Endmember Extraction from Highly Mixed Data Using Minimum Volume Constrained Non-negative Matrix Factorization

Lidan Miao, *Student Member, IEEE*, Hairong Qi, *Senior Member, IEEE*

***Abstract*— Endmember extraction is a process to identify the hidden pure source signals from the mixture. In the past decade, numerous algorithms have been proposed to perform this estimation. One commonly used assumption is the presence of pure pixels in the given image scene, which are detected to serve as endmembers. When such pixels are absent, the image is referred to as the *highly* mixed data, for which these algorithms at best can only return certain data points that are close to the real endmembers. To overcome this problem, we present a novel method without the pure pixel assumption, referred to as the minimum volume constrained non-negative matrix factorization (MVC-NMF), for unsupervised endmember extraction from highly mixed image data. Two important facts are exploited: first, the spectral data are non-negative; second, the simplex volume determined by the endmembers is the minimum among all possible simplexes that circumscribe the data scatter space. The proposed method takes advantage of the fast convergence of NMF schemes, and at the same time eliminates the pure pixel assumption. The experimental results based on a set of synthetic mixtures and a real image scene demonstrate that the proposed method outperforms several other advanced endmember detection approaches.**

***Index Terms*— Spectral unmixing, non-negative matrix factorization, endmember, convex geometry, simplex.**

I. Introduction

Since the first launch of an earth observation satellite, remote sensing imagery has been increasingly utilized for mineral exploration, environmental monitoring, military surveillance, etc. A common problem associated with satellite images is the wide existence of *mixed* pixels [1], within which more than one type of material is present. Thus, the measured spectrum of a single pixel is a mixture of several ground cover spectra known as *endmembers*, weighted by their fractional *abundances*. Very often, to utilize the measured hyperspectral data, one has to decompose these mixed pixels into a set of endmember signatures and their corresponding proportions. This process is called *spectral unmixing* or *mixed pixel decomposition*, which involves two procedures: the first and most challenging step is to identify the endmember signatures (*endmember extraction*); and the second step is to infer the proportions of different endmembers in forming each pixel (*abundance estimation*).

Manuscript received May 09, 2006; revised September 18, 2006. This work was supported in part by Office of Naval Research under grant no. N00014-04-1-0797.

The authors are with the Advanced Imaging & Collaborative Information Processing (AICIP) Group, Department of Electrical and Computer Engineering, University of Tennessee, Knoxville, TN 37996, USA. (e-mail: lmiao1@utk.edu; hqi@utk.edu)

During the past few decades, a great deal of endmember extraction algorithms have been proposed [2]. Three projection pursuit approaches are investigated, including principal component analysis (PCA) [3], independent component analysis (ICA) [4], [5], and singular value decomposition (SVD) [6]. PCA finds a set of orthogonal vectors based on the second-order decorrelation, which best represents the original image data in a least squares sense. To investigate the high-order statistics of independent sources, ICA looks for a transformation in which the data are all-order statistically independent. And, SVD seeks the projections that best represent the image data in a maximum power sense. All these methods have strong mathematical foundations and fast implementations, but they share the same problem that the extracted endmembers are not guaranteed to be non-negative. In addition, the estimates from PCA and SVD do not have the physical meaning of the source signals.

Another group of algorithms exploits the strong parallelism between the linear mixture model and the theory of convex geometry. Under linear mixing, the observations in a hyperspectral scene are within a simplex whose vertices correspond to the endmembers [7], [8]. In this sense, the problem of endmember extraction is equivalent to finding the vertices of a simplex that encloses the data cloud. There are a variety of such approaches. For example, the minimum volume transform (MVT) [8] finds the convex hull that circumscribes the data cloud and then fits a simplex with the minimum volume to it. The problem with this method is that the identification of the convex hull is computationally prohibitive. To speed up the process, some algorithms [9]–[11] assume the presence of *pure* pixels, i.e., pixels containing only one type of endmember materials, which correspond to the simplex vertices. Then, the endmembers are extracted by finding the extreme-valued pixels in the given scene.

Vertex component analysis (VCA) [11] is one of the most advanced convex geometry-based endmember detection methods with the pure pixel assumption. Considering the variations due to the surface topography, VCA models the data using a positive cone, whose projection onto a properly chosen hyperplane is a simplex with vertices being the endmembers. After projecting the data onto the selected hyperplane, VCA projects all image pixels to a random direction and uses the pixel with the largest projection as the first endmember. The other endmembers are identified by iteratively projecting data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember is then

selected as the pixel corresponding to the extreme projection.

Although the assumption of the presence of pure pixels improves the algorithm efficiency, in some cases, such as when the processing data are with low spatial resolution [12] or of specific ground covers [13], it is not reliable to make this assumption, especially for all endmember classes. Finding the endmembers from such *highly* mixed data is therefore a more challenging task. One solution suggested is to use the iterative constrained endmember (ICE) method [13], which formulates an optimization problem with an effort to minimize the reconstruction error regularized by a constraint term, i.e., the sum of variances of the simplex vertices. Within each iteration, if the number of pixels is N, the algorithm then involves N quadratic programmings to estimate the abundances, which makes this method unrealistic for large data sets. Another approach [12] uses a finer spatial resolution image (contains pure pixels) as the auxiliary data to derive the fractional abundances for all image pixels. Then, the endmember signatures are estimated using a least squares method. Although the algorithm shows promising results, the performance highly depends on the registration between the high and the low resolution image. Additionally, the need of high spatial resolution data appears as another limitation.

In recent years, non-negative matrix factorization (NMF) [14], [15] has been applied to hyperspectral data unmixing [16]–[18]. As will be described later, NMF finds a set of non-negative basis vectors that approximates the original data through linear combinations. These basis vectors thus play a similar role as the endmembers. However, the standard NMF algorithms do not impose any constraint on these bases except for non-negativity, which is not sufficient enough to lead to a well-defined problem.

In order to render better estimates, the smoothness constraint is investigated in [18]. The algorithm follows the derivation of the standard multiplicative rule [14] while introduces the smoothness constraint. For each estimated endmember, it finds the closest signature from a spectral library, and then uses the matched laboratory signature as the final endmember. The fractional abundance is then estimated using a constrained least squares method. However, finding the best match from a spectral library is not reliable due to the strong atmospheric and environmental variations.

In this paper, we present a new constrained NMF method, which integrates the least squares analysis and the convex geometry model by incorporating a volume constraint into the NMF formulation, referred to as the minimum volume constrained NMF method (MVC-NMF). Two important facts are exploited: first, the spectral data are non-negative; second, the simplex volume determined by the endmembers is the minimum among all possible simplexes that circumscribe the data scatter space. The proposed cost function consists of two parts. One part measures the approximation error between the observed data and the reconstructions from the estimated endmembers and abundances, and the other part consists of the minimum volume constraint. We can think of these two terms serving as two forces: the *external* force (minimizing the approximation error) drives the estimation to move outward of the data cloud; and the *internal* force (minimizing the simplex volume) acts in the opposite direction by forcing the endmembers to be as close to each other as possible. Through experimental validations, we observe that the balance between these two forces effectively guides the learning process to converge to the true endmember locations.

The rest of the paper is organized as follows. In Section II, we first introduce the linear mixing model widely adopted in spectral unmixing analysis and give a short review of the NMF technique. We then analyze the solution space of pixel unmixing and NMF from a geometric point of view, which leads to the formulation of the volume constrained NMF method. Section III describes the problem formulation and the learning strategies. The performance evaluation using a set of synthetic images and a real hyperspectral scene are presented in Sections IV and V, respectively. Section VI concludes the paper and discusses the future work.

II. Non-negative Matrix Factorization for Spectral Data Analysis

As a blind source separation method, non-negative matrix factorization has been adopted to solve the problem of mixed pixel decomposition [16]–[18]. The motivation is straightforward, considering the non-negative property of spectral measurement and the very similar mathematical modeling between the spectral unmixing analysis and the non-negative matrix factorization. In addition, we present a more informative analysis from a geometric point of view, from which we reason the underlying principle and formulate the proposed MVC-NMF algorithm.

A. Linear Mixing Model

In solving the spectral unmixing problem, the linear mixing model (LMM) has gained significant popularity due to its effectiveness and simplicity. An LMM is valid when the endmembers are distributed as discrete patches, in which different endmembers do not interfere with each other [2], [19]. Mathematically, the model is given by

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{x} \in \mathbf{R}^l$ is an observation vector at a single pixel with l spectral bands. $\mathbf{A} \in \mathbf{R}^{l\times c}$ is the material signature matrix (source matrix) whose columns, $\{\mathbf{a}_j\}_{j=1}^c \in \mathbf{R}^l$, correspond to the spectral signatures of different endmembers, and c is the number of endmembers. The abundance vector is denoted by $\mathbf{s} \in \mathbf{R}^c$, which satisfies two physical constraints, referred to as the abundance non-negative constraint, $s_j \geq 0,\ j = 1, 2, \ldots, c$, and the abundance sum-to-one constraint, $\sum_{j=1}^c s_j = 1$. The possible errors and noises are taken into account by an l-dimensional column vector $\boldsymbol{\epsilon}$.

In most cases, it is reasonable to assume that the entire image consists of a few number of endmembers, and all the image pixels share the same source matrix $\mathbf{A}$. Then, we can arrange the measurement vectors at all pixel locations into the columns of an $l \times N$ data matrix $\mathbf{X}$ (N is the number of pixels), which results in the following model

$$\mathbf{X} = \mathbf{A}\mathbf{S} \tag{2}$$

where the columns of $\mathbf{S} \in \mathbf{R}^{c\times N}$ correspond to the fractional abundances. Note that we have removed the noise term, since it can be incorporated into $\mathbf{A}$ by considering the endmember variations. When written in this form, it is evident that the unmixing problem is to factorize the measured data matrix into two low-rank matrices, subjected to the non-negative and the sum-to-one constraint.

B. Non-negative Matrix Factorization

Given a non-negative matrix $\mathbf{Y} \in \mathbf{R}^{m\times n}$ and a positive integer $r < \min(m,n)$, the task of non-negative matrix factorization is to find two matrices $\mathbf{W} \in \mathbf{R}^{m\times r}$ and $\mathbf{H} \in \mathbf{R}^{r\times n}$ with non-negative elements such that

$$\mathbf{Y} \approx \mathbf{WH} \tag{3}$$

or equivalently, the columns $\{\mathbf{y}_j\}_{j=1}^n$ are expressed as

$$\mathbf{y}_j \approx \mathbf{W}\mathbf{h}_j \tag{4}$$

where $\mathbf{y}_j \in \mathbf{R}^m$, and $\mathbf{h}_j \in \mathbf{R}^r$. The parameter r is the desired rank of matrix $\mathbf{W}$, and normally it is assumed to be known *a priori* or can be determined based on the given data $\mathbf{Y}$. Presumably, the columns of $\mathbf{W}$ represent the latent variables, i.e., physically meaningful non-negative "parts", of the underlying data. This nature has found NMF a wide range of applications in data analysis, dimensionality reduction, feature extraction and target recognition, etc. A comparison between models in Eq. 2 and Eq. 3 clearly shows the potential of applying NMF to decompose mixed pixels.

One natural way to solve the NMF problem is to formulate an optimization problem by minimizing the Euclidean distance between $\mathbf{Y}$ and $\mathbf{WH}$ [14], [15],

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{W},\mathbf{H}) = \frac{1}{2}\|\mathbf{Y}-\mathbf{WH}\|_F^2 \\ \text{subject to} \quad & \mathbf{W} \succeq \mathbf{0}, \mathbf{H} \succeq \mathbf{0} \end{aligned} \tag{5}$$

where the symbol $\succeq$ denotes *componentwise inequality*, i.e., $\mathbf{W} \succeq \mathbf{0}$ means $w_{ij} \geq 0$ for $i = 1,2,\ldots,m$, $j = 1,2,\ldots,r$. The operator $\|\cdot\|_F$ represents the Frobenius norm given by

$$\|\mathbf{Y}-\mathbf{WH}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij}-(WH)_{ij})^2 \tag{6}$$

For the above cost function, many learning strategies have been proposed, which can be found in a recent comprehensive survey paper [20]. The most popular method to solve the optimization problem of Eq. 5 is the multiplicative rule [14], [15]. The algorithm starts from two positive matrices, and multiplies the elements of $\mathbf{W}$ and $\mathbf{H}$ by some positive factors within each iteration. It has been proved that under the multiplicative rule, the distance $\|\mathbf{Y}-\mathbf{WH}\|_F^2$ is monotonically non-increasing [15].

Aiming at speeding up the convergence of the original iterative algorithm, a projected gradient bound-constrained optimization method is adopted in [21]. Although several previous papers [22], [23] have used this technique, they were lack of systematic study and comprehensive comparison with the standard multiplicative update rule. The algorithm in [21] is demonstrated to be computationally simpler and converge faster than the standard learning rules [15].

One hurdle of the NMF problem is the existence of local minima due to the non-convexity of the objective function. It is apparent that for any non-negative invertible matrix pair $\mathbf{D}$ and $\mathbf{D}^{-1}$, the equality $\mathbf{WH} = (\mathbf{WD})(\mathbf{D^{-1}H})$ holds. Thus, the solution highly depends on the initialization of specific learning strategies. The seeding algorithms using the spherical K-Means [24] and the Fuzzy C-Means (FCM) [17] have been demonstrated to present better performance than random initializations. The theoretical study of conditions resulting in unique solution based on positive cone geometry is investigated in [25]. For many applications, the non-uniqueness of solution can be alleviated by adding additional constraints to confine the feasible solution set, such as the smoothness constraint [18] for better spectral signatures, the sparseness constraint [22] and the non-smoothness constraint [26] for part-based structures. It is apparent that the imposed constraints are application-dependent.

C. Geometric Interpretation of NMF and Spectral Unmixing

The above algebraic analysis provides a theoretical interpretation on why the NMF technique could be used to decompose mixed pixels. In this section, we present a geometric interpretation of NMF and spectral unmixing to *visualize* the solution space for both problems. For the multivariate data, such as hyperspectral images, each pixel can be thought of as a point in an l-dimensional space, whose l coordinates are given by the reflectance values at different spectral bands. The factorizations in both Eq. 2 and Eq. 3 reveal that, except for the original Euclidean coordinate system $\{\mathbf{e}_j\}_{j=1}^l \in \mathbf{R}^l$ (only one element of $\mathbf{e}_j$ is 1, all the others are zeros), there exist other sets of basis vectors $\{\mathbf{v}_j\}_{j=1}^c$, $\mathbf{v}_j \succeq \mathbf{0}$ in a subspace $\mathbf{R}^c$, $c < l$, such that all the image points can be approximated as linear combinations of these bases. Due to the inherent constraints involved in pixel unmixing and NMF, these vectors possess different geometric characteristics.

As analyzed in [25], for the problem of non-negative matrix factorization, all data points lie in a positive simplical cone represented by

$$\mathcal{C} = \{\mathbf{x} \mid \mathbf{x} = \sum_j \theta_j\mathbf{v}_j, \boldsymbol{\theta} \succeq \mathbf{0}\} \tag{7}$$

However, the extra abundance sum-to-one constraint in the unmixing model confines the data points to reside within a simplex,

$$\mathcal{S} = \{\mathbf{x} \mid \mathbf{x} = \sum_j \theta_j\mathbf{v}_j, \boldsymbol{\theta} \succeq \mathbf{0}, \mathbf{1}^T\boldsymbol{\theta} = 1\} \tag{8}$$

where $\boldsymbol{\theta}$ is a column vector with its components being the weight of the basis vectors. Fig. 1 illustrates one of the possible cones and simplexes for the given data denoted. It is apparent that they are not unique. For example, you can draw a number of triangles with different vertices to circumscribe the same set of data points. Theoretically, there exist an infinite number of cones and simplexes containing the data cloud, since if $\mathcal{S}$ is a simplex circumscribing the data, for any other simplex $\bar{\mathcal{S}}$ such

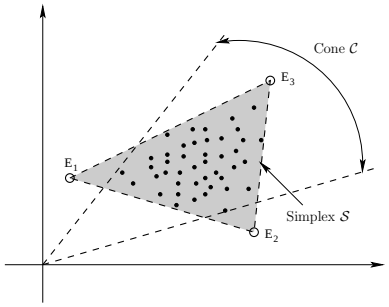


Fig. 1. Geometric illustration of possible cones and simplexes that circumscribe the given data denoted by the dots. The circles $E_1 \sim E_3$ can be considered as possible endmembers.

that $\mathcal{S} \subset \bar{\mathcal{S}}$, $\bar{\mathcal{S}}$ will also contain the data. Then, one immediate and critical question would be what is the criterion for the best simplex solution or what properties the best simplex should possess.

In the convex geometry-based endmember extraction algorithms, the best simplex is either defined as the one that circumscribes the data cloud and at the same time has the minimum volume [8], or defined as the one that inscribes the data cloud with the maximum volume [10]. It is apparent that under the pure pixel assumption [9]–[11], the best simplex is uniquely determined by the pure pixels, which are the vertices of the simplex. The goal of endmember detection is thus to identify the extreme-valued pixels from the given image. However, if all the image pixels are highly mixed, the identified extreme points would not be the true endmembers, even though they are on the boundary of the data cloud and close to the real ones. In order to find the true endmember locations from the highly mixed data, we have to extend the searching space outside the given data cloud. In the meanwhile, we should keep the simplex that circumscribes the data as compact as possible. These observations trigger the development of the proposed algorithm by incorporating the minimum volume constraint into the NMF technique.

III. NMF with Minimum Volume Constraint

In this section, we describe the proposed minimum volume constrained NMF algorithm. To keep the symbols consistent during the algorithm description, we will use the model in Eq. 2 in the rest of the paper.

A. Problem Formulation

Combining the goal of minimum approximation error with the volume constraint, we arrive at the following constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{A},\mathbf{S}) = \frac{1}{2}\|\mathbf{X}-\mathbf{A}\mathbf{S}\|_F^2 + \lambda J(\mathbf{A}) \\ \text{subject to} \quad & \mathbf{A} \succeq \mathbf{0},\ \mathbf{S} \succeq \mathbf{0},\ \mathbf{1}_c^T\mathbf{S} = \mathbf{1}_N^T \end{aligned} \tag{9}$$

where $\mathbf{1}_c$ ($\mathbf{1}_N$) is a c (N)-dimensional column vector of all 1's, and $J(\mathbf{A})$ is the penalty function, calculating the simplex volume determined by the estimated endmembers. The regularization parameter $\lambda \in \mathbf{R}$ is used to control the tradeoff between the accurate reconstruction and the volume constraint. The first term serves as the external force to drive the search to move outward, so that the generated simplex contains all data points with relatively small errors. The second term serves as the internal force, which constrains the simplex volume to be small. A solution is found when these two forces balance each other. One of the expected advantages of the volume constrained NMF would be the resistance to the presence of noise. The noise normally results in a bigger data cloud, which thereby leads to a simplex with larger volume than the clean data, and the simplex vertices deviate from the true endmember locations. By adding the volume constraint, the simplex can be made not to include every data point, particularly, the noisy pixels on the boundary. Therefore, the algorithm would be more robust to the noise effect than the unconstrained NMF approaches.

1) Volume Determination: In order to calculate the volume determined by a set of points, we hereby resort to the connection between the volume and the determinant [27]. Suppose we have c affinely independent points $\mathbf{a}_1,\ldots,\mathbf{a}_c \in \mathbf{R}^{c-1}$, which means that the $c-1$ vectors $\mathbf{a}_2-\mathbf{a}_1,\ldots,\mathbf{a}_c-\mathbf{a}_1$ are linearly independent. The volume determined by these points is calculated by

$$V = \frac{1}{(c-1)!}\left|\det\left([\mathbf{a}_2-\mathbf{a}_1 \quad \ldots \quad \mathbf{a}_c-\mathbf{a}_1]\right)\right| \tag{10}$$

or equivalently,

$$V = \frac{1}{(c-1)!}\left|\det\left(\begin{bmatrix} 1 & \cdots & 1 \\ \mathbf{a}_1 & \cdots & \mathbf{a}_c \end{bmatrix}\right)\right| \tag{11}$$

Problem occurs when these c points are in an l-space, i.e., $\mathbf{a}_1,\ldots,\mathbf{a}_c \in \mathbf{R}^l$ ($c \le l$), as the determinant is not defined for a non-square matrix. In this case, we make an approximation by calculating the volume formed by a new set of points $\tilde{\mathbf{a}}_1,\ldots,\tilde{\mathbf{a}}_c \in \mathbf{R}^{c-1}$, which is the low dimensional transform of the original data; that is, we adopt PCA to reduce the dimensionality of points $\mathbf{a}_1,\ldots,\mathbf{a}_c$ from l to $c-1$ by keeping only the $c-1$ most significant principal components.

It is true that such volume determination can rarely return the actual volume formed by the endmembers. But recall that the real goal of this volume term is to constrain the simplex size as compact as possible, provided that the vertices can approximate the given data set at a certain error tolerance. In this sense, the computed volume is not necessarily the actual value, as long as it approximates the real one and performs as an internal force. The effectiveness of this constraint will be validated by a set of experiments reported in Section IV.

2) Cost Function: For the problem of Eq. 9, the simplex volume is determined by the c endmembers, each of which is an l-dimensional vector (i.e., the columns of $\mathbf{A}$). Then, $J(\mathbf{A})$ is formulated as

$$J(\mathbf{A}) = \frac{1}{2(c-1)!}\det^2\left(\begin{bmatrix} \mathbf{1}_c^T \\ \tilde{\mathbf{A}} \end{bmatrix}\right) \tag{12}$$

where the matrix $\tilde{\mathbf{A}} = (\tilde{\mathbf{a}}_1,\tilde{\mathbf{a}}_2,\ldots,\tilde{\mathbf{a}}_c) \in \mathbf{R}^{(c-1)\times c}$ is a low dimensional transform of $\mathbf{A}$ given by

$$\tilde{\mathbf{A}} = \mathbf{U}^T(\mathbf{A}-\boldsymbol{\mu}\mathbf{1}_c^T) \tag{13}$$

The matrix $\mathbf{U} \in \mathbf{R}^{l\times(c-1)}$ is formed by the $c-1$ most significant principal components (PCs) of data $\mathbf{X}$ through principal component analysis. The column vector $\boldsymbol{\mu}$ is the data mean. Note that both the PCs and the mean vector are calculated from $\mathbf{X}$ instead of $\mathbf{A}$. One reason we formulate it in this way is that the PCs of the given data points reflect important shape information of the true simplex, thus the projection onto this subspace reveals the similarity between the estimated and the true simplex. In addition, from a computational efficiency point of view, the PCs of $\mathbf{X}$ can be obtained through one time calculation, while the PCs of the source matrix $\mathbf{A}$ need to be recalculated constantly as $\mathbf{A}$ changes.

To formulate the penalty term as a function of $\mathbf{A}$, we make the following transformation. Let

$$\mathbf{Z} = \begin{bmatrix} \mathbf{1}_c^T \\ \tilde{\mathbf{A}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_c^T \\ \underline{\mathbf{0}} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{c-1}^T \\ \underline{\mathbf{I}} \end{bmatrix} \tilde{\mathbf{A}} \tag{14}$$

where $\underline{\mathbf{0}}$ is a $(c-1)\times c$ zero matrix, $\underline{\mathbf{I}}$ is a $(c-1)\times(c-1)$ identity matrix, and $\mathbf{0}_{c-1}$ a $(c-1)$-dimensional zero column vector. The corresponding matrices are denoted by

$$\mathbf{C} = \begin{bmatrix} \mathbf{1}_c^T \\ \underline{\mathbf{0}} \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} \mathbf{0}_{c-1}^T \\ \underline{\mathbf{I}} \end{bmatrix} \tag{15}$$

We then have

$$\mathbf{Z} = \mathbf{C} + \mathbf{B}\mathbf{U}^T(\mathbf{A} - \boldsymbol{\mu}\mathbf{1}_c^T) \tag{16}$$

Substituting the above equation into Eq. 12, and then Eq. 9, we finalize the objective function of the proposed MVC-NMF as

$$f(\mathbf{A},\mathbf{S}) = \frac{1}{2}\|\mathbf{X}-\mathbf{A}\mathbf{S}\|_F^2 + \frac{\tau}{2}\det^2(\mathbf{C}+\mathbf{B}\mathbf{U}^T(\mathbf{A}-\boldsymbol{\mu}\mathbf{1}_c^T)) \tag{17}$$

with $\tau = \frac{\lambda}{(c-1)!}$, and the matrices $\mathbf{C}, \mathbf{B}, \mathbf{U}$, and the vector $\boldsymbol{\mu}$ are constants for the given data $\mathbf{X}$.

B. Optimization Algorithm Description

1) Initialization: To solve the formulated optimization problem, the first essential step is to determine the number of endmembers involved in the mixture data, which is always a challenge because it is closely related to the unknown noise. In the current research, we resort to the virtual dimensionality (VD) method [28]. The algorithm aims at finding the minimum number of spectrally distinct signal sources, which characterize the hyperspectral data from the target detection and classification point of view. Based on the Neyman-Pearson detection theory, a hypothesis test is applied to pairs of eigenvalues of the sample covariance matrix and the sample correlation matrix. The key is that when no signal is present in a particular component, the corresponding correlation eigenvalue and covariance eigenvalue should be equal; otherwise, the covariance eigenvalue is less than the corresponding correlation eigenvalue contributed by the correlated signals. To remove the second-order statistical correlation, a noise whitening process is incorporated as a preprocessing step. Although the VD method is effective, it is not guaranteed that the estimation is 100% accurate. Therefore, the sensitivity analysis of the proposed method to the estimated number of endmembers is a very important issue. We will perform this analysis in Section IV.

The next question is how to initialize matrices $\mathbf{A}$ and $\mathbf{S}$. As mentioned earlier, the unconstrained NMF algorithms are sensitive to initializations due to the existence of local optima. The algorithm performance highly depends on the distance between the initial point and the global optimum. In this paper, we will show that by incorporating the proposed volume constraint, this problem is successfully mitigated. We randomly choose c points from the given data and arrange them as the columns of the initial $\mathbf{A}$. The value of $\mathbf{S}$ can also be randomly initialized. In the experiments described in this paper, a zero matrix $\mathbf{S}$ is used to start the learning.

2) Learning Rules: Minimizing the objective function in Eq. 17 with respect to both $\mathbf{A}$ and $\mathbf{S}$ is a combinatorial optimization problem. We here resort to the *alternating non-negative least squares* method [18], [21], [23]. This technique treats the original optimization problem as two sub-problems with the following iterative update rule,

$$\begin{aligned} \mathbf{A}^{k+1} &= \arg\min_{\mathbf{A}} f(\mathbf{A},\mathbf{S}^k) \le f(\mathbf{A}^k,\mathbf{S}^k) \\ \mathbf{S}^{k+1} &= \arg\min_{\mathbf{S}} f(\mathbf{A}^{k+1},\mathbf{S}) \le f(\mathbf{A}^{k+1},\mathbf{S}^k) \end{aligned} \tag{18}$$

that is, we alternatively update one matrix, holding the other one fixed.

For each sub-problem, the projected gradient learning [18], [21], [22] is adopted to impose the non-negative constraint. The method of projected gradient follows the standard gradient learning to update. When the new estimate does not satisfy the constraint, a projective function is used to project the point back to the feasible set. For the non-negative constraint, we use a simple but effective function $\max(0,x)$ to set the negative components to zero and keep the non-negative components unchanged. Then, the update rule is expressed as

$$\begin{aligned} \mathbf{A}^{k+1} &= \max\left(\mathbf{0},\ \ \mathbf{A}^k - \alpha^k \nabla_{\mathbf{A}} f(\mathbf{A}^k,\mathbf{S}^k)\right) \\ \mathbf{S}^{k+1} &= \max\left(\mathbf{0},\ \ \mathbf{S}^k - \beta^k \nabla_{\mathbf{S}} f(\mathbf{A}^{k+1},\mathbf{S}^k)\right) \end{aligned} \tag{19}$$

where the parameters α^k and β^k are the small learning rates (stepsizes) selected based on the Armijo rule [29].

Let α_o be the initial stepsize, $\sigma \in (0, \frac{1}{2})$ be the tolerance, and $\rho \in (0,1)$ the scaling factor that the stepsize is reduced in very iteration. Then, the Armijo rule chooses the stepsize as $\alpha^k = \rho^{m_k}\alpha_o$, where m_k is the first integer such that

$$f(\mathbf{A}^{k+1},\mathbf{S}^k) - f(\mathbf{A}^k,\mathbf{S}^k) \le \sigma\rho^{m_k}\alpha_o \nabla f(\mathbf{A}^k,\mathbf{S}^k)^T(\mathbf{A}^{k+1}-\mathbf{A}^k)$$

The stepsize β^k is selected based on the same procedure.

Another important issue in the formulated problem is the abundance sum-to-one constraint. This is an equality constraint and therefore can be dealt with using the Lagrange multipliers method with closed form solutions. In this paper, we adopt a much simpler but effective method as in [30]. We augment the data matrix $\mathbf{X}$ and the material signature matrix $\mathbf{A}$ by a row of constant denoted by

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \delta\mathbf{1}_N^T \end{bmatrix}, \ \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \delta\mathbf{1}_c^T \end{bmatrix} \tag{20}$$

where δ is a positive number to control the effect of the sum-to-one constraint. The learning of abundance $\mathbf{S}$ described in

Eq. 19 takes these two augmented matrices as inputs. As δ increases, the columns of $\mathbf{S}$ are forced to approach the sum-to-one constraint. In order to balance between the estimation accuracy and the convergence rate, we select a relatively small δ, typically $10 \sim 20$, which results in abundances with the summation varying between 0.998 and 1.002.

To implement the gradient learning, the first order derivative $\nabla_{\mathbf{A}} f(\mathbf{A}, \mathbf{S})$ and $\nabla_{\mathbf{S}} f(\mathbf{A}, \mathbf{S})$ need to be calculated. Since the penalty function is independent of $\mathbf{S}$, the partial derivative $\nabla_{\mathbf{S}} f(\mathbf{A}, \mathbf{S})$ is easy to find,

$$\nabla_{\mathbf{S}} f(\mathbf{A}, \mathbf{S}) = \mathbf{A}^T(\mathbf{AS} - \mathbf{X}) \tag{21}$$

However, more algebraic operations need to be performed to calculate $\nabla_{\mathbf{A}} f(\mathbf{A}, \mathbf{S})$. Taking the partial derivative of Eq. 17 with respect to $\mathbf{A}$, we first get

$$\nabla_{\mathbf{A}} f(\mathbf{A}, \mathbf{S}) = (\mathbf{AS} - \mathbf{X})\mathbf{S}^T + \tau \det(\mathbf{Z}) \frac{\partial \det(\mathbf{Z})}{\partial \mathbf{A}} \tag{22}$$

where $\mathbf{Z}$ is defined in Eq. 16. It is known that the partial derivative of a scalar with respect to a matrix is still a matrix,

$$\frac{\partial \det(\mathbf{Z})}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial \det(\mathbf{Z})}{\partial a_{11}} & \frac{\partial \det(\mathbf{Z})}{\partial a_{12}} & \cdots & \frac{\partial \det(\mathbf{Z})}{\partial a_{1c}} \\ \frac{\partial \det(\mathbf{Z})}{\partial a_{21}} & \frac{\partial \det(\mathbf{Z})}{\partial a_{22}} & \cdots & \frac{\partial \det(\mathbf{Z})}{\partial a_{2c}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \det(\mathbf{Z})}{\partial a_{l1}} & \frac{\partial \det(\mathbf{Z})}{\partial a_{l2}} & \cdots & \frac{\partial \det(\mathbf{Z})}{\partial a_{lc}} \end{bmatrix} \tag{23}$$

Then, the problem is tailored to compute each element of $\frac{\partial \det(\mathbf{Z})}{\partial \mathbf{A}}$, which is given by

$$\frac{\partial \det(\mathbf{Z})}{\partial a_{ij}} = \det(\mathbf{Z}) Tr\left(\mathbf{Z}^{-1} \frac{\partial \mathbf{Z}}{\partial a_{ij}}\right) \tag{24}$$

It is easy to derive that

$$\frac{\partial \mathbf{Z}}{\partial a_{ij}} = \frac{\partial \mathbf{B}\mathbf{U}^T\mathbf{A}}{\partial a_{ij}} \tag{25}$$

Let $\mathbf{D} = \mathbf{B}\mathbf{U}^T$ and $\bar{\mathbf{Z}} = \mathbf{Z}^{-1}$, through algebraic manipulations, we arrive at

$$Tr\left(\mathbf{Z}^{-1} \frac{\partial \mathbf{Z}}{\partial a_{ij}}\right) = \bar{\mathbf{z}}_j^T \mathbf{d}_i \tag{26}$$

where $\bar{\mathbf{z}}_j^T$ is the jth row of $\bar{\mathbf{Z}}$, and $\mathbf{d}_i$ is the ith column of $\mathbf{D}$. Combining Eqs. 23, 24, and 26 gives

$$\frac{\partial \det(\mathbf{Z})}{\partial \mathbf{A}} = \det(\mathbf{Z})(\bar{\mathbf{Z}}\mathbf{E})^T = \det(\mathbf{Z})\mathbf{U}\mathbf{B}^T(\mathbf{Z}^{-1})^T \tag{27}$$

Substitute the above equation into Eq. 22, we obtain the gradient

$$\nabla_{\mathbf{A}} f(\mathbf{A}, \mathbf{S}) = (\mathbf{AS} - \mathbf{X})\mathbf{S}^T + \tau \det^2(\mathbf{Z})\mathbf{U}\mathbf{B}^T(\mathbf{Z}^{-1})^T \tag{28}$$

3) Stopping Conditions: For the NMF learning, different stopping criteria have been exploited. The most commonly used are the maximum iteration number and the error tolerance. Several algorithms [21], [23] choose to stop the learning process whenever the Euclidean norm of the gradient of the objective function is less than a threshold. Some implementations [22] use an infinite loop, which needs user interaction to interrupt when some specific requirements are satisfied. For the proposed minimization problem, we expect the objective function to monotonically decrease. However, in order to avoid being trapped in local minima, we also allow certain steps of increase of the objective value. When the number of successive increasing steps is over a predefined value, the learning is stopped. In addition, we include an iteration limit specified by the prescribed maximum iteration number.

The proposed MVC-NMF approach is summarized in Algorithm 1.

Algorithm 1: Minimum volume constrained NMF

Data : Non-negative mixture data $\mathbf{X} \in \mathbf{R}^{l \times N}$ with each column being an observation vector.

Result: Two non-negative matrices $\mathbf{A} \in \mathbf{R}^{l \times c}$ and $\mathbf{S} \in \mathbf{R}^{c \times N}$ with sum-to-one constraint $\mathbf{1}_c^T \mathbf{S} = \mathbf{1}_N^T$.

//Initialization
Set $\mathbf{S} = \underline{\mathbf{0}}$, i.e., a zero matrix;
Set iteration index $k = 0$;
Estimate the number of endmembers c;
Initialize $\mathbf{A}$ by randomly selecting c data points from $\mathbf{X}$ to form its columns;
//main loop
while *stop conditions are not met* **do**
  $\mathbf{A}^{k+1} = \max(\mathbf{0},\ \mathbf{A}^k - \alpha^k \nabla_{\mathbf{A}} f(\mathbf{A}^k, \mathbf{S}^k))$;
  Augment $\mathbf{A}^{k+1}$ to get $\tilde{\mathbf{A}}^{k+1}$;
  $\mathbf{S}^{k+1} = \max(\mathbf{0},\ \mathbf{S}^k - \beta^k \nabla_{\mathbf{S}} f(\tilde{\mathbf{A}}^{k+1}, \mathbf{S}^k))$;
  Increase k by 1.
end

IV. Evaluation with Synthetic Images

In this section, we conduct a series of experiments to demonstrate the effectiveness of the proposed MVC-NMF using synthetic images. Note that, apart from the extracted endmembers, MVC-NMF deduces the fractional abundance at each pixel as a by-product. In the following, we will evaluate both the estimated endmembers and the inferred abundances. We compare with three existing approaches, namely VCA [5], NMF with smoothness constraint (SCNMF) [18], and the projected gradient NMF method (PGNMF) [21], among which VCA is a deterministic method, and the other two are both NMF-based statistical approaches.

Regarding VCA, since it only identifies endmembers, we use a fully constrained least squares method, called FCLS [30], to find the abundances. The objective of FCLS is to find the abundance vector by minimizing the least squares error $\|\mathbf{x} - \mathbf{As}\|^2$. The algorithm integrates the constraint $\mathbf{1}^T\mathbf{s} = 1$ into the objective function by augmenting both the matrix $\mathbf{A}$ with a row vector of all 1's and the observation $\mathbf{x}$ with an element of 1. In this way, the original problem is converted to the standard non-negative constrained least squares, which is solved using a standard active set method [31]. In the following, the combined process of VCA and FCLS is referred to as VCA-FCLS.

For the two NMF approaches SCNMF [21] and PGNMF [18], the sum-to-one constraint is incorporated by the same augmenting technique as in Eq. 20. In addition, the last two

steps of SCNMF, i.e., finding the best match from a signature database, and calculating the abundance matrix using a least squares method, are eliminated for fair comparison.

A. Performance Metrics

The most widely used metric to measure the shape similarity between the true endmember signature $\mathbf{a}$ and its estimate $\hat{\mathbf{a}}$ is the spectral angle distance (SAD), which is a high dimensional extension to the two-dimensional geometric angle expressed as

$$SAD = \cos^{-1}\left(\frac{\mathbf{a}^T\hat{\mathbf{a}}}{\|\mathbf{a}\|\|\hat{\mathbf{a}}\|}\right) \tag{29}$$

Another metric uses an information theoretic measure, spectral information divergence (SID) [32]. The probability distribution vector associated with each endmember signature is given by $\mathbf{p} = \frac{\mathbf{a}}{\sum_j a_j}$. This vector can be used to describe the variability of the spectral signature. Let $\hat{\mathbf{p}}$ denote the probability distribution vector of the estimate $\hat{\mathbf{a}}$. Then, the similarity between $\mathbf{a}$ and $\hat{\mathbf{a}}$ can be measured by the relative entropy

$$D(\mathbf{a}|\hat{\mathbf{a}}) = \sum_j p_j \log(\frac{p_j}{\hat{p}_j}) \tag{30}$$

Since the relative entropy is not symmetric, the following measure is used

$$SID = D(\mathbf{a}|\hat{\mathbf{a}}) + D(\hat{\mathbf{a}}|\mathbf{a}) \tag{31}$$

Regarding the abundance estimation, we use similar metrics by substituting the endmember signature with the abundance vector of each individual pixel, which gives the following two metrics

$$AAD = \cos^{-1}\left(\frac{\mathbf{s}^T\hat{\mathbf{s}}}{\|\mathbf{s}\|\|\hat{\mathbf{s}}\|}\right) \tag{32}$$

and

$$AID = D(\mathbf{s}|\hat{\mathbf{s}}) + D(\hat{\mathbf{s}}|\mathbf{s}) \tag{33}$$

where $\mathbf{s}$ represents the true abundance vector of a single pixel and $\hat{\mathbf{s}}$ the corresponding estimate. AAD and AID refer to the abundance angle distance and the abundance information divergence, respectively.

B. Creation of Synthetic Images

The synthetic images of size 64×64 are generated using a set of spectral reflectances selected from the USGS digital spectral library [33]. The selection of endmember signatures is arbitrary as long as they are linearly independent. The spectral data contain 224 spectral bands covering wavelengths from 0.38μm to 2.5μm with a spectral resolution of 10nm. To create linear mixtures, we divide the entire image into units of 8×8 small blocks. The pixels within each block are pure and have the same type of ground cover, randomly selected as one of the endmember classes. The resulting image is then degraded by a spatial low pass filter to simulate an image with mixed pixels. With this mixing method, we intend to simulate a hyperspectral scene with endmembers arranged in discrete patches so that a linear mixture model would be appropriate. The low pass filter we used is a simple $k \times k$ averaging filter. Apparently, the value of k controls the degree of mixing. In this paper, we use a 9×9 filter to generate highly mixed data. To further remove pure pixels, we replace all the pixels whose abundance is larger than 80% with a mixture made up of all endmembers of equal abundances; that is, each endmember has an abundance of $1/c$ in the mixture.

To simulate possible errors and sensor noise, we add zero mean Gaussian noise to the mixture data. Assuming the noise is both spatially and spectrally uncorrelated, then the noise covariance matrix is a diagonal matrix $\sigma^2\mathbf{I}$ with $\mathbf{I}$ being an identity matrix. Define the SNR as $SNR = 10\log_{10}\frac{E[\mathbf{x}^T\mathbf{x}]}{E[\mathbf{n}^T\mathbf{n}]}$, then the noise variance is easy to determine for a particular value of SNR, i.e., $\sigma^2 = E[\mathbf{x}^T\mathbf{x}]/(10^{SNR/10}l)$.

C. Algorithm Evaluation

The proposed MVC-NMF algorithm is evaluated through six experiments. The first experiment aims to conduct an overall comparison of different methods, including an illustration of the approximation error and the effect of the introduced volume constraint. In the second experiment, we study the performance dependence of the NMF-based approaches on different methods of initializations. The third experiment is designed to investigate the algorithm robustness to noise corruptions. In the fourth experiment, we perform sensitivity analysis of the proposed method to the accurate knowledge of the number of endmembers. We then vary the number of endmembers to create mixtures in the fifth experiment, aiming at investigating the algorithm generalizations to data sets with relatively large number of endmembers. In the last experiment, we study the algorithm performance when the given data contain pure endmember representations. In all the experiments that follow, the quantitative measures are obtained by averaging 20 random tests. Four endmembers are used to create the synthetic mixtures for rapid computation in all experiments, except when we investigate the algorithm sensitivity to the estimated number of endmembers, and the generalization analysis to data sets with comparatively large numbers of endmembers. All the NMF-based methods are initialized randomly, except when we study the performance dependence on endmember initializations.

1) Overall comparison of different methods: In this experiment, we add 20dB white Gaussian noise to the synthetic mixtures and perform decomposition using different methods. The stopping criteria of PGNMF and MVC-NMF are the same; that is, the algorithms are terminated when the iteration number of successive increase of the objective value is greater than 5, or the maximum iteration number, 150, is reached. The experimental results show that these two learning processes rarely use the maximum allowable number of iterations. Considering the low convergence of SCNMF, we allow the maximum number of iteration to be 300. In all the tests, the SCNMF learning ends up with very large approximation errors at the 300th iteration, when the learning reaches its stable state. The regularization parameter of MVC-NMF is selected as $\tau = 0.01$. The parameters in all the other methods implemented in this paper follow their original work.

Fig. 2 demonstrates the experimental results, in which we show both the mean (bar) and the standard deviation (error

bar). The SID value generated by MVC-NMF (Fig 2(d)) is too small to tell, we thus label the result (mean/standard deviation) aside its bar plot. As can be seen, MVC-NMF produces the smallest means and standard deviations in terms of all measurement metrics used, indicating that this approach generates the most accurate and stable estimates. In addition, VCA-FCLS outperforms SCNMF and PGNMF. In terms of abundance estimation (measured by AAD and AID), SCNMF produces the largest means although with relatively smaller deviations than PGNMF. Regarding the endmember extraction (measured by SAD and SID), it can be seen that the two measures present inconsistent comparisons between SCNMF and PGNMF. Overall, SCNMF yields smaller deviations than PGNMF because of the use of the smoothness constraint, which tends to stabilize the endmember solutions.

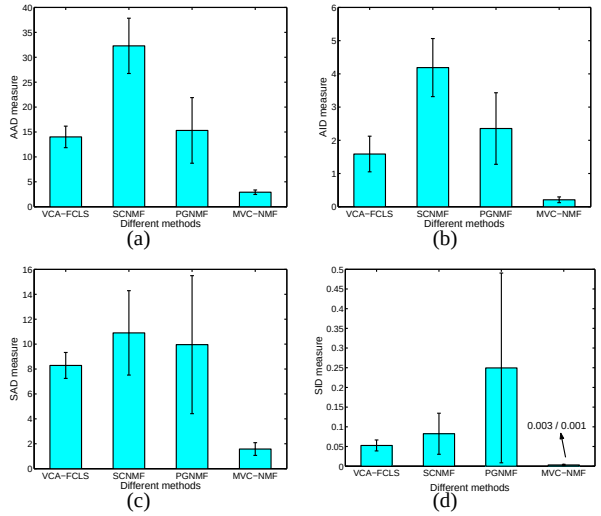


Fig. 2. Performance comparison in terms of different metrics (SNR=20dB) (a) AAD (b) AID (c) SAD (d) SID.

We next fix the iteration number to 150 for the NMF-based methods and study the effect of the constrained volume as well as the approximation error generated. In Fig. 3(a), the two constant profiles correspond to the volume derived by VCA and that calculated using the real endmember set, respectively. As discussed before, VCA is deterministic and the endmembers are chosen as the extreme points of the given data set. Therefore, the volume determined by the VCA estimates is smaller than that of the true endmembers. The other three curves correspond to the results of the NMF-based algorithms. It can be seen that during the first few iterations, both the PGNMF and the MVC-NMF learning expand the simplex rapidly. After that, the introduced volume constraint in MVC-NMF effectively confines the simplex volume to be close to the true value. However, the PGNMF learning keeps increasing the simplex size, which is much larger than the actual volume. And, the volume determined by SCNMF grows very slowly as the learning goes. The approximation error $\frac{1}{2}\|\mathbf{X}-\mathbf{AS}\|_F^2$ is plotted in Fig. 3(b), where we only compare the VCA-FCLS, the PGNMF, and the MVC-NMF learning, as the SCNMF learning error is comparatively much larger even after convergence (more than 5 times that of the VCA-FCLS result). Again, the constant profile illustrates the approximation error of VCA-FCLS. Note that during the first 60 iterations, PGNMF yields smaller approximation errors than MVC-NMF and quickly converges to the stable state. After that, MVC-NMF continues to reduce the approximation error and converges to a state with smaller error than PGNMF. This reveals that the PGNMF learning could be trapped in a local minimum.

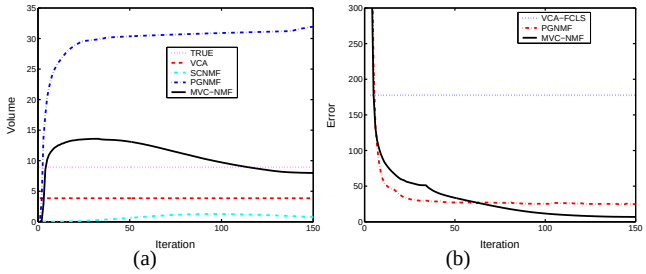


Fig. 3. Comparison of the extracted endmembers and the estimation accuracy in terms of (a) Simplex volume (b) Approximation error.

Based on the above comparisons, we observe that the general smoothness constraint adopted in SCNMF does not perform as well as the proposed volume constraint. In addition, SCNMF demonstrates slow convergence rate and much larger approximation error. Therefore, in the following experiments, we exclude SCNMF for further comparisons and focus on the comparative analysis between the deterministic (VCA-FCLS) and the statistical (PGNMF and MVC-NMF) methods, and between the unconstrained (PGNMF) and the constrained (MVC-NMF) NMF approaches.

2) Dependence analysis on endmember initializations: The purpose of this experiment is to show how different initialization methods could affect the NMF learning process. We study two methods for the endmember initialization, i.e., random initialization and the use of the endmembers identified by VCA. The experimental results are shown in Fig. 4, where we display the deterministic results of VCA-FCLS, and the measures of PGNMF and MVC-NMF with different initialization schemes. One immediate observation we made is that no matter which initialization methods are used, the proposed MVC-NMF always generates the smallest means and standard deviations. In addition, PGNMF outperforms VCA-FCLS when initialized with the endmembers estimated by VCA, which, however, is not the case when using random initializations. Another important observation is that MVC-NMF is less sensitive to the selection of the initial points compared to PGNMF. This observation leads to the conclusion that the introduced volume constraint effectively confines the solution space and converts the original ill-posed problem to a well-posed one.

3) Robustness analysis to noise corruptions: In this experiment, a comparative analysis on the issue of algorithm sensitivity to noise is studied by simulating synthetic data with different noise levels. The SNR is varied to be 15dB, 25dB, 45dB, and infinity (no noise is added). From the previous results, we see that the angle distance and the information divergence metrics give similar comparison trend. It has also been shown that the information divergence measure is more

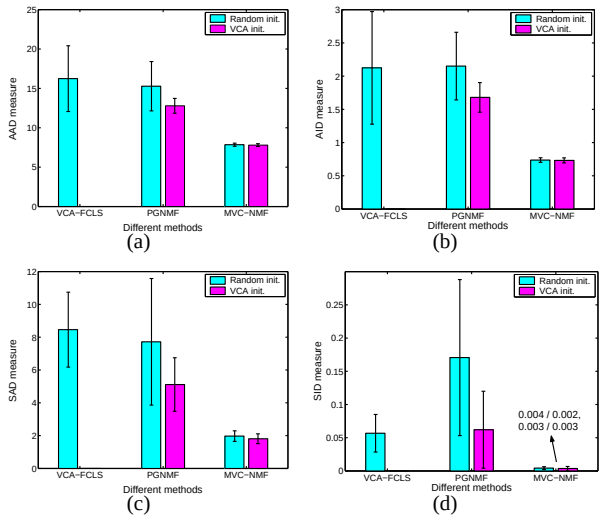


Fig. 4. Performance comparison when PGNMF and MVC-NMF take random and VCA initializations (SNR=10dB) (a) AAD (b) AID (c) SAD (d) SID.

effective in characterizing the spectral features [34]. Thus, in the following experiments, we omit the AAD and SAD results. Fig. 5 compares the performance of different methods at various noise levels. We observe, again, that the proposed MVC-NMF outperforms the others for all SNR cases. As the SNR increases, both VCA-FCLS and MVC-NMF show improved performance. However, it is not the case for PGNMF, which keeps generating bad estimates no matter how clean the data are. The reason is related to the ill-posedness of PGNMF, which is the dominating factor that affects its performance.

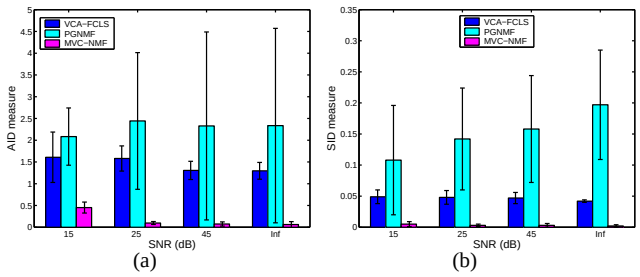


Fig. 5. Performance comparison at different noise levels in terms of (a) AID (b) SID.

4) Sensitivity analysis to the estimated number of endmembers: For a given data set, the number of endmembers cannot be estimated with 100% accuracy, especially when some of the endmembers are less prevalent and only present in a small set of pixels. Therefore, it is of importance to analyze the algorithm sensitivity to the estimated number of endmembers. For this purpose, we create a synthetic image with 6 endmembers and unmix the generated mixture using different numbers of endmembers, varied from 3 to 9. Denote the actual number of endmembers by c ($c = 6$) and the estimated value by $\hat{c}$ ($\hat{c} = 3, 4, \ldots, 9$), we compare the extracted endmembers with the corresponding closest real endmembers when $\hat{c} < c$. For the situation of $\hat{c} > c$, we choose the c estimates which mostly resemble the real endmembers and calculate the differences in between. The abundances of the selected endmembers are then used to compute AID. Fig. 6 illustrates the unmixing performance with different $\hat{c}$ values. The abundance estimation results shown in Fig. 6(a) demonstrate that the best performance of different methods occur when $\hat{c} = c$. As $\hat{c}$ deviates from c, the performances of VCA-FCLS and PGNMF decrease accordingly. For the proposed MVC-NMF, the same trend can be observed when $\hat{c} < c$. However, when $\hat{c} > c$, MVC-NMF produces steady estimates regardless of the value of $\hat{c}$. From the endmember extraction point of view, VCA-FCLS renders comparably stable performance as illustrated in Fig. 6(b). This outcome is inherently related to the essence of the VCA scheme, which finds the extreme-valued pixels based on geometrical projections. Thus, the identified extreme points solely depend on the data distribution. On the other hand, the proposed MVC-NMF aims at finding a convex hull to enclose the data set. The shape of the convex hull is not only dependent on the data distribution, but also affected by the estimated number of endmembers, which determines the number of extreme points of the convex hull. This dependence is much obvious when $\hat{c} < c$ as illustrated by the worse performance of the NMF estimations in Fig. 6(b). When $\hat{c} > c$, the NMF-based methods yield better results compared with the cases when $\hat{c} < c$. This phenomenon can be explained by inspecting the endmember profiles. We have observed that among the extracted $\hat{c}$ endmembers, $\hat{c} - c$ estimates are very similar with the others, indicating that the true dimension of the estimated endmembers is equal to c. Therefore, the increased $\hat{c}$ does not change the shape of the reconstructed convex hull very much, which also explains why MVC-NMF generates steady abundance estimations when $\hat{c} > c$. This example shows that the success of MVC-NMF relies on the correct knowledge of c.

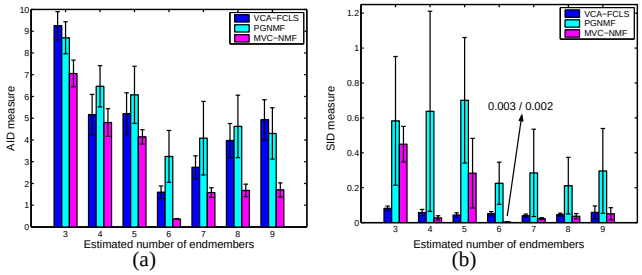
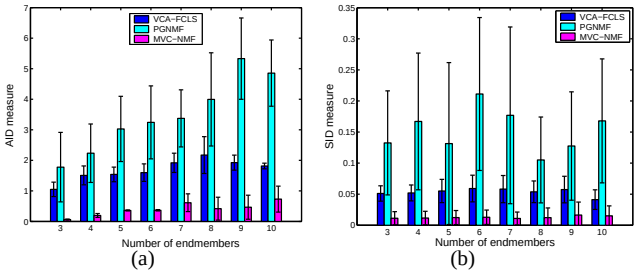


Fig. 6. Illustration of sensitivity to the estimated number of endmembers. The actual number of endmembers is 6 (SNR=20dB). (a) AID (b) SID.

5) Generalization to larger numbers of endmembers: In this experiment, the algorithm is evaluated with images containing different numbers of endmembers. We vary the number of endmembers from $c = 3$ to $c = 10$ and create a set of synthetic images. Fig. 7 illustrates the unmixing performance of different methods, where the SID measures are the logarithmic transforms given by $y = \log(1+x)$, with x denoting the actual SID value and y the transformed result. This transformation is used to suppress the big difference between the PGNMF and the MVC-NMF estimates for better visual comparisons. As can be seen that MVC-NMF provides the best performance for images with different numbers of endmembers, while PGNMF

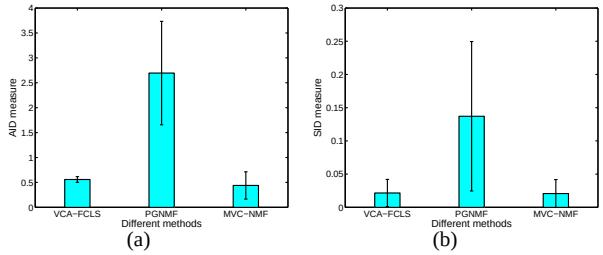
displays the worst results. Examining the results of VCA-FCLS and MVC-NMF, the AID measures increase slightly as the number of endmembers present in the mixture data increases, while the SID measures is less dependent of this parameter by illustrating very similar results when different numbers of endmembers are used.



(a) (b)

Fig. 7. Performance comparison when applied to images created by mixing different numbers of endmembers (SNR=20dB) (a) AID (b) SID (The SID plot displays the logarithmic transforms of actual values for better visual comparisons).

6) Generalization to images containing pure pixel representations: Although the proposed method is motivated by unmixing highly mixed images, it is important to study its generalization to the mixture data containing pure pixels. For this purpose, we take a 7×7 low pass filter to create synthetic mixtures so that the created images contain pure pixels. The experimental results are illustrated in Fig. 8. Similarly, the SID results are displayed in terms of the logarithmic transforms. Note that MVC-NMF generates the smallest mean and slightly worse (in terms of AID) or the same (in terms of SID) standard deviation as VCA-FCLS. This result leads to an important conclusion that MVC-NMF is not specific for unmixing highly mixed data. It can also produce reliable unmixing results when applied to hyperspectral images containing pure pixels, which further broadens the application domain of the proposed MVC-NMF method.



(a) (b)

Fig. 8. Performance comparison when the image contains pure pixels (SNR=20dB) (a) AID (b) SID (The SID results are the logarithmic transforms of actual measures).

V. Evaluation With Real Image Scene

In this section, we apply the MVC-NMF algorithm to real hyperspectral data captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Cuprite, Nevada. The AVIRIS sensor is a 224-channel imaging spectrometer with approximately 10nm spectral resolution covering wavelengths ranging from 0.4μm to 2.5μm. The spatial resolution is 20m. We have chosen this test site for several reasons. First, this site has been extensively used for remote sensing experiments since the 1980s, and many research work have been published with high-accuracy ground truth available [11], [35], [36]; secondly, the Cuprite area is a relatively undisturbed hydrothermal system with many well exposed minerals. More importantly, some of the minerals are prevalent, while others are highly mixed in a small set of pixels. The selected subscene is shown in Fig. 9, which consists of 200 lines and 200 pixels per line. To improve the unmixing performance, we have removed the low SNR bands as well as the water vapor absorption bands (including bands 1-2, 104-113, 148-167 and 221-224) from the original 224-band data cube. So, a total of 188 bands are used in the experiment.

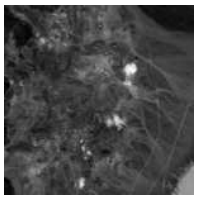


Fig. 9. Band 30 of the real hyperspectral scene

In order to evaluate the performance of MVC-NMF applied to real hyperspectral scenes, we perform two comparisons: the comparison between the extracted endmembers and the laboratory spectra; and the comparison between the derived abundance maps and the published results. For each detected endmember, we find the closest match from the spectral library based on the spectral correlation defined as

$$corr = \frac{(\hat{\mathbf{a}} - \hat{\boldsymbol{\beta}})^T (\mathbf{a}_l - \boldsymbol{\beta}_l)}{\|\hat{\mathbf{a}} - \hat{\boldsymbol{\beta}}\| \|\mathbf{a}_l - \boldsymbol{\beta}_l\|} \tag{34}$$

where $\hat{\mathbf{a}}$ and $\mathbf{a}_l$ correspond to the estimated spectrum and the library spectrum. Their mean values, denoted by $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_l$, are subtracted from the original spectra to give accurate correlation coefficients. The library spectrum with the highest correlation is chosen as the best match. However, this correlation might suffer from mismatch. To mitigate this drawback, we further visually compare the derived abundance maps to the detailed mineral maps provided in [36]. If an obvious mismatch occurs during the finding of the closest signature from the spectral library, we rely on the classification based on the abundance distribution.

The estimated number of endmembers using the VD method [28] is equal to $c = 9$. We use the VCA estimates as the starting points for the MVC-NMF learning. The extracted endmembers by VCA and MVC-NMF are shown in Fig. 10(a) and (b), respectively. Fig. 10(c) displays the closest laboratory spectra to the MVC-NMF results. We have labelled the endmember classes in Fig. 10(b), and the corresponding unlabelled VCA spectra belong to the same mineral classes. The comparisons of MVC-NMF and VCA in terms of spectral angles between extracted endmembers and laboratory signatures are summarized in Table I. Quantitatively, MVC-NMF generates smaller spectral angles than VCA in most cases, which means its estimates better resemble the laboratory data.

By side-by-side visual comparison of the endmembers

TABLE I

SPECTRAL ANGLES BETWEEN EXTRACTED ENDMEMBERS AND LIBRARY SPECTRA BY MVC-NMF AND VCA

	MVC-NMF	VCA
Sphene	4.22	5.54
Nontronite	7.22	7.58
Kaolinite	7.16	5.29
Montmorillonite	8.42	7.28
Chalcedony	6.86	9.54
Desert varnish	7.01	18.41
Alunite	3.52	5.45
Buddingtonite	3.95	7.45
Muscovite	2.34	8.53

extracted by both methods, we observe that they generally identify similar minerals, such as the first six endmembers viewed along the top-down direction. However, MVC-NMF has the potential of providing better results and identifying less prevalent minerals. Specifically, when comparing the *Alunite* spectrum (the seventh endmember), we note that the VCA estimate does not clearly show the characteristic absorption features in the spectral range of 2300nm $\sim$ 2500nm, while the MVC-NMF result shows better resemblance to the laboratory spectrum. Of particular interest are the last two endmembers. VCA identifies *Kaolinite* and *Chalcedony*, which are the same mineral classes as the third and the fifth endmember, respectively. In other words, VCA only identifies seven endmembers although the desired number of endmembers is $c = 9$. The possible cause is that the mineral *Kaolinite* and *Chalcedony* are both prevalent classes, making the data distribution prone to these two endmembers. Therefore, it is very likely that more than one extreme-valued pixel is close to these endmembers. On the contrary, the proposed method derives two new classes, *Buddingtonite* and *Moscovite*, which are close to the laboratory spectra shown in Fig. 10(c).

Fig. 11 illustrates the estimated abundance maps. Compared to the published geologic maps, these estimations present high level of similarity to the published results. The last two maps are of interest, in which the abundance maps present big values in a small area and zero everywhere else. This observation confirms that MVC-NMF is able to identify less prevalent endmembers, which are, however, not detectable using VCA.

VI. CONCLUSION AND FUTURE WORK

This paper addressed a novel endmember extraction method for highly mixed hyperspectral data without the pure pixel assumption. Through the analysis of the connection between the spectral unmixing analysis and the non-negative matrix factorization, we developed an effective and promising constrained NMF algorithm for decomposing mixed pixels. The experimental results with synthetic mixtures showed that the introduced minimum volume constraint results in more accurate endmember estimates, thus more reliable abundance estimates as a direct benefit. A comparison with three advanced algorithms from different perspectives was conducted, which demonstrated that the proposed MVC-NMF outperforms all

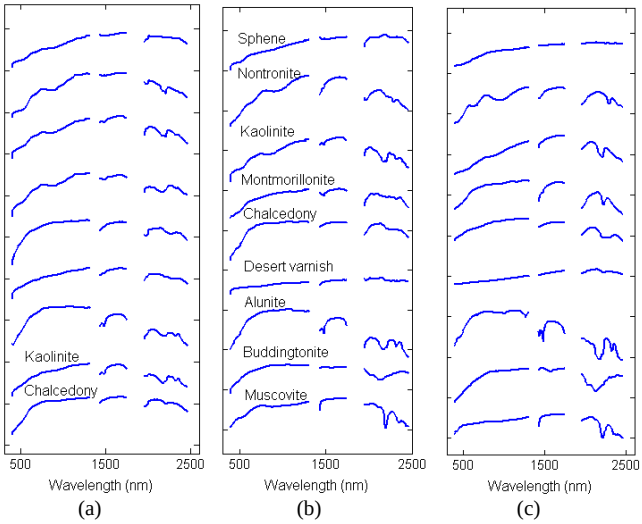


Fig. 10. Comparison of endmember signatures extracted by (a) VCA, (b) MVC-NMF, (c) The closest laboratory spectra to the MVC-NMF estimates. The unlabelled VCA spectra belong to the same minerals as the corresponding endmembers labelled in (b). Note that MVC-NMF detects Buddingtonite and Muscovite that are not detectable using VCA. The order of spectra (from up to down) has been rearranged for easier description instead of using the detection order.

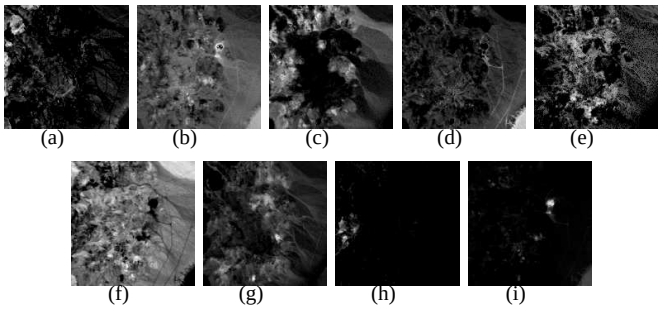


Fig. 11. Abundance maps of different minerals estimated using MVC-NMF. (a) Sphene (b) Nontronite (c) Kaolinite (d) Montnorillonite (e) Chalcedony (f) Desert varnish (g) Alunite (h) Buddingtonite (i) Muscovite.

the other approaches compared. The volume constraint has made the MVC-NMF learning less dependent on the initializations, robust to different levels of noise corruptions, less sensitive to the estimated number of endmembers, and applicable to images containing larger number of endmembers as well as those with pure pixel representations. The proposed method has also been applied to a real hyperspectral scene collected by the AVIRIS sensor. The experimental analyses have indicated that MVC-NMF has the potential of providing more accurate estimates and identifying less prevalent endmembers.

As with any newly developed techniques, there exist some important issues that are in need of further research. First, the current method utilizes all the image pixels during the learning process, which involves heavy computation burden. We suspect that the employment of the boundary pixels alone would be able to generate comparable results. In the future research, we will study how to identify the pixels on or close to the boundary efficiently. In addition, efficient learning algorithms should be investigated. The current learning is gradient-based, which could be improved by adopting, for

example, the second-order methods.

REFERENCES

[1] G. M. Foody, *Remote Sensing Image Analysis: Including the Spatial Domain*. Kluwer Academic Publishers, 2004, ch. 3, pp. 37–49.

[2] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 55–78, 2003.

[3] M. O. Smith, P. E. Johnson, and J. B. Adams, "Quantitative determination of mineral types and abundances from reflectance spectral using principal components analysis," *J. Geophys. Res.*, no. 90, pp. C797–C804, 1985.

[4] J. Bayliss, J. A. Gualtieri, and R. Cromp, "Analyzing hyperspectral data with independent component analysis," in *Proc. of SPIE*, vol. 3240, 1997, pp. 133–143.

[5] J. M. P. Nascimento and J. M. B. Dias, "Does independent component analysis play a role in unmixing hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 1, pp. 175–184, Jan. 2004.

[6] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 6, pp. 2706–2717, 1999.

[7] J. W. Boardman, "Automating spectral unmixing of aviris data using convex geometry concepts," in *Summaries of the fourth annual JPL airborne geoscience workshop*, R. O. Green, Ed., 1994, pp. 11–14.

[8] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 3, pp. 542–552, May 1994.

[9] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of aviris data," in *Summaries of Fifth Annual JPL Airborne Geoscience Workshop*, R. O. Green, Ed., vol. 1, 1995, pp. 23–26.

[10] M. E. Winter, "Fast autonomous spectral endmember determination in hyperspectral data," in *Proc. of 13th international conference on applied geologic remote sensing*, vol. 2, Vancouver, BC, Canada, 1999, pp. 337–344.

[11] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[12] V. F. Haertel and Y. E. Shimabukuro, "Spectral linear mixing model in low spatial resolution image data," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 11, pp. 2555–2562, Nov. 2005.

[13] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "Ice: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[14] D. Lee and S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[15] ——, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, vol. 13. MIT Press, 2001, pp. 556–562.

[16] L. C. P. P. Sajda, S. Du, "Recovery of constituent spectra using non-negative matrix factorization," in *Proceedings of SPIE*, vol. 5207, 2003.

[17] C.-Y. Liou and K. O. Yang, "Unsupervised classification of remote sensing imagery with non-negative matrix factorization," in *ICONIP*, 2005.

[18] V. P. Paura, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *To appear in Linear Algebra and Applications*, 2006.

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

[20] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Preprint submitted to Elsevier*, 2006.

[21] C.-J. Lin, "Projected gradient methods for non-negative matrix factorization." Department of Computer Science, National Taiwan University" Technical report, 2005.

[22] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[23] M. Chu, F. Diele, R. Plemmons, and S. Ragni, "Optimality, computation, and interpretations of nonnegative matrix factorizations," *Available Online: http://www.wfu.edu/ plemmons*, 2004.

[24] S. M. Wild, "Seeding non-negative matrix factorizations with the spherical k-means clustering," M.S. Thesis, Department of Applied Mathematics, University of Colorado, Apr 2003.

[25] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts," in *Advances in Neural Information Processing Systems 16*, S. Thrun and B. Scholkopf, Eds. Cambridge, MA: MIT Press, 2004.

[26] A. Pascual-Montano, J. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsnmf)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 403–415, Mar 2006.

[27] G. Strang, *Linear Algebra and its application*, 3rd ed. Massachusetts Institute of Technology, 1988.

[28] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[29] M. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.

[30] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[31] R. Bro and S. D. Jong, "A fast non-negativity constrained least squares algorithm," *J. Chemom.*, vol. 11, pp. 393–401, 1997.

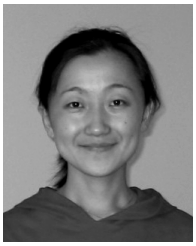
[32] C.-I. Chang and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 3, pp. 1144–1159, May 2000.

[33] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The u.s. geological survey digital spectral library: Version 1:0.2 to 3.0 μm," U.S. Geol. Surv.," Open File Rep. 93-592, 1993.

[34] C. Kwan, B. Ayhan, G. Chen, J. Wang, B. Ji, and C.-I. Chang, "A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 2, pp. 409–419, Feb. 2006.

[35] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing aviris mineral mapping at cuprite, nevada," in *Summaries 3rd Annu. JPL Airborne Geosciences Workshop*, vol. 1, 1992, pp. 47–49.

[36] G. Swayze, "The hydrothermal and structural history of the cuprite mining district, southwestern nevada: An integrated geological and geophysical approach," PhD Dissertation, University of Colorado, Boulder, 1997.



Lidan Miao (S'04) received her BS. and MS. degrees in Electrical Engineering from Sichuan University, China, in 2000 and 2003, respectively. Since 2004, she has been pursuing her PhD degree in the Department of Electrical and Computer Engineering, University of Tennessee, Knoxville, TN.

Her current research interests include signal and image processing, pattern recognition, and remote sensing. She is a student member of IEEE.

Hairong Qi (SM'05) received her Ph.D. degree in Computer Engineering from North Carolina State University in 1999, B.S. and M.S. degrees in Computer Science from Northern JiaoTong University, Beijing, P.R.China in 1992 and 1995 respectively.

She is now an Associate Professor in the Department of Electrical and Computer Engineering at the University of Tennessee, Knoxville. Her current research interests are advanced imaging and collaborative processing in sensor networks, hyperspectral image analysis, and bioinformatics. She has published over 70 technical papers in archival journals and refereed conference proceedings, including a co-authored book in Machine Vision.

Dr. Qi is the recipient of the NSF CAREER award, Chancellor's Award for Professional Promise in Research and Creative Achievement. She serves on the editorial board of Sensor Letters and is the Associate Editor for Computers in Biology and Medicine.