An Abundance Characteristic-Based Independent Component Analysis for Hyperspectral Unmixing

Nan Wang, Bo Du, *Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, and Lifu Zhang, *Member, IEEE*

***Abstract*—Independent component analysis (ICA) has been recently applied into hyperspectral unmixing as a result of its low computation time and its ability to perform without prior information. However, when applying ICA for hyperspectral unmixing, the independence assumption in the ICA model conflicts with the abundance sum-to-one constraint and the abundance nonnegative constraint in the linear mixture model, which affects the hyperspectral unmixing accuracy. In this paper, we consider an abundance matrix composed of Np-dimensional variables, and we propose a new hyperspectral unmixing approach with an abundance characteristic-based ICA model. Two characteristics of the abundance variables are explored, and the model is constructed by these characteristics. A corresponding gradient descent algorithm is also proposed to solve the proposed objective function. Both the synthetic and real experimental results demonstrate that the proposed method performs better than the other state-of-the-art methods in abundance and endmember extraction.**

***Index Terms*—Abundance characteristic, convex geometry, hyperspectral unmixing, independent component analysis (ICA), orthogonal subspace projection.**

I. Introduction

HYPERSPECTRAL images (HSIs) captured by image spectroscopy contain both spatial and spectral information [1], [2], which can help the users to better extract the ground-object information. However, due to the limited spatial resolution and the complexity of the terrain, a single pixel in HSI always contains different materials, i.e., "mixed pixels" always exist in the HSI [3]. These mixed pixels in the HSI are, in fact, the main obstacle for HSI analysis. To solve the problem and make full use of the rich spectral information, spectral unmixing technologies have been developed in recent years [4]. These technologies decompose a mixed pixel into a set of spectra called "endmembers," as well as the corresponding abundances. In other words, the classification procedure is performed at a subpixel scale.

The linear mixture model (LMM) and the nonlinear mixture model (NLMM) are the two basic models used in hyperspectral unmixing [5]. For an overview of recent advances in nonlinear unmixing, see [6]. Compared with the NLMMs, the LMM is more widely applied due to its simplicity and definite physical meaning. More research has therefore been based on the LMM. In this research, four strategies for unmixing have been developed: 1) algorithms based on simplex geometric theory; 2) sparse regression-based methods; 3) algorithms considering the spatial–spectral contextual information; and 4) algorithms based on statistical methods [7]. In simplex geometric theory, the hyperspectral data set can be considered as a simplex, and the vertices correspond to the endmembers in the data set. Many different approaches have been developed to search for the endmembers, such as the N-FINDR [8], simplex growing algorithm (SGA) [9], pixel purity index (PPI) [10], vertex component analysis (VCA) [11], simplex identification via split augmented Lagrangian (SISAL) [12], and so on. N-FINDR and SGA search the simplex with the maximum volume and define the vertices of the simplex as the endmembers, while PPI and VCA search them by a data set projection analysis. SISAL finds a simplex embedding the data set through an augmented Lagrangian method. These approaches mainly focus on the extraction of endmembers, and the estimation of abundances requires the conventional nonconstrained or constrained least squares methods. In recent years, the unmixing problem has often been considered as a sparse problem, with the assumption that the observed pixels can be expressed by linear combinations of a number of pure spectral signatures that are known in advance (a spectral library is used for the known signatures). Since the number of endmembers in one pixel is very small compared with the size of the spectral library, sparse theory can be used for the unmixing [7]. Zare and Gader [13] expanded the iterated-constrained-endmember algorithm that incorporates sparsity-promoting priors to find the correct number of endmembers. Qian *et al.* [14] extended the nonnegative matrix factorization (NMF) method by incorporating an $L_{1/2}$ sparsity constraint, which induces sparsity in the unmixing procedure. Charles *et al.* [15] learned an optimal library under the sparse representation model and used it to learn the spectral signatures of the materials in the image. Some researchers have also added

Manuscript received September 13, 2013; revised December 25, 2013 and February 26, 2014; accepted May 1, 2014. This work was supported in part by the National Basic Research Program of China (973 Program) under Grants 2011CB707105 and 2012CB719905, by the National Natural Science Foundation of China under Grant 61102128, and by the Fundamental Research Funds for the Central Universities under Grant 211-274175.

N. Wang is with the State Key Laboratory of Remote Sensing Sciences, Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, Beijing 100101, China (e-mail: nwangchina@gmail.com).

B. Du is with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: gunspace@163.com).

L. Zhang is with the Remote Sensing Group, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

L. Zhang is with the State Key Laboratory of Remote Sensing Sciences, Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, Beijing 100101, China (e-mail: zhanglf@radi.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2014.2322862

spatial–spectral information for hyperspectral unmixing. Automated morphological endmember extraction [16], as proposed by Plaza *et al*., is based on mathematical morphology, a classic image processing technique that can be applied in the spectral domain of data. Jia and Qian [17] proposed a measure of the spatial complexity to describe the spatial autocorrelation, and the estimated abundances vary smoothly from one pixel to its neighbors. Zare [18] considered a weighted combination of the abundances related to the neighboring pixels as a regularization term to make the abundances smooth.

Statistical approaches can show an effective result if there are no pure pixels in the image, and they can, in general, simultaneously obtain endmembers and the corresponding abundances. Dobigeon *et al.* [19] proposed a new hierarchical Bayesian model to generate results fitting the nonnegative and sum-to-one constraints. Blind source separation (BSS) approaches have also been used for hyperspectral unmixing and can be considered to belong to the statistical approaches.

BSS extracts the signals from the observations, without prior information, and different assumptions about the signal sources derive the different BSS approaches [20]–[23]. Independent component analysis (ICA) [24] and NMF [25] are typical statistical approaches. NMF decomposes a high-dimensional data set into two nonnegative matrices [26]. It does not require a pure-pixel occurrence assumption and naturally keeps the abundance nonnegative constraint (ANC) in the LMM. As a result, NMF has been widely studied in the field of hyperspectral imagery by many researchers [27]. Many different NMF-based approaches have been proposed for the spectral unmixing application. Huck *et al.* [28] proposed a method called minimum dispersion constrained NMF, in which the variance of each spectra is employed to constrain the basic NMF, thereby preserving the possible spectra. Miao and Qi [29] proposed the minimum volume constrained NMF (MVCNMF), which introduces a minimum volume constraint as the "internal force" to make the estimated simplex circumscribe the data cloud. Jia and Qian [30] proposed the piecewise smoothness NMF with a sparseness constraint, which imposes both piecewise smoothness and sparseness constraints on NMF. Liu *et al.* [31] proposed an approach named abundance separation and smoothness constrained NMF, which introduces two constraints, namely, the abundance separation constraint and the abundance smoothness constraint, into the basic NMF. Wang *et al.* [32] proposed an endmember dissimilarity constrained NMF, which defines an endmember dissimilarity function as a constraint to make the estimated endmember spectra dissimilar and smooth.

ICA is another statistical approach for hyperspectral unmixing. Compared with NMF, ICA has a lower computational cost. Bayliss *et al.* [33] applied ICA for hyperspectral unmixing by considering the endmembers' spectra as the original sources. However, because of the limited spectral bands in HSI, there are insufficient statistical characteristics for hyperspectral unmixing with ICA. Therefore, other researchers have considered the abundance vectors as the original sources [34]–[36]. Unfortunately, the independence assumption in ICA contradicts the abundance sum-to-one constraint (ASC) in LMM [34], which affects the performance of ICA. A solution to this problem was recently proposed through the combination of ICA and Bayesian positive source separation [37], which estimates the more meaningful spectral signals. However, the abundances are not estimated simultaneously with the endmember extraction, which may cause a large error between the real abundances and the estimated abundances. Another method is dependent component analysis [38], [39], which assumes that the abundances have a Dirichlet distribution. A result which satisfies both the ASC and ANC can be obtained under this approach. Wei *et al.* [40] also proposed a constrained ICA (CICA) model for hyperspectral unmixing, which adds the ASC and ANC into the ICA model, and they proposed an adaptive abundance modeling technology to characterize the statistical distribution of the data. The CICA model changes the independence assumption in the ICA model so that it is suitable for hyperspectral unmixing. The model is constructed from a mathematical point of view, which will lead to less meaningful results.

By considering the abundance matrix as a p-dimensional variable with N samples, we propose an abundance characteristic-based ICA (ACICA) model for hyperspectral unmixing. Based on the LMM, the abundance variables must be included in the specific convex. Moreover, each ground objective has its own distribution range, which means that the abundance variables should be less independent. According to the two characteristics, we construct the new model from three considerations.

1) We follow the idea of ICA based on geodesic search to keep the data set nonnegative.
2) We apply orthogonal subspace projection theory to formulate an abundance sum-to-one function, and we combine it with ICA based on geodesic search to ensure that the data set is included in the specific convex.
3) We add the mutual information as a regularization item to control the degree of independence of the abundance variables.

Compared with most of the state-of-the-art algorithms based on NMF and ICA, the proposed approach can be implemented without the elaborate selection of a proper initialization, which avoids the problem of the results not being accurate without proper initial values. Experiments with both synthetic and real data sets also confirm the advantages of the proposed method.

The remainder of this paper is organized as follows. Section II briefly presents the LMM and ICA based on geodesic search. Section III details the proposed method. The experiments with synthetic and real hyperspectral data are described in Sections IV and V, respectively. Section VI concludes this paper.

II. Related Works

A. LMM

The LMM assumes that a pixel in a hyperspectral data set is a linear mixture of P known material signatures called endmembers, $\mathbf{A}_{band\times P}=[a_1, a_2, \ldots, a_P]$, where a_i is the endmember spectra with dimension $band$. The corresponding proportion is called the abundance $\mathbf{S}_{P\times N} = [s_1, s_2, \ldots, s_P]^{\mathrm{T}} = [\omega_1, \omega_2, \ldots, \omega_N]$, where s_i is a N-dimension vector. N is the number of image pixels, and P is the number of endmembers.

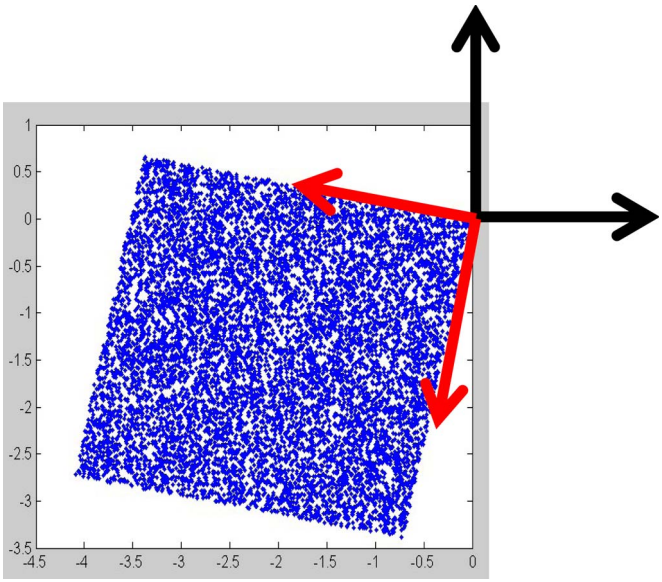

Fig. 1. Illustration of ICA based on geodesic search. The blue points are the whitened data set without mean removal, and the black arrowheads refer to the corresponding coordination system. The red arrows are the rotated coordination system computed by ICA based on geodesic search.

Based on the LMM, each pixel in a HSI data set can be expressed as

$$x = A\gamma\omega + \varepsilon \tag{1}$$

where x is a $band \times 1$ vector representing one pixel in the HSI, γ is a scale factor modeling the illumination variability due to the surface topography [8], and ε is the residual error. According to LMM, the abundance matrix should satisfy the ASC and ANC simultaneously, i.e., $s_1^T + s_2^T + \cdots + s_p^T = \mathbf{1}^{\mathrm{T}}$ and $s_i^{\mathrm{T}} \geq 0$.

B. ICA-Based Geodesic Search

ICA can extract independent sources from the mixed observations, with the assumption that the sources are statistically independent from each other. The basic ICA model can be referred to as (1), where $s = [s_1, s_2, \ldots, s_p]^{\mathrm{T}}$ are the original sources and are independent of each other, $x = [x_1, x_2, \ldots, x_N]^{\mathrm{T}}$ are the observations, and A is the mixing matrix. To find the independent components in the observations, an orthogonal matrix W is searched to make $y = Wx = s$, where it is assumed that there is no noise. As we know, the original sources s are independent of each other, which means that the mutual information between them is zero. Therefore, minimizing the mutual information between $y = Wx$ can help for searching the matrix W and generating the estimated signals. The nonnegative property is common in the real world, but the traditional ICA model cannot achieve this nonnegative constraint. Plumbley [41] proposed ICA based on geodesic search to solve this problem. In the algorithm, it searches an orthogonal and rotational matrix W such that $Y = WZ$ is nonnegative, where Y denotes the estimated signals and Z is the whitened matrix from the original data set X. Fig. 1 shows an illustration of ICA based on geodesic search.

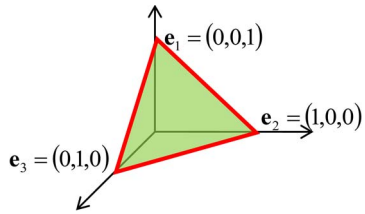

Fig. 2. Specific convex including the abundance variables.

To make the estimated signals nonnegative, the objective function is constructed as follows:

$$\min\ J = \frac{1}{2}\mathrm{trace}\left(Y_- Y_-^T\right) \tag{3}$$

where $Y_- = \min(Y, 0)$ is the negative part of the matrix Y. To find the minimum of the objective function over the matrix W, which is constrained to be orthonormal, a Stiefel manifold [42] is used. The distance between two points on the Stiefel manifold, which is named "geodesic," is defined as

$$W(\tau) = e^{\tau B} W(0) \tag{4}$$

where B is a skew-symmetric matrix and τ is a scalar parameter determining the position along the geodesic. Based on the Stiefel manifold, a steepest descent geodesic algorithm was developed by the authors, which can directly keep the matrix W orthonormal in each iteration. However, this steepest descent geodesic algorithm only searches the orthonormal matrix, which may not generate a steady and satisfactory result for hyperspectral unmixing. The details of this algorithm can be found in [41, Sec. V].

III. Proposed Method

This section proposes a new model for hyperspectral unmixing. The motivation behind the improved objective function is from two aspects.

1) According to the abundance constraints in LMM, the sample abundance variables should constitute a specific convex, which is

$$\mathrm{conv}\{\mathbf{e}_1, \ldots \mathbf{e}_p\} = \left\{ s = \sum_{i=1}^{p} \theta_i \mathbf{e}_i \middle| \boldsymbol{\theta} \in R_+^p, \quad \mathbf{1}_p^T \boldsymbol{\theta} = 1 \right\} \tag{5}$$

in which $I = [\mathbf{e}_1, \ldots \mathbf{e}_p]$ is a $p \times p$ identity matrix.

Fig. 2 shows the specific convex constructed by the abundance variables. However, only satisfying these constraints will not correctly extract the abundances due to the fact that many abundance pairs can fit the constraints. Therefore, the other characteristic needs to be explored to avoid this problem.

2) In the real world, each ground object has a smooth distribution in the scene; therefore, the abundance variables corresponding to the different ground objects have a dispersed distribution. Fig. 3 illustrates the distribution of the abundance variables in the 2-D convex. We can see that the abundance variables are included and dispersed in a convex, which shows that they are correlated, i.e., not completely independent.

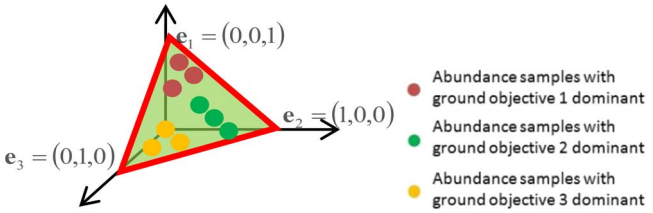

Fig. 3. Distribution of the abundance variables in the 2-D specific convex.

Above all, we can extract the variables with less independence in the specific convex as the abundance maps. The new objective function is proposed as follows:

$$\min\ F(W) = G1 + G2 + \mu G3,\ \text{where}\ G1 = \frac{1}{2}\text{trace}\left(Y\_Y\_^T\right). \tag{6}$$

In (6), function $G1$ is the ICA model. We add functions $G2$ and $G3$ to achieve the aforementioned two motivations, where minimizing function $G1 + G2$ can keep the estimated data in the specific convex and minimizing function $G3$ can make the estimated data less independent. μ is the regularization parameter, which controls the intensity of the degree of independence. $G2$ is an ASC function constructed by orthogonal subspace theory [48], and $G3$ is a mutual information function measured by the degree of the independence of the abundance variables. The details of the objective function are as follows.

A. ASC Function

According to the LMM, the ASC can be seen as an affine set

$$\text{affine}\{\mathbf{e}_1, \ldots, \mathbf{e}_p\} = \left\{ s = \sum_{i=1}^{p} \theta_i \mathbf{e}_i, \mathbf{1}_p^T \boldsymbol{\theta} = 1 \right\} \tag{7}$$

where $[\mathbf{e}_1, \ldots \mathbf{e}_p] = I_{p\times p}$, $I_{p\times p}$ is the identity matrix. The affine set can also be presented as follows [43]:

$$\text{affine}\{\mathbf{e}_1, \ldots \mathbf{e}_p\} = \{s = C\boldsymbol{\alpha} + \mathbf{d} | \boldsymbol{\alpha} \in R^{p-1}\} = \Lambda(C, \mathbf{d}). \tag{8}$$

Due to $\{\mathbf{e}_1, \ldots \mathbf{e}_p\}$ being affinely independent, we can get the matrix

$$C = [\mathbf{e}_1 - \mathbf{e}_p, \ldots, \mathbf{e}_{p-1} - \mathbf{e}_p]_{L\times(p-1)} \tag{9}$$

$$\mathbf{d} = \mathbf{e}_p. \tag{10}$$

According to (8), the vector s included in the affine set can satisfy the following equation:

$$s - \mathbf{d} = C\boldsymbol{\alpha}. \tag{11}$$

Therefore, we can measure the distance from a vector to the affine set by projecting the vector to the orthogonal subspace of matrix C. The distance function of the estimated abundance vector y can be defined as follows:

$$g = (y - \mathbf{d})^T C^{\perp} (y - \mathbf{d}) \tag{12}$$

$$C^{\perp} = I - C(C^T C)^{-1} C^T \tag{13}$$

where $y_{p\times 1}$ is a column vector in the abundance matrix Y and $C^{\perp}$ is the orthogonal subspace of matrix C. Under (12), if the abundance vector is included in the affine set, the function value g is zero; otherwise, the function value is a nonnegative value since the matrix $C^{\perp}$ is a nonnegative-definite matrix. Therefore, for all the vectors, we can define the whole distance function $G2$ as

$$G2 = \frac{1}{N}\sum_{i=1}^{N} (y_i - \mathbf{d})^T C^{\perp} (y_i - \mathbf{d}) \tag{14}$$

where y_i is any column vector in the abundance matrix Y. Equation (14) can also be presented in a matrix style as follows:

$$G2 = \frac{1}{N}\text{trace}\left[(Y - D)C^{\perp}(Y - D)^T\right] \tag{15}$$

where matrix D is written as

$$D = \begin{bmatrix} 0, \ldots, 0 \\ \vdots, \ldots, \vdots \\ 0, \ldots, 0 \\ 1, \ldots, 1 \end{bmatrix}_{p\times N}. \tag{16}$$

For (15), the greater the number of vectors that fall into the affine set, the closer the function is to zero. Therefore, through minimizing (15), the ASC can be satisfied for hyperspectral unmixing.

B. Mutual Information Function

Mutual information is usually used to measure the degree of independence between signals. The mutual information function is defined as follows:

$$MI = D_{KL}\left(p(\mathbf{y}) \left\| \prod_{i=1}^{p} p(y_i)\right.\right) \tag{17}$$

where $\mathbf{y} = [y_1, y_2, \ldots y_p]$ is a random vector, $p(\mathbf{y})$ is the joint probability density, and $p(y_i)$ is the probability density of y_i. According to (17), the mutual information is used to measure the independence between the random variables. The range value of the MI function is $[0, +\infty]$, and when the value is zero, the variables are independent. For $MI > 0$, the variables have a limited degree of correlation. In practice, the abundance variables are usually dispersed in a convex, and they also present a low independence. Therefore, the independence degree between the abundance variables can be well described by the mutual information. In this paper, when minimizing the objective function, we can adjust the parameter to control the mutual information to be a proper value fitting the degree of dispersion. Therefore, the dispersed property of the abundance variables can be satisfied.

Due to the fact that the distribution of the signals is usually unknown in advance, the mutual information needs to be estimated. One of the methods is to use the cumulant of the signals. The estimation formula used in this paper is presented in detail in Section III-C.

C. Proposed ACICA

In this section, more details of the ACICA model are presented, and the corresponding update rules and the stopping conditions are also developed.

Combining (6), (15), and (17), the objective function is as follows:

$$\min\ F(W) = \frac{1}{2}\text{trace}\left(Y_{-}Y_{-}^{T}\right) + \frac{1}{N}\text{trace}\left[(Y-D)C^{\perp}(Y-D)^{T}\right] + \mu D_{KL}\left(p(\mathbf{y})\left\|\prod_{i=1}^{p}p(y_i)\right.\right) \tag{18}$$

where $Y = WZ$. $Z_{P\times N}$ is the dimensionally reduced matrix from the original data set X, and it can be found in (28) (see Section III-D).

In (18), the first term on the right can constrain the estimated variables to be nonnegative, and the second term can constrain the estimated variables to be included in a specific convex. Minimizing the first and second terms, we can constrain the estimated variables in the convex in which the real abundance variables are found. As mentioned earlier, variables included in the specific convex are not unique; therefore, the third term in (18) is necessary. By minimizing the third term, the estimated variables are separated from each other, which fits the real distribution of the abundance maps. Additionally, through changing the parameter μ, we can control the degree of the independence to improve the final accuracy.

To optimize (18), we choose the gradient descent algorithm

$$W \leftarrow W - \lambda\Delta W \tag{19}$$

where $\Delta W = \partial F(W)/\partial W$ and λ is the step size. We calculate the gradient ΔW as follows.

1) The gradient of the first item is

$$\Delta W_1 = \frac{\partial G1(W)}{\partial W} = Y_{-}Z^{T}. \tag{20}$$

2) The gradient of the second item is

$$\Delta W_2 = \frac{\partial G2(W)}{\partial W} = \frac{2}{N}C^{\perp}(Y-D)Z^{T}. \tag{21}$$

In the third term of (18), because we do not know the real probability density function (pdf) of the abundance distributions, the pdf of the abundance distributions needs to be estimated. Several studies [49], [50] have provided possible ways to estimate the pdf. One way is that the pdf can be expressed by a series expansion with respect to a Gaussian distribution. Therefore, the estimation is more accurate when the estimated pdf is close to a Gaussian distribution. According to Amari *et al.* [20], the gradient of the third term can be defined as follows:

$$\Delta W_3 = \frac{\partial G3(W)}{\partial W} = \left(QZ^{T} - (W^{T})^{-1} * N\right)/N \tag{22}$$

where

$$Q = (f * \mathbf{1}).*(Y.^{\wedge}2) + (g * \mathbf{1}).*(Y.^{\wedge}3) \tag{23}$$

$$f = -\frac{3}{48}(8k_3./\sigma^3 - 12k_3.*k_4./\sigma^7)$$

$$g = -\frac{4}{48}(2k_4./\sigma^4 - 9k_4^2./\sigma^8 - 6*k_3^2./\sigma^6). \tag{24}$$

In (23) and (24), the symbols "./" and ".∗" represent dot division and dot multiplication, respectively. ".^" represents the power computing of each element in the matrix, and "∗" represents the multiplication computing. k_3 and k_4 are the third- and fourth-order cumulants, and σ is the variance of the variable in matrix Y, which is updated at every iteration.

Combining (20)–(22), the gradient of the objective function can be written as

$$\Delta W = Y_{-}Z^{T} + \frac{2}{N}C^{\perp}(Y-D)Z^{T} + \mu\left(QZ^{T} - (W^{T})^{-1}*N\right)/N. \tag{25}$$

The stopping condition of this algorithm is that the distance between the new and the old objective function value is smaller than a threshold ρ

$$|F^{k-1} - F^{k}| < \rho \tag{26}$$

where $k-1$ and k are the previous and current iterations, relatively. We can then optimize the objective function by (19) and (25), and the estimated abundance matrix Y can finally be obtained. We can calculate the endmember matrix A by the nonnegative least squares algorithm.

D. Data Preprocessing—DR

Before performing the proposed method, the number of endmembers needs to be estimated if there is no prior information. Eches *et al.* [51] provided an algorithm to estimate the number of endmembers. Our method also needs dimensionality reduction (DR) to make matrix W square. Some researchers have, in fact, proposed DR algorithms [52]–[55] for HSIs. In the traditional ICA, whitening is an important step before the ICA iteration procedure. This is because it can do half of the work of ICA, removing any second-order dependences in the data set. However, for hyperspectral unmixing, the whitening step should be modified in two aspects: 1) The abundance variables are always nonnegative, so to ensure that most of the estimated abundance values are also nonnegative, we do not remove the mean of the original data set X; and 2) the abundance variables are always correlated with each other. Therefore, we whiten the original data set with matrix $\sum$, as defined in (27), to keep the variables correlated. The whitening step in this paper is as follows.

First, define the matrix of the original data set

$$\sum = XX^{T}/N \tag{27}$$

where N is the number of pixels.

Second, calculate the eigenvalue matrix D and eigenvector matrix E of the matrix $\sum$, and reduce the dimension of the original data set to p dimensions with

$$Z = \left(D_p^{-1/2}E_p^{T}\right)X \tag{28}$$

where D_p is a diagonal matrix with the first p eigenvalues in the diagonal line and E_p is the corresponding vectors.

After the DR, $ACICA$ can be performed to optimize the objective function. Additionally, compared with many of the statistical algorithms based on NMF or ICA, the proposed method in this paper does not need a proper initialization.

E. Procedure of ACICA

The flowchart of ACICA is as follows:

Input:

Data information: original hyperspectral data set $\boldsymbol{X}$ and the number of endmembers P.

Parameter: the regularization parameter μ and the step size λ in the gradient descent algorithm, and the threshold ε.

Begin:

Undertake the dimension reduction by (27) and (28);

Set the matrix W as the identity matrix.

for k=1: maximum iteration

update the matrix W by (19) and (25) until satisfying (26)

end

Calculate the endmember matrix A based on the original data set X and the estimated abundance matrix Y by constrained least squares minimization.

End

IV. Synthetic Data Set Experiment

In this section, ACICA is evaluated by the use of synthetic data. Several classical hyperspectral unmixing methods are used for the comparison, including the CICA [40], MVC-NMF [29], SISAL [12], and VCA with fully constrained least squares [11].

The two criteria of the spectral angle distance (SAD) [5] and root-mean-square error (RMSE) [44] are used to evaluate the unmixing results. These two criteria measure the similarity between the estimated results and the reference values. SAD is defined as

$$SAD = \cos^{-1} \frac{a_i^T \cdot \hat{a}_i}{\left|a_i^T\right| \left|\hat{a}_i\right|} \tag{29}$$

where a_i is an estimated endmember signature and $\hat{a}_i$ is the corresponding reference signature.

RMSE measures the similarity between the estimated abundance and the corresponding reference abundance, which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(s_j - \hat{s}_j)^2} \tag{30}$$

where s_j is the estimated abundance and $\hat{s}_j$ is the corresponding reference abundance.

The spectral signatures in the synthetic data are chosen from the United States Geological Survey (USGS) digital spectral library. Fig. 4 shows these five endmember signatures and their names. The generation of abundances is similar to the method in [11], with the abundance fractions being randomly generated based on the Dirichlet distribution given by

$$p(\alpha_1, \alpha_2, \ldots \alpha_p) = \frac{\Gamma(\psi_1 + \psi_2 + \cdots + \psi_p)}{\Gamma(\psi_1)\Gamma(\psi_2)\ldots\Gamma(\psi_p)} \times \alpha_1^{\psi_1 - 1}\alpha_2^{\psi_2 - 1}\ldots\alpha_p^{\psi_p - 1} \tag{31}$$

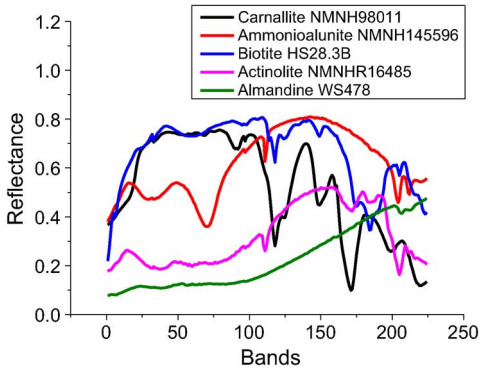

Fig. 4. Endmember spectra of the synthetic data.

where $0 < \alpha_i < 1$, $\sum_i^p \alpha_i = 1$, $E(\alpha_i) = \psi_i \sum_k^p \psi_k$ is the expected value of the ith abundance fraction, and $\Gamma(\psi_i)$ is the Gamma function, whose value is decided by ψ_i. In the synthetic experiments, the parameter γ controlling the illumination variability follows a $Beta(\beta_1, \beta_2)$ distribution. β_2 is fixed as 1, and β_1 is used for changing parameter γ. We employ the parameter η to denote the maximum value of the abundance fractions in the data set. The noise in the synthetic data is Gaussian noise, while the SNR is defined as

$$SNR = 10\log_{10}\frac{E[r^T r]}{E[\varepsilon^T \varepsilon]} \tag{32}$$

where r denotes the signal information without noise, $E[r^T r]$ denotes the signal intensity, and $E[\varepsilon^T \varepsilon]$ denotes the noise intensity.

Extensive experiments on the simulated data set are conducted to evaluate the effectiveness of ACICA from five important aspects: 1) In the first experiment, the algorithms are evaluated with varying SNR to analyze the robustness with regard to different levels of noise; 2) in the second experiment, the parameter β_1, which controls the degree of illumination due to the surface topography, is used to evaluate the algorithms; 3) in the third experiment, the algorithms are evaluated with regard to the absence of pure pixels. The parameter η denotes the maximum value of abundance fractions in the data set. If a pure pixel exists, η equals 1, and η will decrease with the absence of pure pixels; 4) in the fourth experiment, the number of endmembers P is varied in the scene in order to evaluate the algorithm performance under different numbers of land objects; and 5) in the final experiment, the algorithms are evaluated by different numbers of pixels N in the HSI data set, corresponding to different scales of data set.

Additionally, at the close of this section, we also analyze the effect of the regularization item, the convergence of the objective function under our optimization algorithm, the calculation complexity, and the effect of the parameters.

A. Experiment A (Noise Robustness Analysis)

In this experiment, the SNR changes from 10 to 30 dB. The hyperspectral scene has 1296 pixels and 5 endmembers. Parameter γ is Beta distributed with $\beta_1 = 10$ and $\beta_2 = 1$. Parameter

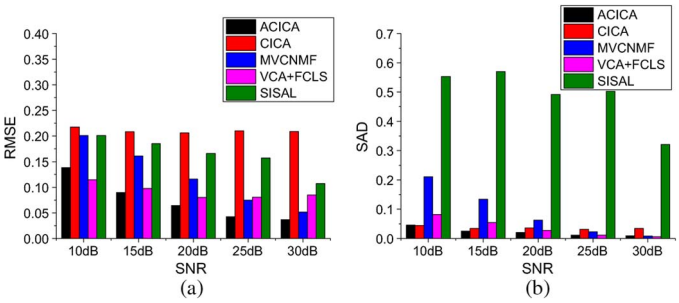

Fig. 5. Experimental results with different SNR. (a) RMSE. (b) SAD.

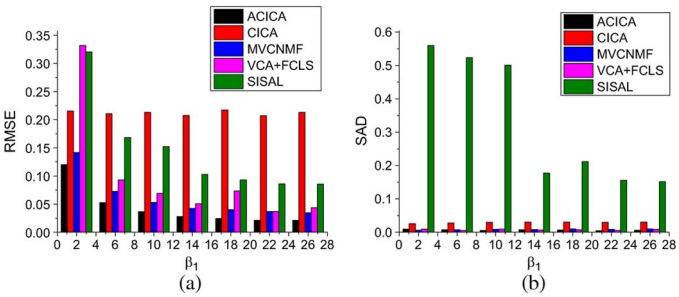

Fig. 6. Experimental results with different values of the degree of illumination. (a) RMSE. (b) SAD.

η is fixed as 1. Fig. 5(a) and (b) presents the mean values of RMSE and SAD for all abundances and endmembers, respectively, and the unit for the SAD angle is radiance. In terms of RMSE, the proposed ACICA gives the best performance at 15–30 dB, while vertex component analysis + full constrained least square (VCA+FCLS) gives the second-best performance. The performance of MVCNMF deteriorates rapidly with the decrease in SNR. SISAL also gives a good performance, while CICA could not give a satisfactory performance in these terms. In terms of SAM, ACICA gives the best performance at 10–25 dB, followed by CICA, VCA, MVCNMF, and SISAL. Overall, the proposed method performs best with the varying levels of noise.

B. Experiment B (Degree of Illumination Robustness Analysis)

Due to fluctuations in surface topography, the degree of illumination is different for different images. In this experiment, parameter γ distributed with $Beta(\beta_1, \beta_2)$ is used to describe the illumination cases, and β_1 and β_2 are used to change the parameter γ. In the experiment, the range of β_1 is 2–26, $\beta_2 = 1$, and the quality of the data sets decreases as β_1 decreases. The size of the abundance map N is 1296, the number of endmembers P is 5, and the SNR is fixed to 30 dB in these data sets. Fig. 6(a) and (b) illustrates the average values of RMSE and SAD, respectively. In terms of the RMSE values, ACICA gives the best performance. VCA+FCLS, SISAL, and MVCNMF also perform well at $\beta > 10$. In terms of SAD, ACICA also gives a stable and satisfactory performance. VCA+FCLS, MVCNMF, and CICA also perform well, but SISAL does not perform very well in this experiment. It is concluded that the proposed method is robust with regard to the variation in the degree of illumination.

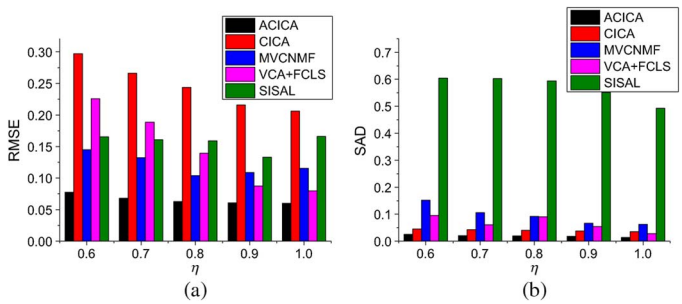

Fig. 7. Experimental results with different degrees of mixing. (a) RMSE. (b) SAD.

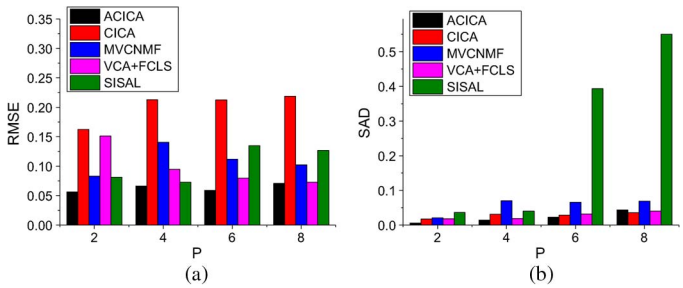

Fig. 8. Experimental results with different numbers of endmembers. (a) SAD. (b) RMSE.

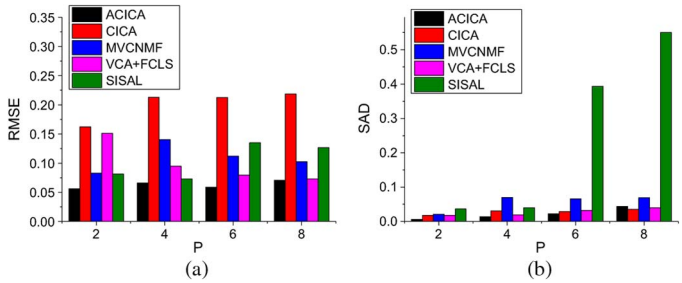

Fig. 9. Experimental results with different number of pixels. (a) RMSE. (b) SAD.

C. Experiment C (Robustness Analysis to Degree of Mixing)

In this experiment, the parameter η controlling the degree of mixing is changed from 0.6 to 1. The number of pixels in the scene is 1296, and the number of endmembers is 5. The parameter γ is decided by $\beta_1 = 10$ and $\beta_2 = 1$, and the SNR is fixed as 20 dB. The average values of RMSE and SAD are plotted in Fig. 7(a) and (b). The SAD and RMSE values show that ACICA performs best. In terms of RMSE, MVCNMF also performs well, followed by VCA+FCLS, SISAL, and CICA. In terms of SAD, VCA+FCLS has the second-best performance, followed by CICA and MVCNMF. From Fig. 7, we can see that no matter whether pure pixels exist or not, ACICA performs the best.

D. Experiment D (Robustness Analysis to the Number of Endmembers)

In this experiment, the number of endmembers is changed from 2 to 8 in the scene, and the number of pixels is fixed at 1296. The parameter γ is decided by $\beta_1 = 10$ and $\beta_2 = 1$, the SNR is fixed as 20 dB, and the parameter η is fixed as 1. Fig. 8(a) and (b) plots the average values of RMSE and SAD.

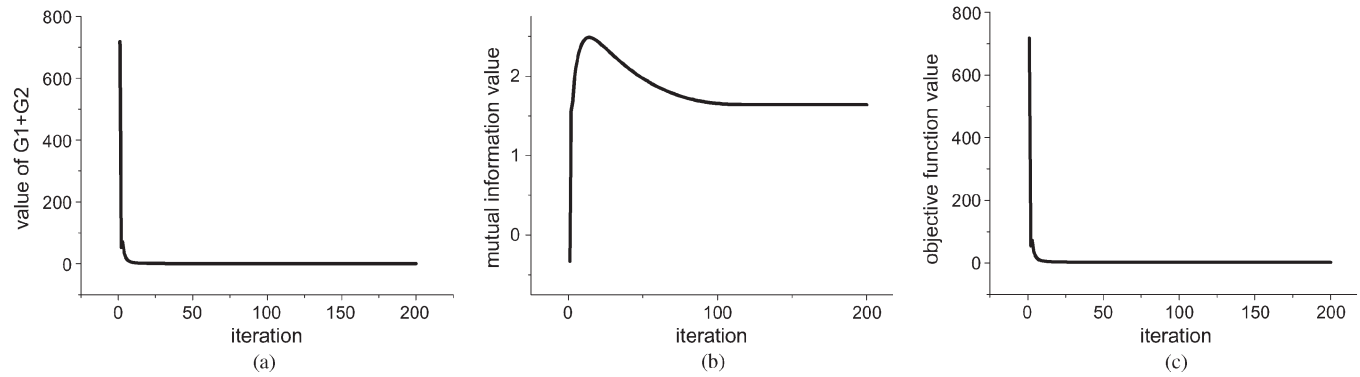

Fig. 10. Values in the objective function with the first 200 iterations. (a) Value of the first item plus second item in the objective function. (b) Mutual information value (third item in the objective function). (c) Value of the objective function.

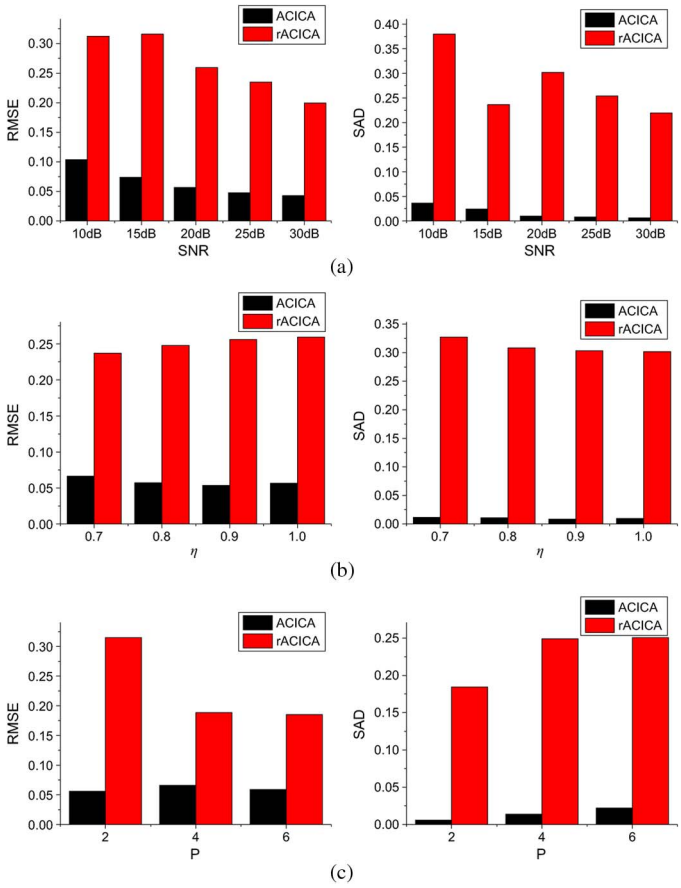

Fig. 11. Effectiveness test for the removal of ACICA. (a) Results with different SNR. (b) Results under different purity levels. (c) Results under different numbers of endmembers.

From the values of RMSE, it can be seen that ACICA performs best, and VCA+FCLS performs better with large numbers of endmembers. MVCNMF also performs well, followed by SISAL and CICA. In terms of SAD, ACICA gives the best performance in most cases, but with high numbers of endmembers, and CICA and VCA+FCLS perform better than ACICA and MVCNMF in terms of SAD. SISAL performs well at low numbers of endmembers. The aforementioned observations are proof that the proposed method achieves the best performance in abundance estimation with complex distributions of ground objects, and the method also presents an acceptable performance in endmember extraction.

TABLE I
COMPUTATIONAL COMPLEXITY OF THE ALGORITHMS

Algorithm	*Computational complexity*
ACICA	$o(P^2(4N+P))$
MVCNMF	$o(2[P^2(L+N)+PLN]+P^2(L+P)+P!)$
VCA	$2P^2N$
CICA	$o(P^2(4N+5P))$
SISAL	$o(NP)$

E. Experiment E (Robustness Analysis to the Number of Pixels)

In this experiment, the number of pixels is changed from 100 to 10 000 in the scene, and the number of endmembers is fixed as 6. The parameter γ is decided by $\beta_1 = 10$ and $\beta_2 = 1$, the SNR is fixed as 20 dB, the number of endmembers $P = 6$, and parameter η is fixed as 1. The average values of SAD and RMSE are plotted in Fig. 9(a) and (b). Both the SAD values and RMSE values show that ACICA achieves the best performance among all the methods. In terms of RMSE, VCA+FCLS has the second-best performance, followed by MVCNMF, CICA, and SISAL. For the SAD values, CICA has the second-best performance, followed by VCA+FCLS, MVCNMF, and SISAL. We therefore conclude that the proposed method is suitable for different sizes of hyperspectral data set.

F. Convergence Analysis, Regularization Item Analysis, Calculation Complexity, and Parameter Analysis

The convergence results under our update rules are shown in Fig. 10. Fig. 10(a) shows the value of the first two items in the objective function $G1 + G2$, which means the degree of the estimated abundance vectors included in the specific convex. We can see that, with the growth in the iterations, the value is close to zero, which shows that most of the estimated abundance vectors are included in the convex. Fig. 10(b) shows the value of the mutual information $G3$ in the objective function, which shows the degree of independence between the variables. We can see that, after the preprocessing, the abundance variables show a low mutual information value, but the value is stable with the growth of the iterations and it is greater than 0, which shows that the estimated abundance variables are not completely independent. From Fig. 10(a) and (b), we can see

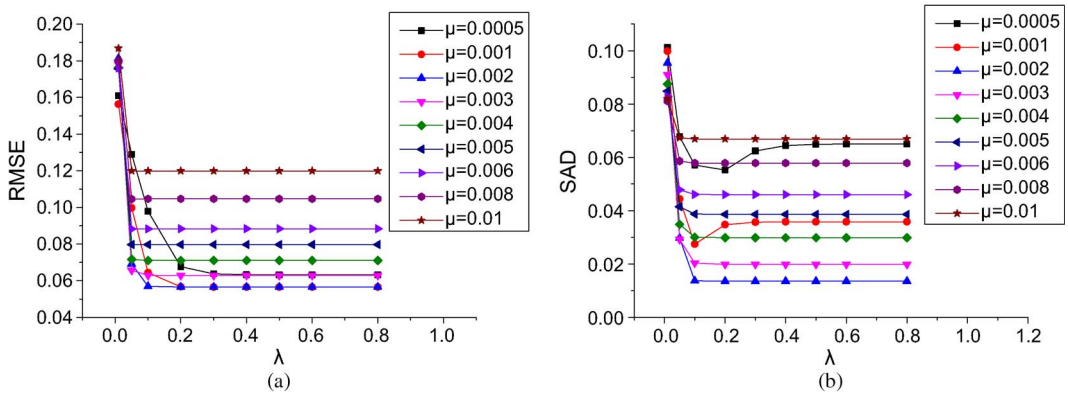

Fig. 12. Parameter analysis.

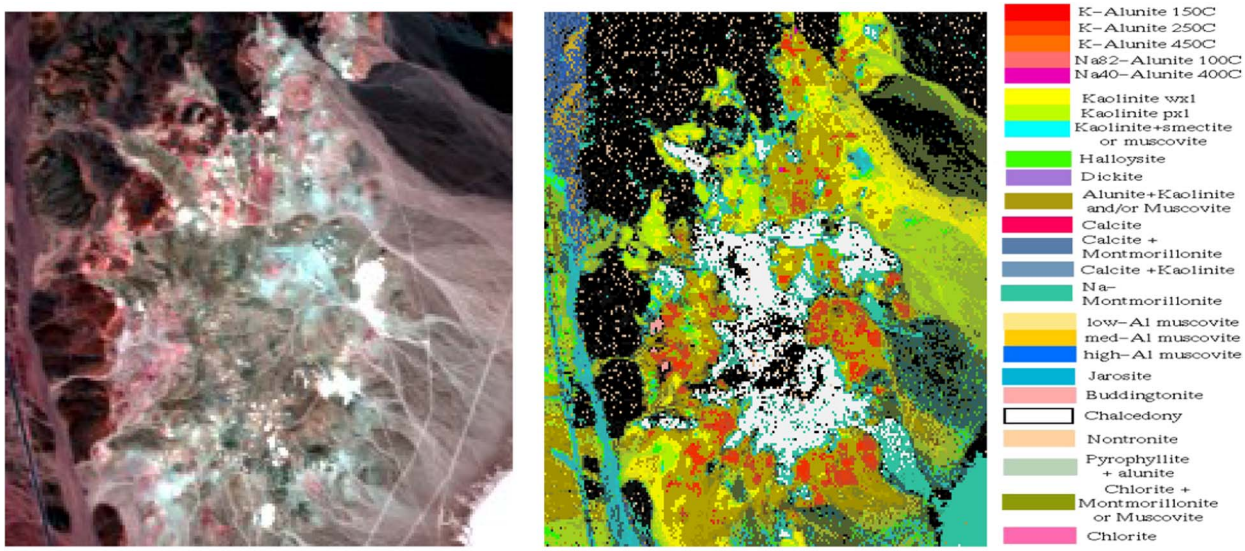

Fig. 13. (a) Subscene extracted from the Cuprite data set. (b) Nevada Cuprite reference map by the USGS, 1995.

that, under our update rule, the characteristics can be satisfied well. Fig. 10(c) shows the change in the objective function values with the iterations, and we can see that they converge with the iterations.

We also test the effectiveness of the mutual information item in the objective function. We remove the mutual information regularization item from the ACICA model (we refer to it as rACICA) to evaluate its effect on ACICA. Fig. 11 shows the experimental results under several different cases. From the experimental results, we can see that, without the regularization item, the algorithm shows a poor accuracy, which proves that applying the dispersed characteristic of the abundances is important to improve the accuracy.

The computational complexity of the algorithms is shown in Table I. From the table, we can see that MVCNMF has the highest computational complexity, while VCA has the lowest. ACICA has an average computational complexity.

We now analyze the effect of the parameters. There are three parameters in the paper: μ which controls the mutual information in the objective function, λ which is the step size of the gradient descent, and ε which is the criterion of the convergence. In the experiments, it was found that the algorithm can converge well, and ε is fixed at 0.0001 in all the experiments; therefore, we only discuss μ and λ here. We select a typical synthetic data set to analyze the parameters: $P = 5$, $N = 1296$, $\beta_1 = 10$, $\beta_2 = 1$, SNR $= 20$ dB, and purity $= 0.8$. We change parameter $\mu \in [5 \times 10^{-4}, 1 \times 10^{-2}]$ with $\lambda \in [0.05, 0.8]$ and plot the SAD and RMSE values, considering λ as the $\boldsymbol{x}$ coordinate as Fig. 12.

We can see that, when we fix parameter μ, the SAD and RMSE values are basically stable, with λ falling in the range [0.05,0.8]. This shows that, when the selection of λ falls in this interval, the parameter will not affect the accuracy. However, if λ is set too large, the iterations will diverge. λ is set as 0.5 in most of the experiments. Only for the Hyperspectral

TABLE II
SPECTRAL ANGLE BETWEEN ESTIMATED ENDMEMBERS AND REFERENCE ENDMEMBERS IN THE CUPRITE DATA SET

Mineral Name	*ACICA*	*CICA*	*MVCNMF*	*VCA+FCLS*	*SISAL*
Buddingtonite	0.0918	0.1116	0.0981	**0.0763**	-
Muscovite	**0.0579**	0.1530	0.0762	0.1057	0.1048
Alunite	**0.0561**	0.3173	0.1376	0.2790	0.3611
Chalcedony	0.1698	**0.1223**	0.2182	0.1604	1.263
Kaolin	0.0941	**0.0827**	0.1871	0.1761	0.9556
Montmorillonite#1	**0.0786**	0.0803	0.1006	0.0747	0.1057
Dumortierite	0.1217	**0.0969**	0.1853	0.1652	0.2485
Montmorillonite#2	0.1183	**0.0874**	0.1011	0.1113	0.1718
Desert Varnish	**0.1929**	0.2601	0.6038	0.3560	2.6298
Average	**0.1077**	0.1457	0.1898	0.1672	0.7300

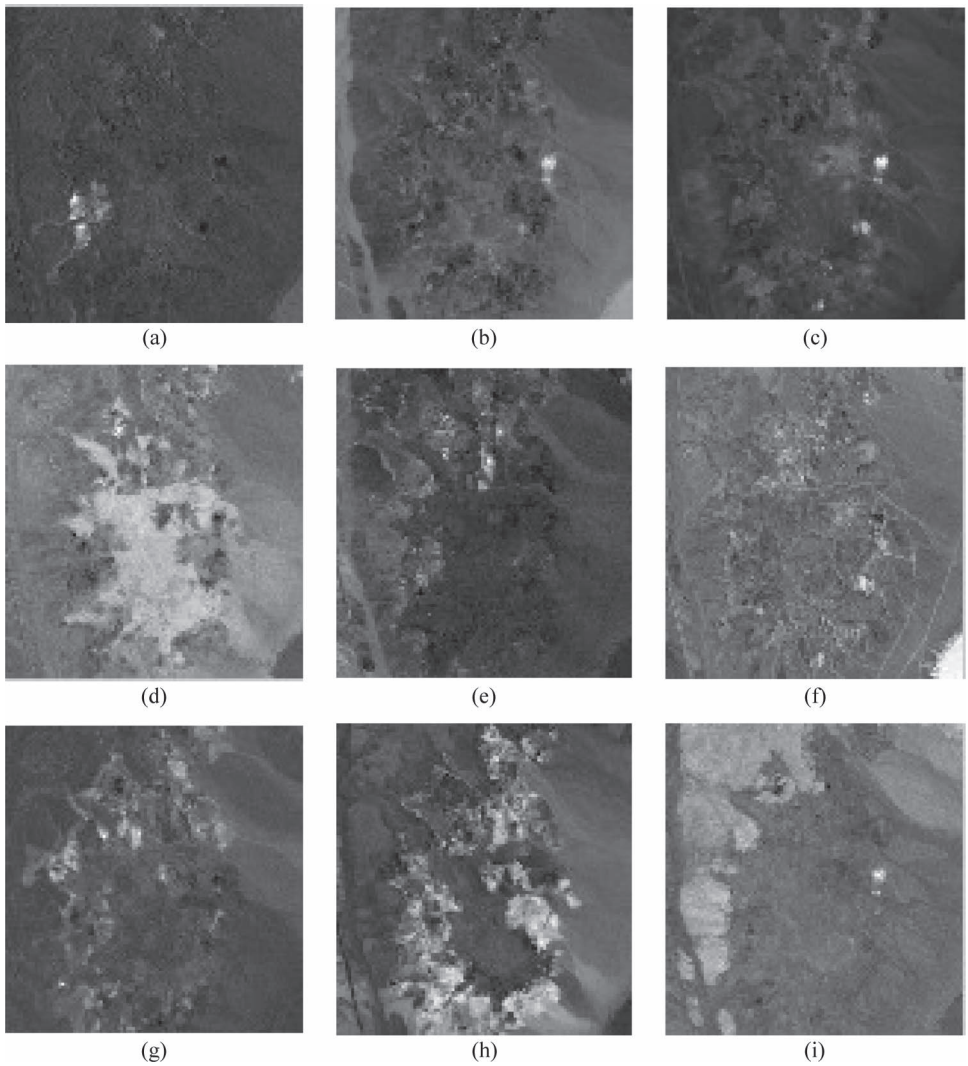

Fig. 14. Abundance maps of ACICA in the Cuprite data set. (a) Buddingtonite. (b) Muscovite. (c) Alunite. (d) Chalcedony. (e) Kaolin. (f) Montmorillonite#1. (g) Dumortierite. (h) Montmorillonite#2. (i) Desert varnish.

Fig. 15. Subscene extracted from the urban data set.

Digital Imagery Collection Experiment (HYDICE) data set will 0.5 lead to divergence; therefore, λ is set as 0.08 to keep the convergence. If we constrain SAD as < 0.05 and RMSE as < 0.09, we can see that μ should be set in the range [0.001,0.006].

TABLE III
SPECTRAL ANGLE BETWEEN ESTIMATED ENDMEMBERS AND REFERENCE ENDMEMBERS IN THE URBAN DATA SET

Mineral Name	*ACICA*	*CICA*	*MVCNMF*	*VCA+FCLS*	*SISAL*
Roof1	0.7529	**0.2401**	0.9507	0.8635	1.0919
Grass	**0.0899**	0.0909	0.5215	0.4219	0.5682
Roof2	0.3109	**0.0763**	0.3080	0.2937	0.3218
Tree	**0.0529**	0.0582	0.1194	0.0954	1.0464
Street	**0.0396**	0.0766	0.1265	0.2277	2.5837
Average	0.2492	**0.1084**	0.4052	0.3804	1.1224

From the parameter analysis, μ is important for the accuracy. In most of the experiments, we set μ as 0.003. Only for the data sets with a very poor quality should μ be set to be larger to lead to a better accuracy. The reason for this is that, for a normal data set with a certain correlation t among the abundances, a fixed value of μ can extract the abundances accurately. However, if the quality of the data set is very poor, e.g., the SNR is very low, under a fixed value of parameter μ, the estimated abundances will include much noise, and in fact, the correlation among the abundances cannot reach t. Therefore, we should set a larger μ to overcome the effect of the poor-quality data set.

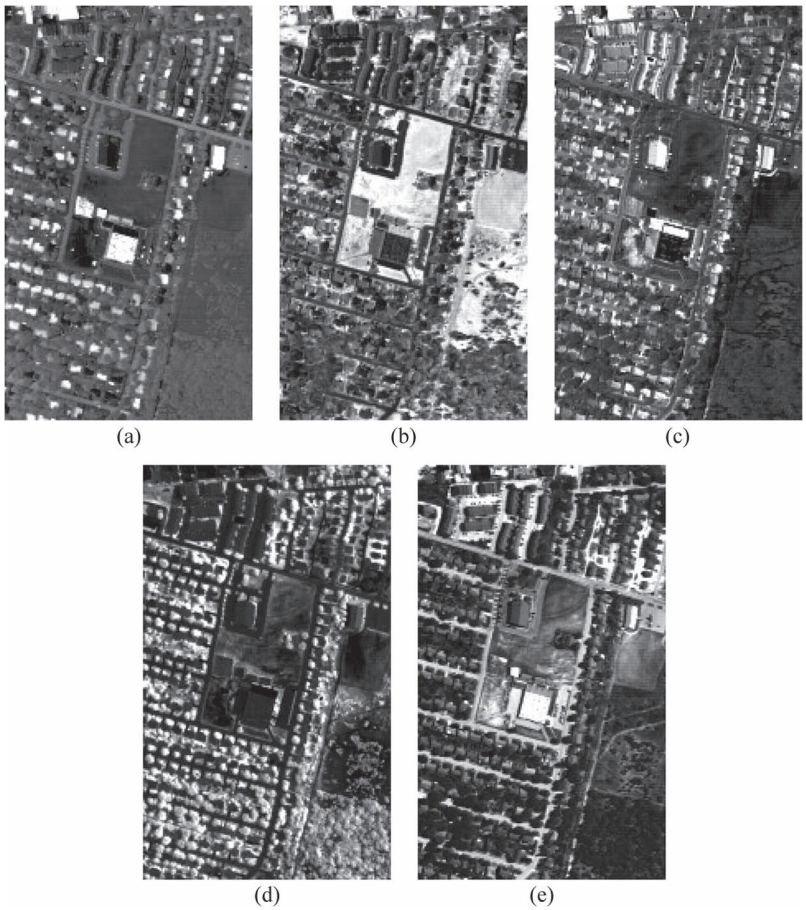

Fig. 16. Abundance maps of ACICA in the urban data set: (a) roof#1, (b) grass, (c) roof#2, (d) tree, and (e) street.

V. Experimental Results and Analysis on Real Hyperspectral Data

A. AVIRIS Data Set

In this section, two real hyperspectral data sets are used to evaluate the algorithms. The first data set was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in June 1997, covering Cuprite in the state of Nevada, USA. There are 224 bands in the data set, covering the wavelength range of 0.37–2.48 μm. The spectral resolution is 10 nm. Swayze and Clark have reported on the ground truth of the area [46], [47]. The USGS mineral spectral library contains the main mineral spectra in the area and is used for the reference endmember signatures [11]. For our experiment, a block with the size of 200×190 is cut from the original data. The false-color image and the corresponding reference map captured by the USGS in 1995 are shown in Fig. 13(a) and (b), respectively.

Before unmixing, the noisy bands (1–3 and 221–224) and water absorption bands (104–115 and 148–170) are removed, leaving 182 bands. The number of the endmembers is computed as 9, according to the virtual dimension method [45]. The SAD values between the reference endmembers and the estimated endmembers by the different methods are shown in Table II, while the average for these methods' performances is also shown in the last line. The best performance is denoted by a bold font. In the results, montmorillonite is divided into two endmembers due to the signature variability. From the SAD values in Table II, ACICA and CICA have the largest number of best performance cases; therefore, the average values are calculated for the evaluation. We can see that the proposed algorithm has the lowest average value, while CICA also performs better than VCA+FCLS, MVCNMF, and SISAL. The abundance maps with ACICA are shown in Fig. 14.

B. HYDICE Data Set

The second hyperspectral data set, covering an urban area, was collected by the HYDICE sensor, which has 210 spectral channels, with a spectral resolution of 10 nm, ranging between 0.4–2.4 μm. A highly mixed area is cut from the original data set in this experiment, with the size of 200×150. Fig. 15 shows the true-color image composed of R-band 1, G-band 89, and B-band 162.

The subscene includes five types of ground objects: two types of roof, tree, grass, and street. Five pixels are selected manually, corresponding to the five ground objects, and serve as the reference endmembers. The SAD values between the reference endmembers and the estimated endmembers by the different methods are shown in Table III. The best performance is denoted by a bold font. In the results, roof is divided into two endmembers, due to the different materials. From the SAD values in Table III, ACICA has the largest number of best performance cases in the five classes. It cannot perform best in the class "Roof," which may be because roof has a high spectral variability while

our method has no initialization. CICA also gives a better performance in endmember extraction than VCA+FCLS, MVC-NMF, and SISAL, and CICA has the best average value. The abundance maps with ACICA are shown in Fig. 16.

VI. Conclusion

This paper proposes a new approach for hyperspectral unmixing. Two characteristics of the abundance variables are explored: 1) Based on the LMM, the abundance variables must be included in the specific convex, and 2) the abundance variables are dispersed, i.e., less independent. To satisfy the first characteristic, this paper applies orthogonal subspace projection theory to formulate an abundance sum-to-one function and combines it with the idea of ICA based on geodesic search to ensure that the data set is included in the specific convex. To satisfy the second characteristic, we add the mutual information as a regularization item to control the degree of independence of the abundance variables. We also develop a corresponding gradient descent algorithm to solve the proposed objective function. The proposed method can also perform well without initialization, which avoids the problem of the results being incorrect without a proper initialization.

The proposed ACICA is evaluated with several classical hyperspectral unmixing methods: CICA, MVCNMF, and VCA+FCLS, for both synthetic and real hyperspectral data sets. The synthetic experiment results show that the proposed method performs the best in the abundance extraction, and it also perform the best, in most cases, for the endmember extraction. For the real hyperspectral data set results, we can see that the estimated abundance maps are clear and accurate, while the proposed method also has the most best performance cases, which proves that the proposed method is effective for hyperspectral unmixing.

References

[1] A. F. H. Goetz, G. Vane, J. Solomon, and B. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, Jun. 1985.

[2] Q. Tong, Y. Xue, and L. Zhang, "Progress in hyperspectral remote sensing science and technology in China over the past three decades," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 70–91, Jan. 2014.

[3] G. Foody, *Remote Sensing Image Analysis: Including the Spatial Domain*. Norwell, MA, USA: Kluwer, 2004, ch. 3, pp. 37–49.

[4] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Plenum, 2003.

[5] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[6] N. Dobigeon *et al.*, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 89–94, Jan. 2014.

[7] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[8] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrometry V*, 1999, vol. 3753, pp. 266–275.

[9] C.-I. Chang, C.-C. Wu, W. Liu, and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Oct. 2006.

[10] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. AVIRIS Workshop*, 1993, vol. 1, pp. 11–14, JPL Pub. 93-26.

[11] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[12] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 1st Workshop WHISPERS*, 2009, pp. 1–4.

[13] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jun. 2007.

[14] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via l_1,2 sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[15] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 963–978, Sep. 2011.

[16] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.

[17] S. Jia and Y. Qian, "Spectral and spatial complexity-based hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3867–3879, Dec. 2007.

[18] A. Zare, "Spatial-spectral unmixing using fuzzy local information," in *Proc. IEEE IGARSS*, Oct. 2011, pp. 1139–1142.

[19] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Process.*, vol. 89, no. 12, pp. 2657–2669, Dec. 2009.

[20] S. I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, vol. 8. Cambridge, MA, USA: MIT Press, 1996, pp. 757–763.

[21] M. Li, Y. Liu, G. Feng, Z. Zhou, and D. Hu, "OI and fMRI signal separation using both temporal and spatial autocorrelations," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1917–1926, Aug. 2010.

[22] K. J. Friston, "Experimental design and statistical parametric mapping," in *Human Brain Function*, R. S. Frackowiak *et al.*, Ed., 2nd ed. San Diego, CA, USA: Academic, 2004, pp. 599–634.

[23] A. Zepeda, C. Arias, and F. Sengpiel, "Optical imaging of intrinsic signals: Recent developments in the methodology and its applications," *J. Neurosci. Meth.*, vol. 136, no. 1, pp. 1–21, Jun. 2004.

[24] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY, USA: Wiley, 2001.

[25] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neur. Inf.*, 2000, vol. 13, pp. 556–562.

[26] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts," in *Proc. Adv. Neur. Inf. Process. Syst.*, 2003, pp. 1–8.

[27] G. X. Zhou *et al.*, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.

[28] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010.

[29] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[30] S. Jia and Y. T. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.

[31] X. S. Liu, W. Xia, B. Wang, and L. M. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 757–772, Feb. 2011.

[32] N. Wang, B. Du, and L. Zhang, "An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 6, no. 2, pp. 554–569, Apr. 2013.

[33] J. Bayliss, J. A. Gualtieri, and R. F. Cromp, "Analyzing hyperspectral data with independent component analysis," *Proc. SPIE*, vol. 3240, pp. 133–143, Mar. 1998.

[34] J. Nascimento and J. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[35] C. Chen and X. Zhang, "Independent component analysis for remote sensing study," *Proc. SPIE*, vol. 3871, pp. 150–158, Dec. 1999.

[36] S.-S. Chiang, C.-I. Chang, and I. W. Ginsberg, "Unsupervised hyperspectral image analysis using independent component analysis," in *Proc. IGARSS*, 2000, vol. 7, pp. 3136–3138.
[37] S. Moussaoui *et al.*, "On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation," *Neurocomputing*, vol. 71, no. 10–12, pp. 2194–2208, Jun. 2008.
[38] J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," in *Proc. IEEE IGARSS*, Jul. 2007, pp. 4033–4036.
[39] J. M. Bioucas-Dias and J. Nascimento, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 863–878, Mar. 2012.
[40] X. Wei, X. S. Liu, B. Wang, and L. M. Zhang, "Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2165–2179, Jun. 2011.
[41] M. D. Plumbley, "Algorithm for nonnegative independent component analysis," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 534–543, May 2003.
[42] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.
[43] T. H. Chan, C. Y. Chi, Y. M. Huang, and W. K. Ma, "Convex analysis-based minimum volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.
[44] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 650–663, Mar. 2004.
[45] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
[46] D. Landgrebe, *Multispectral Data Analysis: A Signal Theory Perspective*. West Lafayette, IN, USA: School Elect. Comput. Eng., Purdue Univ., 1998.
[47] R. N. Clark and G. A. Swayze, "Evolution in imaging spectroscopy analysis and sensor signal-to-noise: An examination of how far we have come," in *Proc. 6th Annu. JPL Airborne Earth Sci. Workshop*, Mar. 1996, pp. 49–53.
[48] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.
[49] M. G. Kendall and S. Alan, *The Advanced Theory of Statistics, Vols. II and III*. New York, NY, USA: Hafner, 1961.
[50] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–231, Apr. 1994.
[51] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 582–591, Jun. 2010.
[52] N. Acito, M. Diani, and G. Corsini, "Hyperspectral signal subspace identification in the presence of rare signal component," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1940–1954, Apr. 2010.
[53] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[54] P. Bajorski, "Second moment linear dimensionality as an alternative to virtual dimensionality," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 672–678, Feb. 2011.
[55] T. Zhou and D. Tao, "Double shrinking sparse dimension reduction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 244–257, Jan. 2013.

Nan Wang received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University, Wuhan, China, in 2014.

She is currently a Postdoctoral Researcher with the State Key Laboratory of Remote Sensing Sciences, Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, Beijing, China. Her research interests include hyperspectral image processing and signal processing.

Bo Du (M'12) received the B.S. degree from Wuhan Univeristy, Wuhan, China, in 2005 and received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University, in 2010.

He is currently an Associate Professor with the School of Computer, Wuhan University. His major research interests include pattern recognition, hyperspectral image processing, and signal processing.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 260 research papers. He is the holder of five patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Electrical Engineers, Executive Member (Board of Governor) of the China National Committee of the International Geosphere-Biosphere Programme, Executive Member of the China Society of Image and Graphics, etc. He regularly serves as a Cochair of the series International Society for Optics and Photonics Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and Geoinformatics Symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Journal of Geo-spatial Information Science*, and *Journal of Remote Sensing*.

Lifu Zhang (S'04–M'05) received the B.E. degree in photogrammetry and remote sensing from the Department of Airborne Photogrammetry and Remote Sensing, Wuhan Technical University of Surveying and Mapping (WTUSM), Wuhan, China, in 1992, the M.E. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, WTUSM, Wuhan, in 2000, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, in 2005.

He was a Visiting Researcher with the Department of Information and Computer Sciences, Nara Women's University, Nara, Japan, from 2003 to 2004, a Postdoctoral Researcher with the Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing, China, from 2005 to 2007, and an Advanced Visiting Researcher with the Earth Science and Resource Engineering, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia, in 2011. In 2007, he joined the Institute of Remote Sensing Applications, Chinese Academy of Sciences, where he is currently a Professor and the Head of the Hyperspectral Remote Sensing Laboratory. His main research interests focus on hyperspectral remote sensing, imaging spectrometer system development and its applications, etc.

Dr. Zhang is a member of SPIE and the Academy of Space Science of China, a committeeman of the Chinese National Committee of the International Society for Digital Earth (CNISDE), a Vice-Chairman of the Hyperspectral Earth Observation Committee of CNISDE, and a Standing Committeeman of the Expert Committee of China Association of Remote Sensing Applications.