problems, rather than using feature extraction methods that are based on signal representation, such as the conventional PCA approach.**

***Index Terms*—Feature extraction, hyperspectral, linear unmixing, remote sensing, wavelet.**

I. Introduction

ONE OF THE important applications of the remote sensing technology is target detection and classification. It is typically assumed that a pixel in a remotely sensed image represents a distinct ground cover material and can be uniquely assigned to a ground cover class. However, this assumption cannot be guaranteed because of both the limited spatial resolution of remote sensors and the heterogeneous nature of target surfaces. Therefore, it becomes necessary to investigate the subpixel information when a pixel's corresponding spectrum is composed of a mixture of multiple materials. This leads to a research area of spectral unmixing, which is described as a quantitative analysis procedure to recognize constituent ground cover materials (or endmembers) and obtain their mixing proportions (or abundances) from a mixed pixel.

The spectral unmixing problem has been extensively investigated for the past two decades [1]–[9]. In general, mathematical models for spectral unmixing are divided into two broad categories: linear mixture models (LMMs) and nonlinear mixture models (NLMMs). The NLMM typically has a relatively more accurate simulation of physical phenomena and, thus, could result in a better unmixing performance for certain applications [8], [9]. However, the NLMM model is usually complicated and application dependent, which 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 [1], the forest studies [2], [3], and the vegetation studies [4], [5]. Based on the LMM, many mature techniques, such as least squares estimation (LSE) [6], [7], can be applied to the spectral unmixing problem. In fact, the LSE has been commonly accepted to solve the generalized linear spectral unmixing problem [1]–[7].

One requirement for 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 [6], [7]. To a certain extent, this condition limits the use of multispectral data. While this limitation has less effect on the use of hyperspectral data, it is questionable that simply using all hyperspectral bands for linear unmixing can lead to a satisfactory result. Moreover, the use of hyperspectal data causes another problem called the Hughes phenomena [10], which requires more training data for a supervised classification system to obtain accurate results. It has been realized that the use of feature extraction (or dimensionality reduction) can avoid the Hughes phenomena and improve the classification performance [10]–[14]. Naturally, the question is whether the linear unmixing performance can be improved by the use of appropriate features. Answering this question becomes a major concern of this paper. Among various feature extraction techniques, discrete wavelet transform (DWT)-based method has been our research focus. Thus, we are interested in knowing whether the feature extraction based on the DWT can improve the linear unmixing performance, particularly the LSE of endmember abundances using hyperspectral signals.

Feature extraction approaches based on principal component analysis (PCA) and discrete cosine transform (DCT) are also investigated for the purpose of comparison. A conventional way to extract PCA- or DCT-based features is to use the first several large-amplitude PCA or DCT coefficients. Experimental results in this paper show that while this conventional feature extraction method reduces the dimensionality of hyperspectral data, it does not help improve the abundance estimation. Note that this conventional approach is based on techniques that provide superior energy compaction. 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

Manuscript received May 19, 2003; revised November 19, 2003.

The author is with Strategic Technology Development, Baker Hughes, Inc., Houston, TX 77073 USA (e-mail: jiang.li@bakerhughes.com).

Digital Object Identifier 10.1109/TGRS.2003.822750

coefficients may not be adequate. For this reason, alternative approaches to feature extraction based on PCA and DCT are investigated in this paper. For the same reason, the DWT-based approach is proposed for the linear unmixing of hyperspectral signals.

II. Error Analysis of Abundance Estimation

The LSE is a mathematical technique that is used to implement an optimum estimation of parameters based on certain known information. The estimation is regarded as optimum in the sense of minimizing the total energy of estimation errors. In the context of linear spectral unmixing, the LSE is implemented under an assumption of the LMM to obtain an optimum abundance estimation, given the information of mixed pixels and constituent pure pixels. In general, the LMM is described using a linear equation

$$\vec{y} = A\vec{x} + \vec{e} \tag{1}$$

where $\vec{y} = [y_1, y_2, \ldots, y_N]^T$ is a mixed-pixel spectrum; $\vec{x} = [x_1, x_2, \ldots, x_M]^T$ contains true abundances of endmembers; $\vec{e} = [e_1, e_2, \ldots, e_N]^T$ is random measurement error; and $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_M]$ contains endmember spectra, with $\vec{a}_i = [a_{i1}, a_{i2}, \ldots, a_{iN}]^T$ for $i = 1, 2, \ldots, M$. The superscript T stands for a vector transpose; N is the number of spectral bands; and M is the number of endmembers. Using the LSE method, an optimum estimate, $\vec{x}_{LS}$, of endmember abundances can be derived from (1)

$$\vec{x}_{LS} = (A^T A)^{-1} A^T \vec{y} \tag{2}$$

provided that $(A^T A)^{-1}$ exists. It can be shown that if the endmember spectra are linearly independent, then $(A^T A)^{-1}$ always exists [15].

To quantitatively evaluate the effect of the feature extraction on the endmember abundance estimation, a criterion of mean square error (MSE), Γ_x, is first defined

$$\Gamma_x = E\left[\frac{\vec{e}_x^T \vec{e}_x}{M}\right] \tag{3}$$

where $\vec{e}_x = \vec{x}_{LS} - \vec{x}$ is defined as an error of abundance estimate, and $E[\cdot]$ calculates the expectation value of a random variable. Note that in this definition the average value of the total error energy is utilized. Its advantage is that the MSE is no longer scaled by the number of endmembers. Thus, it allows a fair comparison of the MSE results from experiments with various numbers of endmembers. According to the definition of the trace of a matrix, (3) can be further derived as,

$$\Gamma_x = \frac{1}{M} E[\mathrm{Tr}[\vec{e}_x^T \vec{e}_x]] \tag{4}$$

where $\mathrm{Tr}[\cdot]$ represents a matrix trace operation. Substituting (1) and (2) into (4) and utilizing the property of the matrix trace operation that given any two matrices, U and V, there exists $\mathrm{Tr}[UV] = \mathrm{Tr}[VU]$, Γ_x can be further derived as

$$\Gamma_x = \frac{1}{M} \mathrm{Tr}[(A^+)^T A^+ R] \tag{5}$$

where $A^+ = (A^T A)^{-1} A^T$ is defined as a pseudoinverse of matrix A, and $R = E[\vec{e}\vec{e}^T]$ is the autocorrelation matrix of random measurement errors. Suppose that $\vec{e}$ have the following two statistical characteristics:

- *Assumption 1:* Random variables, e_j, in $\vec{e}$ have zero means, for $j = 1, 2, \ldots, N$.
- *Assumption 2:* e_j are uncorrelated and have different variances σ_j^2.

Then, (5) can be further derived as

$$\Gamma_x = \frac{1}{M} \sum_{j=1}^{N} p_j \sigma_j^2 \tag{6}$$

where p_j are the main diagonal elements in the symmetric matrix $(A^+)^T A^+$.

To solve (1) for abundances, the endmember spectra 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 mixed pixels. These differences could be caused by many factors such as various acquisition conditions and inaccurate library endmember spectra. However, it is reasonable to make the following generalized assumption:

Assumption 3: Random measurement errors in LMM stem from differences between the library endmember spectra and the true endmember spectra constituting mixed pixels.

Then, (1) can be rewritten as

$$y = A\vec{x} + \Delta_A \vec{x} \tag{7}$$

where $\Delta_A = [\vec{\delta}_1, \vec{\delta}_2, \ldots, \vec{\delta}_M]$, with $\vec{\delta}_i = [\delta_{i1}, \delta_{i2}, \ldots, \delta_{iN}]^T$, represents the differences defined in *Assumption 3*. Comparing (1) and (7), the random measurement errors, $\vec{e}$, can be expressed as

$$\vec{e} = \Delta_A \vec{x}. \tag{8}$$

Furthermore, suppose that $\vec{\delta}_i$ have the following three statistical characteristics.

- *Assumption 4:* $\vec{\delta}_i$ are linearly independent.
- *Assumption 5:* Random variables, δ_{ij}, in $\vec{\delta}_i$ have zero means.
- *Assumption 6:* δ_{ij} (for a fixed i) are uncorrelated, and have different variances σ_{ij}^2.

Then, σ_j^2 in (6) can be further derived as

$$\sigma_j^2 = \sum_{i=1}^{M} \sigma_{ij}^2 x_i^2. \tag{9}$$

Substituting (9) into (6), Γ_x can be further computed as

$$\Gamma_x = \frac{1}{M} \sum_{j=1}^{N} \sum_{i=1}^{M} p_j \sigma_{ij}^2 x_i^2. \tag{10}$$

Note that Γ_x is only affected by two variables, p_j and σ_{ij}^2. The variable p_j is uniquely determined by the endmember spectra. The variable σ_{ij}^2 indicates the variances of endmember spectra, which are referred to as the within-endmember variances in this paper.

A special case of *Assumption 2* is as follows.

Assumption 7: e_j are uncorrelated and have the same variance σ^2.

Then, (5) can be reduced to

$$\Gamma_{xs} = \frac{1}{M}\mathrm{Tr}\left[R_A^{-1}\right]\sigma^2 \tag{11}$$

where Γ_{xs} denotes a simplified version of Γ_x, and $R_A = A^T A$ is an autocorrelation matrix of endmember spectra. A special case of *Assumption 6* is as follows.

Assumption 8: δ_{ij} (for a fixed i) are uncorrelated, and have the same variances σ_i^2.

Then, (9) can be reduced to

$$\sigma^2 = \sum_{i=1}^{M} \sigma_i^2 x_i^2. \tag{12}$$

Substituting (12) into (11), Γ_{xs} can be further derived as

$$\Gamma_{xs} = \frac{1}{M}\mathrm{Tr}\left[R_A^{-1}\right]\sum_{i=1}^{M} \sigma_i^2 x_i^2. \tag{13}$$

Note that Γ_{xs} is represented by two separated terms multiplied by a constant factor $1/M$. One term is a weighted sum of within-endmember variances σ_i^2. The other term $\mathrm{Tr}[R_A^{-1}]$ is affected by separabilities (or dissimilarities) of endmember spectra, which are referred to as the between-endmember variances in this paper. Experimentally it can be shown that increasing the between-endmember variances can reduce the value of $\mathrm{Tr}[R_A^{-1}]$. Even though this trend is not a simple monotonic relationship, it does provide a possibility to reduce the value of $\mathrm{Tr}[R_A^{-1}]$ via adjusting the endmember spectra, e.g., via feature extraction. It has been proven possible to decrease the within-class (or within-endmember) variances and increase the between-class (or between-endmember) variances by using DWT-based features [12], [13]. According to (13), therefore, it is possible to reduce the MSE of endmember abundance estimation.

III. Feature Extraction for Abundance Estimation

A. Feature Extraction Based on DWT

Wavelet transform of a signal is defined as an inner product of the signal and wavelet bases. The fine-scale and large-scale information of hyperspectral signals can be simultaneously investigated by projecting signals onto a set of wavelet bases with various scales, which is referred to as a multiresolution analysis (MRA) of signals [16]. Extracting appropriate features from the wavelet-based MRA information provides a potential to effectively discriminate hyperspectral signals in feature space. The idea has been verified in our previous research for target detection and classification problems using hyperspectral signals [12], [13].

In practice, a fast wavelet transform (FWT) algorithm exists for a computationally efficient implementation of the DWT. The basic idea is to represent the mother wavelet as a set of high-pass and lowpass filters (G and H) in a filter bank. Then the signal is passed through the filter bank. Following the filtering, the signal is decimated by a factor of two for a dyadic DWT. The outputs of the high- and lowpass branches at scale s are called wavelet detail and approximation coefficients (D_s and C_s), respectively. This filtering process followed by decimation is referred to as single-step wavelet decomposition, as shown in Fig. 1. 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.

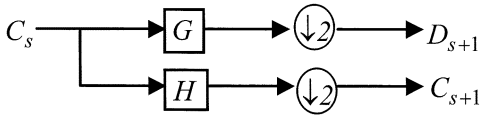


Fig. 1. Iterative implementation of the dyadic FWT.

A large variety of features could be extracted from the DWT results, such as the energy of DWT coefficients, the DWT coefficients themselves, and any combination of the coefficients. Considering the linear unmixing problem, it is important to ensure the LMM remains after the DWT and feature extraction. Otherwise, the linear unmixing problem will no longer exist. For example, it has been experimentally shown that the use of the nonlinear wavelet energy features actually reduces the abundance estimation performance [15], though this result does not exclude the possibility that nonlinear wavelet energy features might be helpful in the case of the NLMM. Since the DWT is a linear transform, the DWT coefficients at specific scales could be directly used as linear features. That is, a linear DWT feature can be formed using the all detail coefficients D_s or approximation coefficients C_s, for a fixed s.

These 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 these DWT scalar subset features provides a potential for improving the LSE of endmember abundances. Note that the dimensionality of these linear wavelet features, which is equal to the number of wavelet detail or approximation coefficients at a specific scale, decreases with the increase of wavelet decomposition scale. Theoretically, the number of elements in D_{s+1} or C_{s+1} is half of the number of elements in D_s or C_s. Thus, the use of these linear wavelet features is also associated with a dimensionality reduction of hyperspectral signals.

B. DWT-Based Linear Unmixing System

Based on the DWT-based features extraction, a linear unmixing system is designed for the endmember abundance estimation of a mixed-pixel spectrum. 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 preprocessing 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 LSE (CLSE) technique using a quadratic programming (QP) algorithm [17] is performed to implement the abundance estimation. The reason for using the CLSE is that in the practical implementation of the LSE of endmember abundances, two constraints are applied to make the

estimated abundances physically meaningful. One is nonnegativity, i.e., physically abundances should not be negative numbers. The other 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.

The proposed system is a supervised linear unmixing system. First, the system needs to be trained to determine DWT-based optimal feature sets (OFS). 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. Note that the optimum is based on the given training data. For different sets of training data, the optimal results may be different. However this is the essential idea of the supervised system. Second, the OFS determined in the system training phase need to be tested to investigate how well they can work in practical applications. A system testing procedure is performed on a set of testing data, which is mutually exclusive from the training data set. This avoids the bias introduced by the use of any information from the training data. Finally, the testing results are quantitatively and/or qualitatively evaluated to investigate the estimation system performance. In this paper, a root mean square error (RMSE) of the endmember abundance estimation is defined for a quantitative evaluation. Given a set of testing data consisting of K mixed-pixel spectra, the RMSE, Ω, of abundance estimations for all the mixed-pixel spectra in the set is computed as

$$\Omega = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\Theta_k} \tag{14}$$

where Θ_k represents the average error energy of the abundance estimation corresponding to the kth mixed pixel in the set. In general, Ω indicates an average deviation of the abundance estimate from the true abundance. The RMSE is also utilized as a criterion to determine OFS during the system training phase. That is, among the investigated DWT-based feature sets, the one resulting in the smallest RMSE is regarded as the OFS.

C. Feature Extraction Based on PCA and DCT

The preprocessing is the first and core step of the proposed DWT-based linear unmixing system. However, the DWT followed by the feature extraction is not the only choice for the preprocessing. As a comparison, two other preprocessing methods based on the PCA and DCT are investigated. Since both the PCA and the DCT are linear transform, their transform coefficients are directly used as features to retain the validity of the LMM after feature extraction.

Conventionally, the PCA- or DCT-based features are formed using the first several large-amplitude PCA or DCT coefficients, since most of the energy (or information) of the original signal concentrates in these coefficients. For example, given a sequence of PCA or DCT coefficients $Z(n)$, the first feature set would be $[Z(0)]^T$, the second one would be $[Z(0), Z(1)]^T$, and so on. Since this conventional method is not appropriate for the endmember abundance estimation as discussed in Section I, an alternative approach, called the sliding-window method, is proposed for selecting a subset of PCA or DCT coefficients. This alternative approach utilizes a sliding window of size P to select coefficient subsets. That is, given a shifting stepsize q, the first feature set would be $[Z(0), \ldots, Z(P)]^T$, and the second one would be $[Z(q), \ldots, Z(q+P)]^T$, etc. For this paper, the window sizes of $P = 3, 5, 10,$ and 20 and the shifting stepsize of $q = 1$ are investigated.

TABLE I
OFS AND THE RMSE OF ENDMEMBER ABUNDANCE ESTIMATION

	2-endmember		3-endmember			2-endmember		3-endmember	
	OFS	RMSE	OFS	RMSE		OFS	RMSE	OFS	RMSE
Haar	D7	0.0531	D7	0.1035	Bior1.3	D7	0.0551	D2	0.1074
Db3	D6	0.1177	D6	0.1020	Bior2.2	D6	0.1274	D6	0.1732
Db6	D6	0.0978	D6	0.1748	Bior3.5	D5	0.0557	D3	0.1080
Db8	D4	0.0718	D5	0.1702	Bior4.4	D6	0.1192	D7	0.1180
Sym3	D6	0.1177	D8	0.1020	C-PCA	{9}	0.1045	{19}	0.1423
Sym6	D6	0.0742	D6	0.1681	A-PCA	{10, 0}	0.1045	{20, 0}	0.1423
Coif1	D6	0.1107	D7	0.1101	C-DCT	{19}	0.1047	{20}	0.1424
Coif2	D6	0.1107	D6	0.1683	A-DCT	{5, 63}	0.0528	{20, 53}	0.1046
ORG	/	0.1045	/	0.1423	/	/	/	/	/

IV. EXPERIMENTS AND RESULTS

In general, experiments of the linear unmixing can be divided into two broad categories: two-endmember and multiendmember linear unmixing. This paper takes an agriculture application as an example. 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. For the two-endmember case, the unmixing of soybean and soil is investigated. For the multiendmember case, the unmixing of soybean, grass and soil is investigated.

Hyperspectral reflectance spectra are measured using a handheld spectroradiometer of Analytical Spectral Devices (ASD). The ASD's instrument has an ability to measure the electromagnetic radiance (and consequently derive the reflectance) ranging from 350–2500 nm with an average 10-nm bandwidth for each spectral band. In this paper, part of the spectrum ranging from 354–1753 nm (1400 samples) is investigated. 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 U.S. Department of Agriculture's Agricultural Research Service near Stoneville, MS, in June, July, and August 2000. The measurements were calibrated using a white reference panel every few minutes to ensure the accuracy of measurement results. 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.

Hyperspectral reflectance spectra from the handheld 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. Using these pure-pixel spectra, mixed-pixel spectra with known abundances can be synthesized. For this paper, the values of true abundance vary from 0.0 to 1.0 with an increment of 0.1. The endmember spectra are synthesized through averaging all pure-pixel spectra in the training data set just for simplicity.

In addition, the mixed-pixel spectra with various noise levels are synthesized to simulate the case where the atmospheric in-

TABLE II
OFS AND THE RMSE OF ENDMEMBER ABUNDANCE ESTIMATION WHEN MIXED-PIXEL SPECTRA INCLUDE VARIOUS-LEVEL NOISES

	2-endmember						3-endmember					
	SNR = 10		SNR = 100		SNR = 1000		SNR = 10		SNR = 100		SNR = 1000	
	OFS	RMSE	OFS	RMSE	OFS	RMSE	OFS	RMSE	OFS	RMSE	OFS	RMSE
Haar	D8	0.0781	D7	0.0574	D7	0.0530	D9	0.1371	D9	0.1186	D7	0.1050
Db3	D7	0.0802	D8	0.0667	D6	0.1237	A1	0.1449	D8	0.1154	D8	0.1031
Db6	A1	0.1039	D6	0.1081	D6	0.1008	A1	0.1449	A1	0.1424	D6	0.1820
Db8	A1	0.1039	A1	0.1047	D6	0.0842	A1	0.1449	A1	0.1424	D5	0.1344
Sym3	D7	0.0802	D8	0.0667	D6	0.1237	A1	0.1449	D8	0.1154	D8	0.1031
Sym6	A1	0.1040	D6	0.1252	D6	0.0904	A2	0.1449	A6	0.1418	D6	0.1746
Coif1	D7	0.0790	D8	0.0647	D6	0.1125	A1	0.1448	D7	0.1204	D7	0.1113
Coif2	A1	0.1040	A1	0.1047	D6	0.1244	A1	0.1448	A5	0.1418	D6	0.1765
Bior1.3	D8	0.0774	D7	0.0585	D7	0.0552	A1	0.1448	D7	0.1199	D7	0.1065
Bior2.2	D7	0.0908	D8	0.0712	D6	0.1335	A1	0.1448	D8	0.1168	D8	0.1034
Bior3.5	A1	0.1040	A1	0.1047	D6	0.0829	A2	0.1450	A6	0.1411	A6	0.1395
Bior4.4	D7	0.0837	D7	0.0677	D6	0.1283	A1	0.1448	D7	0.1314	D7	0.1195

fluence exists. An additive white Gaussian noise with zero mean and α^2 variance is investigated. The SNR of the synthesized spectra is calculated as the ratio of the average energy of original spectra to α^2. For this paper, SNR $= 10, 100,$ and 1000 are investigated to simulate the high, middle, and low noise levels. Besides, to simulate a practical situation where the available endmember spectra are not pure, mixed endmember spectra are synthesized. The mixed ratios (MRs) investigated are $9:1$ and $8:2$ for the two-endmember case, and $9:0.5:0.5$ and $8:1:1$ for the three-endmember case. For example, the MR $= 9:1$ means that the mixed spectrum of soybean consists of 90% and 10% pure spectra of soybean and soil, respectively; and the mixed spectrum of soil consists of 90% and 10% pure spectra of soil and soybean, respectively. These two additional experiments are designed to test the robustness of the proposed DWT-based features.

Table I summarizes the testing RMSE results of endmember abundance estimation. The wavelet types investigated include Haar, Daubechies 3 (Db3), Db6, Db8, Symlet 3 (Sym3), Sym6, Coiflet 1 (Coif1), Coif 2, Biorthogonal 1.3 (Bior1.3), Bior 2.2, Bior 3.5, and Bior4.4 wavelets [18]. The OFS are obtained from the system training. For the DWT-based features, notation D7 represents the OFS is from the wavelet detail coefficients at scale 7. For the conventional PCA- or DCT-based features (C-PCA or C-DCT), notation {19} represents the OFS is from the first 19 PCA or DCT coefficients. For the alternative (or sliding window) PCA- or DCT-based features (A-PCA or A-DCT), notation {20, 53} represents the OFS is obtained when the window size is 20 and the window shift is 53.

The results in Table I show that for the two-endmember case the DWT-based OFS from the Haar wavelet produces the best RMSE result among the 12 investigated wavelets. It leads to a $\sim 50\%$ improvement on the RMSE compared to the use of original hyperspectral signals (ORG). While the OFS from C-PCA, A-PCA, and C-DCT do not lead to the RMSE improvement, the OFS from A-DCT does. Similar results can be observed for the three-endmember case, where an $\sim 30\%$ RMSE improvement is obtained when using the DWT-based OFS from the Haar, Db3, and Sym3 wavelets, or using the OFS from A-DCT. The results in Table I also show that, in general, there is not a single wavelet performing well for all experiment cases. However, it is interesting to note that the DWT-based features from the Haar and Bior1.3 wavelets perform well in both investigated experiment cases. This can also be observed in the following testing results of the robustness of the DWT-based features.

TABLE III
OFS AND THE RMSE OF ENDMEMBER ABUNDANCE ESTIMATION WHEN ENDMEMBER SPECTRA ARE NOT PURE

	2-endmember				3-endmember			
	MR = 9:1		MR = 8:2		MR = 9:0.5:0.5		MR = 8:1:1	
	OFS	RMSE	OFS	RMSE	OFS	RMSE	OFS	RMSE
Haar	D7	0.0823	D7	0.1265	D5	0.1215	D5	0.1379
Db3	D6	0.1381	D6	0.1693	D8	0.1220	D8	0.1553
Db6	D6	0.1202	D6	0.1545	D5	0.1317	D5	0.1507
Db8	D4	0.0978	D4	0.1380	D5	0.1893	D5	0.2145
Sym3	D6	0.1381	D6	0.1693	D8	0.1220	D8	0.1553
Sym6	D6	0.1021	D6	0.1426	D6	0.1817	D6	0.2005
Coif1	D6	0.1313	D6	0.1642	D3	0.1226	D3	0.1429
Coif2	D6	0.1325	D6	0.1663	D6	0.1827	D6	0.2049
Bior1.3	D7	0.0824	D7	0.1261	D5	0.1306	D5	0.1484
Bior2.2	D6	0.1479	D6	0.1773	D6	0.1932	D3	0.1380
Bior3.5	D5	0.0718	D5	0.1149	D3	0.1141	D3	0.1285
Bior4.4	D6	0.1404	D6	0.1717	D7	0.1321	D7	0.1590

Table II shows the testing RMSE results of endmember abundance estimation when the mixed-pixel spectra include various-level noises, where notation A6 represents the OFS is from the wavelet approximation coefficient at scale 6. In general, the DWT-based features from the Haar and Bior1.3 wavelets perform consistently, while the DWT-based features from other investigated wavelets do not. This consistence appears in two aspects: 1) the Haar and Bior1.3-based features perform well in both experiment cases; and 2) as the noise level increases, the performance degrades, but is still satisfactory even at the high noise level (SNR $= 10$). This consistence also shows the robustness of the DWT-based features from certain wavelets, particularly the Haar and Bior1.3 wavelets in this paper.

Table III shows the testing RMSE results of endmember abundance estimation when the endmember spectra are not pure. In general, the RMSE results are worse than the case when the endmember spectra are pure. However, note that the DWT-based features from certain wavelets, particularly Haar and Bior1.3, consistently perform well. This further verifies the robustness of the proposed DWT-based features.

V. CONCLUSION

This paper both experimentally and theoretically shows that the use of appropriate features, such as DWT-based and nontraditional DCT-based features, can improve the LSE of endmember abundances using remotely sensed hyperspectral signals. Experimental results show that the use of DWT-based features reduces the abundance estimation deviation by 30% to 50% on average, as compared to the use of original hyperspectral signals or conventional PCA- and DCT-based

features. Furthermore, experimental results from two special cases, where mixed-pixel spectra include various-level noises and endmember spectra are not pure, verify the robustness of the proposed DWT-based features. Theoretical analyses reveal that a fundamental reason leading to this improvement of endmember abundance estimation stems from the increase of endmember separability after the use of appropriate features. This also provides a generic criterion to design a feature extraction technique for improving the endmember abundance estimation. In conclusion, the remote sensing community needs to investigate feature extraction (or dimensionality reduction) methods that are based on signal classification, such as the proposed DWT approach, for linear unmixing problems, rather than using feature extraction methods that are based on signal representation, such as the conventional PCA and DCT approaches.

In the future, it will be interesting to investigate the multichannel DWT (MDWT) and the wavelet packet (WP) techniques [19] for the feature extraction. The MDWT allows a more complete MRA of signals, and feature sets could be obtained from wavelet coefficients at various scales and filtering channels. The WP provides an iterative decomposition of both approximation and detail signals, and feature sets could be obtained from optimal WP trees. Second, it will be interesting to investigate the applications of the proposed DWT-based linear unmixing system to hyperspectral images, where both spectral and spatial information can be utilized for the feature extraction. Third, it will be interesting to investigate the weighted LSE based on the feature extraction for the endmember abundance estimation, because the extracted features can be weighted according to their different contributions to the LSE. In addition, the idea of this paper could be utilized to design a signal-classification-based compression scheme for hyperspectral imagery, which could lead to a better compression result than the scheme based on conventional signal representation concepts.

ACKNOWLEDGMENT

The author would like to thank L. Bruce, J. Fowler, Y. Huang, R. King, and N. Younan for many helpful discussions. The author also acknowledges Remote Sensing Technology Center at Mississippi State University for supporting this project.

REFERENCES

[1] 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.

[2] 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 Sens.*, vol. 12, no. 5, pp. 1119–1129, 1991.

[3] 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, pp. 1067–1080, Sept. 1994.

[4] 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 Sens. Environ.*, vol. 31, pp. 1–26, 1990.

[5] 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 Sens.*, vol. 17, no. 17, pp. 3373–3400, 1996.

[6] 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, pp. 16–20, Jan. 1991.

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

[8] 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. 19 419–19 425, Aug. 1998.

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

[10] L. O. Jimenez and D. A. Landgrebe, "Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. Syst., Man, Cybern. C*, vol. 28, pp. 39–54, Jan. 1998.

[11] ——, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2653–2667, Nov. 1999.

[12] 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, pp. 977–980, Apr. 2002.

[13] J. Li, L. M. Bruce, J. Byrd, and J. Barnett, "Automated detection of Pueraria montana (kudzu) through Haar analysis of hyperspectral reflectance data," in *Proc. IGARSS*, vol. 5, 2001, pp. 2247–2249.

[14] 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," in *Proc. IGARSS*, vol. 1, 2001, pp. 527–529.

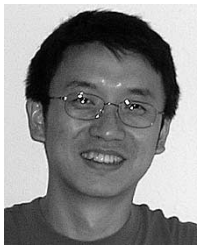
[15] J. Li, "Linear unmixing of hyperspectral signals via wavelet feature extraction," Ph.D. dissertation, Mississippi State Univ., Mississippi State, 2002.

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

[17] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 2000.

[18] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[19] S. Mallat, *A Wavelet Tour of Signal Processing*. Orlando, FL: Academic, 1999.



Jiang Li (S'99–M'04) received the B.S. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1994, the M.S. degree from the University of Nevada, Las Vegas, in 1999, and the Ph.D. degree from the Mississippi State University, Mississippi State, in 2002, all in electrical engineering.

From 1994 to 1997, he was an Electrical R&D Engineeer with China Aerospace Science and Industry Corporation, Beijing. He is currently a Scientist with Baker Hughes, Inc., Houston, TX. His research interests include signal and image processing, data compression, noise suppression, and data transmission and telemetry.

Dr. Li is a member of Eta Kappa Nu, SEG, and SPWLA.