Multiple Clustering Guided Nonnegative Matrix Factorization for Hyperspectral Unmixing

Wenhong Wang, Yuntao Qian, *Member, IEEE*, and Hongfu Liu, *Member, IEEE*

***Abstract*—Spectral unmixing is an important technique for quantitatively analyzing hyperspectral remote sensing images. Recently, constrained nonnegative matrix factorization (NMF) has been demonstrated to be a powerful tool for spectral unmixing. However, acquiring the problem-dependent prior knowledge and incorporating it into NMF as effective constraints is a challenging task. In this article, a multiple clustering guided NMF unmixing approach is proposed under a self-supervised framework, which has been used to effectively learn high-level semantic information from the data with a surrogate task in many applications. Specifically, in order to provide self-supervised information to guide the NMF-based unmixing model, multiple clustering is integrated into the optimization process of NMF. Moreover, by introducing interaction between each clustering and the unmixing procedure, more accurate proximate endmember signatures and proximate abundance distributions are expected to be acquired and used to impose self-supervised constraints on endmembers and abundances, respectively. Consequently, effective prior information about endmember signatures and abundance distributions is captured and simultaneously integrated into NMF as valuable constraints to promote unmixing performance. Experiments are conducted on both synthetic data and real hyperspectral images, and the superior performance of our method is shown by comparing it with several state-of-the-art algorithms.**

***Index Terms*—Hyperspectral image (HSI) unmixing, multiple clustering, nonnegative matrix factorization (NMF), self-supervised learning.**

I. Introduction

HYPERSPECTRAL images (HSIs) are usually captured using many different electromagnetic bands and can contain rich spatial and spectral information about the observed scene, and thereby have many real applications. Limited by the low spatial resolution of the sensors, each pixel of an HSI usually cover a relatively large ground area, so that its spectral information may be a mixture of several pure spectra (i.e., endmembers) of the materials present in this area [1]. These mixed pixels sometimes have serious implications for quantitative analysis of HSIs. One of the techniques to tackle this problem is hyperspectral unmixing (HU), by which a set of endmembers included in an HSI and the corresponding proportions (i.e., abundances) of them in each pixel are estimated. To solve HU problem, two types of spectral mixture models are commonly used for HU: the linear spectral mixture model (LSMM) [1] and the nonlinear spectral mixture model (NLSMM) [2]. NLSMM is mainly adopted to express a complex spectral mixing mechanism caused by multiple scattered effects of source radiation among several endmembers. LSMM assumes that source radiation is only reflected once by the ground materials before being captured by the sensor. Owing to the flexibility and tractability, LSMM is a widely used model to express the mixture mechanism of HSIs.

Under LSMM, numerous HU methods have been proposed over decades [2]. A type of representative methods is developed based on the principle of convex geometry, which supposes that the observed pixels can constitute a simplex in the Euclidean space with vertices denoting endmembers. Then, the task of unmixing is transformed to find a group of vertices that can form a simplex with maximum volume, such as N-FINDR algorithm [3], pixel purity index [4], vertex component analysis (VCA) [5], and so on. Recently, nonnegative matrix factorization (NMF) [6], [7] based HU approaches have attracted much attention because of its several advantages, such as the natural nonnegativity and good interpretability of unfixing results. NMF regards HU as a blind source separation problem and aims to decompose an HSI into two nonnegative factor matrices. Thus, it can unmix the endmembers and the corresponding abundances simultaneously. However, because of the nonconvexity of its objective function, the standard NMF is prone to fall into local minima, and also vulnerable to the effects of initial values [8]. One effective measure to tackle these issues is to impose additional constraints on the factor matrices. For example, Miao and Qi [9] introduced geometrical principle-based constraint on endmembers, so that the obtained minimum volume constrained NMF (MVCNMF) can unmix the highly mixed data. Qian *et al.* [10] proposed a $L_{1/2}$ sparsity-constrained NMF ($L_{1/2}$-NMF) unmixing method by employing an effective $L_{1/2}$ regularizer that can enforce the estimated abundances to be more accurate and sparse. Following the same tactic, graph regularized NMF variant (GLNMF) [11] and spatial group sparsity regularized NMF (SGSNMF) [12] have been proposed in pursuit of abundance estimation with structured sparsity. In addition, spatial information has also been used to construct effective constraints for the NMF model,

Manuscript received May 19, 2020; revised July 8, 2020 and July 31, 2020; accepted August 18, 2020. Date of publication August 31, 2020; date of current version September 16, 2020. This work was supported in part by the 2030 National Key AI Program of China under Grant 2018AAA0100500 and in part by the Shandong Provincial Natural Science Foundation under Project ZR2017MF028. *(Corresponding author: Wenhong Wang.)*

Wenhong Wang is with the College of Computer Science, Liaocheng University, Liaocheng 252059, China (e-mail: wangwenhong@lcu.edu.cn).

Yuntao Qian is with the Institute of Artificial Intelligence, College of Computer Science, Zhejiang University, Hangzhou 310027, China (e-mail: ytqian@zju.edu.cn).

Hongfu Liu is with the Volen National Center for Complex Systems, Department of Computer Science, Brandeis University, Waltham, MA 02453 USA (e-mail: hongfuliu@brandeis.edu).

Digital Object Identifier 10.1109/JSTARS.2020.3020541

such as abundance separation [13], piecewise smooth [14], and substance dependence constraints [15]. Differ from the NMF-based unmixing framework, some recent works tackle the HU problems by decomposing a third-order tensor of the target HSI via tensor factorization, so that the intrinsic spectral-spatial joint structure of HSIs can be fully utilized. For example, Qian *et al.* [16] proposed a matrix-vector nonnegative tensor factorization (MVNTF) method by integrating the characteristics of canonical polyadic decomposition and Tucker decompositions. Xiong *et al.* [17] introduced total variation regularization into MVNTF-based unmixing model, by which the global spectral-spatial information and local spatial information were simultaneously exploited. Using the idea of increasing additional constraints, Feng and Wang [18] improve the plain MVNTF method by integrating sparseness, volume, and nonlinearity constraints into the cost function. Besides, low-rank constraints for abundance and endmember tensors have also been adopted in NTF-based unmixing methods [19], [20].

Some NMF variants for HU are developed by making use of semisupervised information. Recently, the prior signature information of partial known endmembers has been recognized as another important problem-dependent knowledge for HU. This is reasonable due to the fact that some land covers can be known in advance, for example, by field investigation [21], in real applications. Under such circumstances, spectral signatures of known materials can be obtained from the spectral library [2], or even be retrieved directly from the image regions that contain only pure endmembers [22]. Based on the prior spectra, different regularizations have been devised to impose effective constraints on endmembers in NMF-based unmixing models [21]–[24]. For example, Tang *et al.* [21] proposed an unmixing method, referred to as NMFupk, in which the known endmembers in the endmember matrix are constrained to be fixed in the optimized process, whereas the remainder elements of the endmember matrix need to be updated iteratively. To improve NMFupk, Wang *et al.* [23] explored the correlation between the known and unknown endmembers by introducing a function into the unmixing model, which describes the property that the known and unknown endmembers lie in the nearly orthogonal planes. This made the known endmembers can be used to help the estimation of the unknown endmembers. Tong *et al.* [22] developed an NMF-based unmixing method with partially known endmembers (called NMF-PPK) by incorporating a regularization that can minimize the difference between the known endmembers and the estimated ones. Although these works have improved the unmixing performance of NMF effectively, they also suffer from the following three problems: first, the spectra in the spectral libraries are seldom acquired under the same conditions with real HSIs [25]. As a result, there may be a large deviation when the spectra from these libraries are used as known endmember signatures; second, the completeness of these libraries is not guaranteed [22], which limits the successful application of these methods; third, it is often the case that prior knowledge of only rare endmembers is available on the target scene in practical applications, so that the performance enhancement of these methods is more confined. Therefore, how to accurately mine the implicit signature information of all the endmembers from the data, and take it as semisupervised information to devise effective regularizations for NMF remains a challenging problem.

In this article, to take full advantage of the self-learning semisupervised information about endmembers and abundances, a novel NMF-based unmixing approach, named multiple clustering guided NMF (MCG-NMF), is proposed. Our work assumes that the distributions of practical land covers show significant locality [10], [26], and lots of mixed pixels in an HSI may include only one or very few kinds of dominant ground objects [27]. Therefore, these pixels might carry the signature information of endmembers to a large extent and can be extracted to guide the unmixing. According to this observation, in this article, the proximate endmember signatures (PESs) and proximate abundance distributions (PADs) are captured to supply valuable cues for more accurate estimation of endmembers and abundances in NMF-based unmixing under the framework of self-supervised learning [28]. As an important type of machine learning approach, self-supervised learning can mine the high level semantic information from the nature of data with a surrogate task, and has been successfully applied in many fields, such as video hashing [28], classification of HSIs [29], and learning visual representation [30]. With self-supervised learning, PESs and PADs as semisupervised information can be effectively learned from the data via multiple clustering tasks, rather than supplied by the users. Specifically, to provide more accurate PESs for the NMF-based unmixing model, we conduct multiple clustering to the target HSI with the optimization progress of NMF. Although several clustering methods, such as hierarchical clustering [31] and iterative self-organizing data [32], can be used to make effective cluster analysis to HSIs, we adopt the K-means method to achieve multiple clustering considering its ease of implementation and computational effectiveness. Furthermore, based on the obtained PESs, the corresponding PADs are effectively calculated and integrated into NMF as semisupervised guidance for abundance estimation. This can not only effectively confine the solution space of NMF to meet the need of HU application, but also promote more accurate estimation results as well as stable unmixing performance. For the sake of clarity, the contributions of this article are summarized as follows.

1) We propose a novel NMF method for HU based on the guidance of multiple clustering. By conducting clustering analysis on target HSI, PESs and PADs are obtained as semisupervised information and used to construct effective regularizations for NMF. This is reasonable considering that only one or few kinds of substances play a dominant role in a local region because of the locality property of HSIs in spatial domains [26], [27]. Thus, effective clustering method can be used to extract valuable endmember and abundance information to guide the unmixing process.
2) Our method integrates both self-supervised learning [29], [30] and matrix factorization into a unified framework, in which more accurate semisupervised information used by NMF can be automatically learned via interactions between the multiple clustering and the procedure of matrix

factorization. Consequently, the proposed method can be effectively applied in scenarios where the endmember and corresponding abundances are not known *a priori*.

3) The performance of our approach is validated on both synthetic and real HSIs. The experimental results demonstrate that the proposed method can outperform several state-of-the-art algorithms.

The remainder of this article is structured as follows. A brief description of LSMM and NMF is given in Section II. Then, we present the formulation of our proposed MCG-NMF and its optimization, as well as some implementation issues in Section III. Next, extensive experimental results and analysis on synthetic data and real HSIs are discussed in Section IV. Finally, we conclude this article and provide possible future research directions in Section V.

II. Related Work

In this section, we briefly describe the important research basis of the proposed MCG-NMF method on HU, including LSMM and NMF.

A. Linear Spectral Mixture Model

LSMM is widely used in HU research because of its simplicity and effectiveness. According to LSMM, an observed pixel spectrum of an HSI can be represented as a linear combination of a set of endmembers included, in which the weight of each endmember corresponds to the fractional abundance of this pixel spectrum. To be specific, LSMM can be formalized as follows:

$$\mathbf{y}_n = \mathbf{E}\mathbf{a}_n + \mathbf{v}_n \tag{1}$$

where vector $\mathbf{y}_n$ denotes the nth pixel spectrum and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_P] \in \mathbb{R}^{L\times P}$ is the endmember matrix with each column representing an endmember signature; $\mathbf{a}_n = (a_{n1}, a_{n2}, ..., a_{nP})^\top$ stands for the abundance vector corresponding to $\mathbf{y}_n$; $\mathbf{v}_n = (v_{n1}, v_{n2}, ..., v_{nL})^\top$ denotes an additive noise vector, in which the components can be assumed to subject to independent and identically distributed (i.i.d.) zero-mean Gaussian distribution [33]. In order to have explicit physical meaning, the abundance vector is usually subject to the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC) [2], which are, respectively, defined as

$$\text{ANC:}\quad a_{np} \geq 0, p = 1, ..., P \tag{2}$$

$$\text{ASC:}\quad \sum_{p=1}^{P} a_{np} = 1. \tag{3}$$

Using matrix operations, (1) can be rewritten as more compact form, i.e.,

$$\mathbf{Y} = \mathbf{E}\mathbf{A} + \mathbf{V} \tag{4}$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N] \in \mathbb{R}^{L\times N}$ is the HSI matrix with N pixels and L bands; $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N] \in \mathbb{R}^{P\times N}$ denotes the abundance matrix; and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N] \in \mathbb{R}^{L\times N}$ represents the noise matrix. Note that the number of endmembers included in a real HSI is usually much smaller than the number of bands, i.e., $P \ll L$. Considering that both $\mathbf{E}$ and $\mathbf{A}$ are unknown in practice, the task of HU can be regarded as a matrix factorization problem that aims to decompose a high-dimensional observation matrix into two low-dimensional factor matrices with the subjection of abundance matrix to ANC and ASC.

B. Nonnegative Matrix Factorization

The NMF method was originally proposed by Lee and Seung [6] in *Nature* for the purpose of reducing the dimension of high-dimensional data. Thereafter, it has been developed to be a powerful tool for statistical analysis [8]. Specifically, NMF aims to factorize a nonnegative data matrix into the product of two factor matrices with nonnegative entries, so that the reconstructed data matrix can approximate the original one as far as possible. Given that $\mathbf{Y}$ denotes the nonnegative observation matrix with each column as a data sample, the standard NMF can be formally modeled as

$$\min_{\mathbf{E},\mathbf{A}} F(\mathbf{E}, \mathbf{A}) = \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2, \quad \text{s.t. } \mathbf{E}, \mathbf{A} \succeq 0 \tag{5}$$

where $\mathbf{E}$ and $\mathbf{A}$ represent nonnegative matrix factors; $\|\cdot\|_F$ and "$\succeq$" stand for the Frobenius norm and the componentwise greater-or-equal relation of a matrix, respectively.

To optimize both matrix factors $\mathbf{E}$ and $\mathbf{A}$ of the objective function in (5), Lee and Seung [7] presented the famous multiplicative update rules (MURs) as follows:

$$\mathbf{E} \leftarrow \mathbf{E} \odot (\mathbf{Y}\mathbf{A}^\top \oslash \mathbf{E}\mathbf{A}\mathbf{A}^\top) \tag{6}$$

$$\mathbf{A} \leftarrow \mathbf{A} \odot (\mathbf{E}^\top\mathbf{Y} \oslash \mathbf{E}^\top\mathbf{E}\mathbf{A}) \tag{7}$$

where $(\cdot)^\top$ denotes the matrix transpose; "$\odot$" and "$\oslash$" represent elementwise multiplication and division between two matrices, respectively. In addition, some representative optimization methods for NMF were also investigated, such as the active set method [34], projected gradient method [35], projected nonnegative least squares [36], and optimal gradient method [37]. However, the numerical minimization of the objective function $F(\mathbf{E}, \mathbf{A})$ in (5) still encounters some challenges owing to the nonconvexity of $F(\mathbf{E}, \mathbf{A})$ w.r.t. matrices $\mathbf{E}$ and $\mathbf{A}$ simultaneously. This causes the solutions of standard NMF to be prone to being trapped into local minima. Additionally, the standard NMF also suffers from the problem of nonunique solution [38]. To alleviate these problems, various constraints, such as manifold [39], sparseness [40], and low-rank constraints [41], as well as smooth constraint [42], have been introduced into the standard NMF according to the requirements of different applications, and the resulting variants of NMF are called constrained NMF (CNMF). Generally, the CNMF can be expressed as

$$\min_{\mathbf{E},\mathbf{A}} C(\mathbf{E}, \mathbf{A}) = \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 + \alpha J_1(\mathbf{E}) + \beta J_2(\mathbf{A}),$$
$$\text{s.t. } \mathbf{E}, \mathbf{A} \succeq 0 \tag{8}$$

where functions $J_1(\mathbf{E})$ and $J_2(\mathbf{A})$ are regularization terms with the role of enforcing certain constraints on both matrix factors; α and β are two regularization parameters used to adjust the effects of the corresponding regularization term on the objective function $C(\mathbf{E}, \mathbf{A})$.

In order to solve the CNMF problem defined by (8), the Lagrange multipliers method is one of the commonly used methods to deduce the update rules of matrix factors under the alternating optimization framework [11], [39]. Considering the objective function $C(\mathbf{E}, \mathbf{A})$ in (8), the Lagrange function associated with it is

$$\Gamma(\mathbf{E}, \mathbf{A}, \Phi, \Psi) = \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 + \alpha J_1(\mathbf{E}) + \beta J_2(\mathbf{A}) + \mathrm{Tr}(\Phi\mathbf{E}^\top) + \mathrm{Tr}(\Psi\mathbf{A}^\top) \tag{9}$$

where Φ and Ψ are two Lagrange multipliers matrices associated with $\mathbf{E}$ and $\mathbf{A}$, respectively. Then, the following conditions, called Karush–Kuhn–Tucker (KKT) conditions [43], for the CNMF problem have to hold:

$$\begin{cases} \frac{\partial \Gamma}{\partial \mathbf{E}|_{\mathbf{E}=\mathbf{E}^*}=0, \frac{\partial \Gamma}{\partial \mathbf{A}}|_{\mathbf{A}=\mathbf{A}^*}=0} \\ \Phi \odot \mathbf{E}^* = 0, \Psi \odot \mathbf{A}^* = 0 \\ \Phi \succeq 0, \Psi \succeq 0 \\ \mathbf{E}^* \succeq 0, \mathbf{A}^* \succeq 0 \end{cases} \tag{10}$$

where $(\mathbf{E}^*, \mathbf{A}^*)$ denotes a local minimum of the CNMF problem.

III. MCG-NMF Unmixing Method

Here, we first address the regularization terms introduced for NMF, as well as the proposed MCG-NMF unmixing model that can exploit the self-learning semisupervised information during the unmixing process. Then, the MURs and the corresponding optimization technique are introduced. Then, the implementation issues and the MCG unmixing algorithm are presented.

As aforementioned, incorporating prior signatures of particular endmembers into NMF as semisupervised information plays an important role in improving the unmixing performance. The related methods [21]–[24] commonly assume that the signatures of endmembers known in advance can be acquired from the available spectral libraries. However, the practical applications of these methods suffer from the incompleteness of dependent libraries or the diversity between the spectra of the libraries and the corresponding actual endmember signatures of HSIs [25]. As a result, the performance of such methods is not always satisfactory. Fortunately, a recently proposed weakly supervised NMF [44] that has been successively applied in cluster analysis provides the chance to introduce imprecise prior information about endmembers and abundances into NMF-based unmixing model. However, specifying prior information for endmember signatures and abundance distributions remains a hard work in practice. In this article, inspired by the principle of self-supervised learning [28]–[30], we aim to extract prior information about signatures of all the endmembers and corresponding abundance distributions by conducting multiple clustering analysis on HSIs during the optimization process of NMF, so that problem-dependent knowledge about endmembers and abundances can be acquired via self-learning means, and used as effective semisupervised guidance for NMF-based unmixing procedure.

A. Self-Supervised Constraint for Endmembers

Because of the locality of material distribution in a real scene, all endmembers in general contribute unequally to a specific pixel in HSIs [10], [26], [45], so that many pixels are dominated by only one or very few kinds of endmembers [27]. Thus, effective cluster analysis technology can be used to approximately estimate endmember signatures from the data with the number of clusters and endmembers being the same [45]–[47]. For example, in [46], endmember signatures used to construct desired constraints for Bayesian NMF are estimated from data via subspace clustering. In another study [45], spectral unmixing is formulated as a clustering problem, and endmember extraction is implemented by a modified K-means method. Considering its computational efficiency and ease of implementation, the K-means method is integrated into NMF to help estimating PESs of all the endmembers in this study.

First, we present how to extract PESs of all the endmembers by conducting cluster analysis on the target HSI via the K-means method. Based on the clustering results, we construct a group of PESs for each type of endmember by first generating a mean pixel in the corresponding cluster, and then select a set of pixels that have better similarity to this mean pixel. Specifically, clustering analysis is conducted on the data matrix $\mathbf{Y}$ first, and a set of clusters $\{C_1, C_2, ..., C_P\}$ is obtained, in which each cluster includes a subset of all pixels in the data and the intersection of C_i and $C_j, i \neq j$ is the empty set. Next, the mean pixel $\mathbf{r}_p$ of each C_p is calculated by

$$\mathbf{r}_p = \frac{1}{|C_p|} \sum_{\mathbf{y}_i \in C_p} \mathbf{y}_i, p = 1, 2, ..., P \tag{11}$$

where $|C_p|$ denotes the number of pixels in cluster C_p. Then, based on $\mathbf{r}_p$, we select K pixels, i.e., $G_p = \{\mathbf{y}_p^1, \mathbf{y}_p^2, ..., \mathbf{y}_p^K\}$, from each cluster C_p according to the principle of K-nearest neighbors [48], which consider the Euclidean distance between $\mathbf{r}_p$ and the other pixels $\mathbf{y}_i \in C_p$, i.e., $\|\mathbf{r}_p - \mathbf{y}_i\|$. Here, without loss of generality, we let $\|\mathbf{r}_p - \mathbf{y}_p^k\| \leq \|\mathbf{r}_p - \mathbf{y}_p^{k+1}\|$, where $k = 1, 2, ..., K-1$. Accordingly, for each endmember $\mathbf{e}_p$, $K+1$ pixels, i.e., $\{\mathbf{r}_p\} \bigcup G_p$, are obtained and regarded as its PESs. Finally, we construct $K+1$ PES matrices $\mathbf{R}^k, k = 1, 2, ..., K+1$, as follows:

$$\mathbf{R}^k = \begin{cases} [\mathbf{y}_1^k, \mathbf{y}_2^k, ..., \mathbf{y}_p^k, ..., \mathbf{y}_P^k], & \text{for} k = 1, 2, ..., K \\ \mathbf{U}, & \text{for} k = K+1 \end{cases} \tag{12}$$

where $\mathbf{U} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_p, ..., \mathbf{r}_P]$; $\mathbf{y}_p^k$, $p = 1, 2, ..., P$ and $k = 1, 2, ..., K$, represents a PES with index k taken from the corresponding G_p. As an illustration, the generation procedure of PESs and the PES matrices is shown in Fig. 1.

In order to make full use of the estimated PES matrices in NMF-based HU, we introduce a self-supervised constraint for endmembers, by which each estimated endmember is compelled to be similar with the corresponding PESs in all PES matrices. To be specific, a function $J(\mathbf{E})$ that measures the average difference between each estimated endmember $\mathbf{e}_p$ in endmember matrix $\mathbf{E}$ and the corresponding signatures $\mathbf{r}_p^k$ in matrices $\mathbf{R}^k, k = 1, 2, ..., K+1$, is constructed via the F-norm

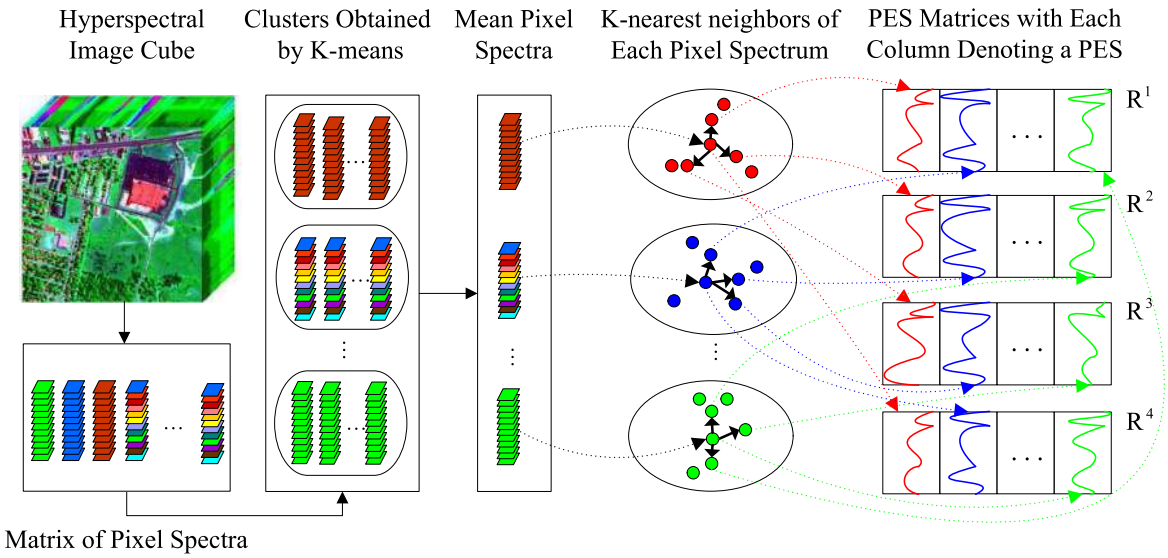


Fig. 1. Illustration of the generation procedure of PESs and the PES matrices with $K = 3$. By conducting clustering analysis on the data matrix via K-means, a set of clusters and corresponding mean pixel spectra are obtained. Then, the K-nearest neighbors of each mean pixel spectrum are found with $K = 3$ according to the Euclidean distance between this mean pixel spectrum and the other pixels within the same cluster. Consequently, each mean pixel spectra and their K-nearest neighbors are regarded as a group of PESs of a certain type of endmember. Based on the obtained PESs, each PES matrix $\mathbf{R}^k, k = 1...4$, is constructed by selecting a PES from different PES group, except for $\mathbf{R}^4$ that is consist of all the mean pixel spectra.

of the matrix, which is formalized as

$$J(\mathbf{E}) = \frac{1}{K+1}\sum_{k=1}^{K+1}\|(\mathbf{R}^k - \mathbf{E})\|_F^2. \tag{13}$$

Then, the function $J(\mathbf{E})$, as a regularization term, is incorporated into the objective function of standard NMF, and the obtained objective function is given by

$$F(\mathbf{E},\mathbf{A}) = \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 + \frac{\alpha}{K+1}\sum_{k=1}^{K+1}\|(\mathbf{R}^k - \mathbf{E})\|_F^2 \tag{14}$$

where α is the regularization parameter that controls the influence of $J(\mathbf{E})$. Note that the second term at the right-hand side of (14) can act as a driving force to impel each estimated endmember located near the corresponding PES in the input space, whereas the first term can minimize error between the original data matrix and the reconstructed one. To achieve a better compromise between these two acting forces, a suitable value should be chosen for α.

Although several groups of PESs have been extracted and used to impose the self-supervised constraint on endmembers, we can further acquire another type of valuable information from the clustering results to designate the significance of each set of PESs. In other words, the proximate degree of each PES set $\{\mathbf{r}_p\}\bigcup G_p$ is worth exploring and using to optimize the design of function $J(\mathbf{E})$ as follows:

$$J_{\text{weight}}(\mathbf{E}) = \frac{1}{K+1}\sum_{k=1}^{K+1}\|(\mathbf{R}^k - \mathbf{E})\mathbf{W}\|_F^2 \tag{15}$$

where $\mathbf{W}$ stands for the diagonal matrix and is referred to as proximate matrix; $w_{p,p}, p = 1, 2, ..., P$, denotes the proximate degree of PES set $\{\mathbf{r}_p\}\bigcup G_p$. In order to determine the diagonal elements of the proximate matrix $\mathbf{W}$, we consider the similarity relationship between each pixel in $\{\mathbf{r}_p\}\bigcup G_p$ and their mean pixel $\mathbf{m}_p = \frac{1}{|\{\mathbf{r}_p\}\bigcup G_p|}\sum_{\mathbf{y}_i\in\{\mathbf{r}_p\}\bigcup G_p}\mathbf{y}_i$. If the average similarity is high, this means the PES set $\{\mathbf{r}_p\}\bigcup G_p$ has high purity, and hence has a great probability of containing the proximate signatures of a latent endmember. Accordingly, a bigger weight value should be assigned to the corresponding $w_{p,p}$ so that the estimated endmember $\mathbf{e}_p$ holds a smaller distance with each pixel in $\{\mathbf{r}_p\}\bigcup G_p$. To achieve this effect, inspired by the approach of generating hyperedge weight in the hypergraph model [49], the diagonal elements $w_{p,p}$ is calculated by

$$w_{p,p} = \sum_{\mathbf{y}_j\in\{\mathbf{r}_p\}\bigcup G_p}\exp\left(-\frac{\|\mathbf{y}_j - \mathbf{m}_p\|^2}{\sigma^2}\right) \tag{16}$$

where $\sigma = \frac{1}{(K+1)P}\sum_{p=1}^{P}\sum_{\{\mathbf{y}_j\in\{\mathbf{r}_p\}\bigcup G_p\}}\|\mathbf{y}_j - \mathbf{m}_p\|$ represents the average distance between the mean pixel $\mathbf{m}_p$ and all the pixels $\mathbf{y}_j \in \{\mathbf{r}_p\}\bigcup G_p$. With this strategy, the value of each diagonal element $w_{p,p}$ is determined by making a better tradeoff between the similarity in each class and the average similarity among all the classes.

By replacing the second term of the objective function in (14), we obtain the following objective function:

$$F(\mathbf{E},\mathbf{A}) = \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 + \frac{\alpha}{K+1}\sum_{k=1}^{K+1}\|(\mathbf{R}^k - \mathbf{E})\mathbf{W}\|_F^2. \tag{17}$$

B. Self-Supervised Constraint for Abundances

Based on the obtained PESs of all the endmembers and class information of each pixel spectrum, we estimate the PADs for each type of endmember, so that effective constraint is imposed on the abundance matrix of the NMF-based unmixing model. Specifically, a matrix $\hat{\mathbf{A}}$ is first obtained using the least squares method as follows:

$$\hat{\mathbf{A}} = \max(0, ((\bar{\mathbf{R}})^\top \bar{\mathbf{R}})^{-1} (\bar{\mathbf{R}})^\top \mathbf{Y}) \tag{18}$$

where function $\max(\cdot)$ is used to guarantee the nonnegativity of $\hat{\mathbf{A}}$, whereas $\bar{\mathbf{R}}$ denotes the average PES matrix given by

$$\bar{\mathbf{R}} = \frac{1}{(K+1)} \sum_{k=1}^{K+1} \mathbf{R}^k. \tag{19}$$

In order to acquire more accurate PADs, the class information of each pixel spectrum is further used to estimate the PADs via the binary masking technique [50], which aims to capture the dominant material present in a pixel. Specifically, we construct another matrix $\mathbf{A}_c \in \mathbb{R}^{P \times N}$ based on the clusters $\{C_1, C_2, ..., C_p, ..., C_P\}$ as follows:

$$[\mathbf{A}_c]_{pi} = \begin{cases} 1, & \text{if pixel spectrum} \mathbf{y}_i \in C_p, i = 1, 2, ..., N \\ 0, & \text{otherwise} \end{cases}. \tag{20}$$

So far, two kinds of strategies to estimate the PADs are introduced. Each strategy, if used separately, has its own advantages and disadvantages. To integrate the estimated PADs effectively, a common measure is to formulate them as a linear summation with data-dependent weights. Nevertheless, this will lead to more complex unmixing model and introduce another algorithm parameter requiring to be set. To balance the simplicity of the proposed unmixing model and the accuracy of the fused PADs, based on the obtained matrices $\hat{\mathbf{A}}$ and $\mathbf{A}_c$, we generate the integrated PAD matrix $\mathbf{H}$ by

$$\mathbf{H} = \frac{1}{2}(\hat{\mathbf{A}} + \mathbf{A}_c) \tag{21}$$

which leverage the value of the two kinds of PADs in a balanced way. It is worth noting that each column vector of $\mathbf{A}_c$ only includes one nonzero element, which enforces that each pixel spectrum only contains one kind of endmember corresponding to the main component in that pixel. Moreover, the obtained PAD matrix $\mathbf{H}$ has a certain degree of sparseness owing to the existence of zeros elements in matrices $\mathbf{A}_c$ and $\bar{\mathbf{A}}$, which is in accordance with the nature of abundance distributions of ground objects in a real scene.

Considering that abundance matrix $\mathbf{A}$ should meet the ASC constraint, the divergence-based function is used to quantify the approximation degree of $\mathbf{H}$ from abundance matrix $\mathbf{A}$. Specifically, it is expressed as [7], [39]

$$D(\mathbf{H}\|\mathbf{A}) = \sum_{n=1}^{N} \sum_{p=1}^{P} \left(h_{pn} \log \frac{h_{pn}}{a_{pn}} - h_{pn} + a_{pn} \right). \tag{22}$$

This function reduces to the Kullback–Leibler divergence or relative entropy when $\sum_{p=1}^{P} h_{pn} = \sum_{p=1}^{P} a_{pn}, n = 1...N$ holds [39], so that each column of $\mathbf{H}$ and $\mathbf{A}$ can be regarded as the probability distributions.

By introducing (22) into the objective function in (17) as a self-supervised constraint for abundances, the unmixing model of MCG-NMF is given as follows:

$$\min_{\mathbf{E},\mathbf{A}} F(\mathbf{E}, \mathbf{A}) = \|\mathbf{Y} - \mathbf{E}\mathbf{A}\|_F^2 + \frac{\alpha}{K+1} \sum_{k=1}^{K+1} \|(\mathbf{R}^k - \mathbf{E})\mathbf{W}\|_F^2 + \beta D(\mathbf{H}\|\mathbf{A}), \quad \text{s.t. } \mathbf{E}, \mathbf{A} \succeq 0 \tag{23}$$

where β is the regularization parameter that controls the impact of the corresponding regularization term. Note that rather than keeping fixed, the matrix $\mathbf{R}^k, k = 1...(K+1)$, $\mathbf{W}$, as well as $\mathbf{H}$ will be updated during the optimization procedure of MCG-NMF based on a self-learning process, which is presented in Section III-D.

C. Multiplicative Update Rules

Considering the nonconvexity of the objective function of MCG-NMF in (23), we solve the proposed unmixing model by alternately optimizing [51] each factor matrix with another being fixed. Next, we deduce the MURs of MCG-NMF using the method of Lagrange multipliers.

First, the objective function in (23) is re-expressed as

$$\begin{aligned} F(\mathbf{E},\mathbf{A}) &= \mathrm{Tr}((\mathbf{Y} - \mathbf{E}\mathbf{A})(\mathbf{Y} - \mathbf{E}\mathbf{A})^\top) + \beta D(\mathbf{H}\|\mathbf{A}) \\ &\quad + \frac{\alpha}{K+1} \sum_{k=1}^{K+1} \mathrm{Tr}((\mathbf{E}\mathbf{W} - \mathbf{R}^k\mathbf{W})(\mathbf{E}\mathbf{W} - \mathbf{R}^k\mathbf{W})^\top) \\ &= \mathrm{Tr}(\mathbf{Y}\mathbf{Y}^\top) - 2\mathrm{Tr}(\mathbf{Y}\mathbf{A}^\top\mathbf{E}^\top) + \mathrm{Tr}(\mathbf{E}\mathbf{A}\mathbf{A}^\top\mathbf{E}^\top) \\ &\quad + \beta \sum_{p=1}^{P} \sum_{n=1}^{N} (h_{pn} \log \frac{h_{pn}}{a_{pn}} - h_{pn} + a_{pn}) + \alpha \mathrm{Tr}(\mathbf{E}\mathbf{W}\mathbf{W}^\top\mathbf{E}^\top) \\ &\quad + \frac{\alpha}{K+1} \sum_{k=1}^{K+1} (\mathrm{Tr}(\mathbf{R}^k\mathbf{W}\mathbf{W}^\top(\mathbf{R}^k)^\top) - 2\mathrm{Tr}(\mathbf{E}\mathbf{W}\mathbf{W}^\top(\mathbf{R}^k)^\top)) \end{aligned} \tag{24}$$

where the first equality is based on the definition of matrix Frobenius norm $\|\mathbf{D}\|_F^2 = \mathrm{Tr}(\mathbf{D}\mathbf{D}^\top)$ with $\mathrm{Tr}(\mathbf{X})$ denoting the trace of matrix $\mathbf{X}$, and the second equality uses the properties of matrix trace, i.e., $\mathrm{Tr}(\mathbf{B}\mathbf{D}) = \mathrm{Tr}(\mathbf{D}\mathbf{B})$ and $\mathrm{Tr}(\mathbf{D}) = \mathrm{Tr}(\mathbf{D}^\top)$. Since the factor matrices in (23) must meet $\mathbf{E}_{lp} \geq 0$ and $\mathbf{A}_{pn} \geq 0$, the Lagrange multipliers corresponding to them can be defined as $\Phi_{lp} \geq 0$ and $\Psi_{pn} \geq 0$, respectively. Then, the Lagrange function is defined by

$$\begin{aligned} \Gamma &= \mathrm{Tr}(\mathbf{Y}\mathbf{Y}^\top) - 2\mathrm{Tr}(\mathbf{Y}\mathbf{A}^\top\mathbf{E}^\top) + \mathrm{Tr}(\mathbf{E}\mathbf{A}\mathbf{A}^\top\mathbf{E}^\top) + \alpha \mathrm{Tr}(\mathbf{E}\mathbf{W}\mathbf{W}^\top\mathbf{E}^\top) \\ &\quad + \frac{\alpha}{K+1} \sum_{k=1}^{K+1} (-2\mathrm{Tr}(\mathbf{E}\mathbf{W}\mathbf{W}^\top(\mathbf{R}^k)^\top)) + \mathrm{Tr}(\mathbf{R}^k\mathbf{W}\mathbf{W}^\top(\mathbf{R}^k)^\top) \\ &\quad + \mathrm{Tr}(\Phi\mathbf{E}^\top) + \mathrm{Tr}(\Psi\mathbf{A}^\top) + \beta \sum_{p=1}^{P} \sum_{n=1}^{N} \left(h_{pn} \log \frac{h_{pn}}{a_{pn}} - h_{pn} + a_{pn} \right). \end{aligned} \tag{25}$$

By taking the partial derivatives of Γ w.r.t. $\mathbf{E}$, we can obtain

$$\frac{\partial \Gamma}{\partial \mathbf{E}} = -2\mathbf{Y}\mathbf{A}^\top + 2\mathbf{E}\mathbf{A}\mathbf{A}^\top + 2\alpha\mathbf{E}\mathbf{W}\mathbf{W}^\top - \frac{2\alpha}{K+1}\sum_{k=1}^{K+1}\mathbf{R}^k\mathbf{W}\mathbf{W}^\top + \Phi. \tag{26}$$

Next, we deduce the partial derivative of Γ w.r.t. $\mathbf{A}$. Considering the last term of (25), its derivative is calculated as

$$\frac{\partial \beta D(\mathbf{H}\|\mathbf{A})}{\partial a_{pn}} = \beta\left(-\frac{h_{pn}}{a_{pn}} + 1\right). \tag{27}$$

Thus, we obtain the partial derivative $\frac{\partial \Gamma}{\partial \mathbf{A}}$ as follows:

$$\frac{\partial \Gamma}{\partial \mathbf{A}} = -2\mathbf{E}^\top\mathbf{Y} + 2\mathbf{E}^\top\mathbf{E}\mathbf{A} + \Psi - \beta(\mathbf{H} \oslash \mathbf{A} - \mathbf{Q}) \tag{28}$$

where $\mathbf{Q} \in \mathbb{R}^{P\times N}$ is an all-ones matrix. According to the KKT conditions $\Phi_{lp}\mathbf{E}_{lp} = 0$ and $\Psi_{pn}\mathbf{A}_{pn} = 0$, the following formulas concerning $\mathbf{E}_{lp}$ and $\mathbf{A}_{pn}$ can be deduced:

$$\alpha(\mathbf{E}\mathbf{W}\mathbf{W}^\top)_{lk}\mathbf{E}_{lp} + (\mathbf{E}\mathbf{A}\mathbf{A}^\top)_{lp}\mathbf{E}_{lp} = \frac{\alpha}{K+1}\sum_{k=1}^{K+1}(\mathbf{R}^k\mathbf{W}\mathbf{W}^\top)_{lp}\mathbf{E}_{lp} + (\mathbf{Y}\mathbf{A}^\top)_{lp}\mathbf{E}_{lp} \tag{29}$$

$$(\mathbf{E}^\top\mathbf{E}\mathbf{A})_{pn}\mathbf{A}_{pn} + \frac{\beta}{2}\mathbf{Q}_{pn}\mathbf{A}_{pn} = (\mathbf{E}^\top\mathbf{Y})_{pn}\mathbf{A}_{pn} + \frac{\beta}{2}(\mathbf{H}\oslash\mathbf{A})_{pn}\mathbf{A}_{pn}. \tag{30}$$

Using these two equations, the MURs of matrices $\mathbf{E}$ and $\mathbf{A}$ can be deduced as follows:

$$\mathbf{E}_{lp} \leftarrow \mathbf{E}_{lp}\frac{(\mathbf{Y}\mathbf{A}^\top)_{lp} + \frac{\alpha}{K+1}\sum_{k=1}^{K+1}(\mathbf{R}^k\mathbf{W}\mathbf{W}^\top)_{lp}}{(\mathbf{E}\mathbf{A}\mathbf{A}^\top)_{lp} + \alpha(\mathbf{E}\mathbf{W}\mathbf{W}^\top)_{lp}} \tag{31}$$

$$\mathbf{A}_{pn} \leftarrow \mathbf{A}_{pn}\frac{(\mathbf{E}^\top\mathbf{Y})_{pn} + \frac{\beta}{2}(\mathbf{H}\oslash\mathbf{A})_{pn}}{(\mathbf{E}^\top\mathbf{E}\mathbf{A} + \frac{\beta}{2}\mathbf{Q})_{pn}}. \tag{32}$$

D. Algorithm Implementation

In this section, several implementation issues of our algorithm are presented. As the first issue, we address the multiple clustering strategy adopted in the proposed algorithm. As aforementioned, each type of PES is extracted based on the corresponding cluster given by K-means, and it is expected that the mean vector of each cluster is as similar as possible to the corresponding real endmember. However, the clustering performance of K-means is clearly affected by its initial values. To tackle this problem and inspired by the hyperspectral clustering method developed in [52], in our algorithm, we use the phased unmixing results of NMF to construct effective initial values for K-means, and run it multiple times on the target HSI with the optimization progress of NMF. Specifically, let matrix $\mathbf{U} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_P]$ comprises the mean vector $\mathbf{r}_p, k = 1, 2, ..., P$ of corresponding clusters given by K-means the last time. After the two matrix factors of NMF are optimized NI times iteratively, the K-means algorithm is executed once more with its initial values taken from the columns of matrix $\frac{1}{2}(\mathbf{E} + \mathbf{U})$, by which the estimated endmembers so far and the clusters given by K-means the last time are expected to contribute equally to obtaining better initial values for K-means. Based on the clustering results, we then update $\mathbf{R}^k$, $\mathbf{W}$, as well as $\mathbf{H}$ by (12), (16), and (21), respectively. Note that when the K-means algorithm is run for the first time, we initialize it by selecting P data points uniformly at random from the range of HSI matrix $\mathbf{Y}$, that is, each of P initial values is generated by $\min(\mathbf{Y}) + \text{rand}(0,1) * (\max(\mathbf{Y}) - \min(\mathbf{Y}))$, where $\text{rand}(0,1)$ denotes the function producing the uniformly distributed random numbers in the interval [0,1].

The second issue concerns the initialization of the proposed algorithm. Given that the objective function of MCG-NMF in (23) is nonconvex, the initial values of matrices $\mathbf{E}$ and $\mathbf{A}$ will make a noticeable impact on the unmixing performance as demonstrated by relevant research. Therefore, effective initial values should be provided for these two matrices. According to the characteristics of our objective function, we initialize $\mathbf{E}$ and $\mathbf{A}$ using two methods: random initialization and clustering-based initialization. By random initialization that provides values for $\mathbf{E}$ and $\mathbf{A}$ randomly in the interval [0,1], we aim to test our algorithm on synthetic data under general conditions. In the experiments on real HSIs, MCG-NMF employs clustering-based initialization, which provides initial values for $\mathbf{E}$ and $\mathbf{A}$ as follows.

1) The clustering analysis is conducted on the HSI matrix $\mathbf{Y}$ using K-means.
2) Set $\mathbf{E} = \mathbf{U}$, where $\mathbf{U}$ comprises the mean pixels given by the first run of K-means and is constructed according to (12).
3) Using the obtained matrix $\mathbf{E}$, the initial value of $\mathbf{A}$ is provided by the fully constrained least squares (FCLS) [53] algorithm.

The third issue is about the ASC constraint in LSMM. Considering that the abundance matrix $\mathbf{A}$ is subject to the ASC constraint according to LSMM, we tackle this issue following the method in [10] and [53], by which the ASC constraint is gradually satisfied during the optimization process of MCG-NMF. Specifically, the HSI matrix $\mathbf{Y}$ and the endmember matrix $\mathbf{E}$ are, respectively, augmented as

$$\mathbf{Y}_a \leftarrow \begin{bmatrix}\mathbf{Y} \\ \delta\mathbf{1}_N^\top\end{bmatrix} \tag{33}$$

$$\mathbf{E}_a \leftarrow \begin{bmatrix}\mathbf{E} \\ \delta\mathbf{1}_P^\top\end{bmatrix} \tag{34}$$

where δ is the punishment coefficient that adjusts the impact of the ASC constraint on the objective function; $\mathbf{1}_N$ and $\mathbf{1}_P$ are all-ones vectors. Then, in each update of $\mathbf{A}$ with MUR in (32), the matrices $\mathbf{Y}_a$ and $\mathbf{E}_a$ are employed to replace $\mathbf{Y}$ and $\mathbf{E}$, respectively. Note that an appropriate value for δ should be given to balance its effects between the ASC constraint and the convergence speed of the proposed algorithm. In our experiments, it is set to 15 as suggested in [54].

The fourth issue concerns how to determine the number of endmembers P. Recently, some effective methods have been proposed to estimate the number of endmembers in HSIs, such as hyperspectral signal subspace identification by minimum error

Algorithm 1: MCG-NMF Algorithm For HU

Input: The HSI matrix $\mathbf{Y} \equiv [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{R}^{L \times N}$, the algorithm parameter NI, the regularization parameters α and β;

Output: The endmember matrix $\mathbf{E}$ and the abundance matrix $\mathbf{A}$;

1 Determine the number of endmembers P;
2 Initialize matrices $\mathbf{E}$ and $\mathbf{A}$;
3 Normalize the column of $\mathbf{A}$ to unit;
4 Augment $\mathbf{Y}$ to generate $\mathbf{Y}_a$ according to Eq. (33);
5 Conduct clustering analysis on $\mathbf{Y}$ via K-means, then obtain mean pixels $\mathbf{r}_p, p = 1, 2, ..., P$ by Eq. (11);
6 Set matrix $\mathbf{U} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_P]$;
7 Obtain $\mathbf{R}^k, k=1,...,K+1$ and $\mathbf{W}$ by Eqs. (12) and (16) respectively;
8 Generate $\mathbf{H}$ following Eq. (21);
9 Set $iterCounter = 0$;
10 **while** *the algorithm stopping criteria are not satisfied* **do**
11 **if** $iterCounter == NI$ **then**
12 Conduct clustering analysis on $\mathbf{Y}$ via K-means with initial values $\frac{1}{2}(\mathbf{E} + \mathbf{U})$;
13 Generate mean pixels $\mathbf{r}_p, p=1, ..., P$ by Eq. (11);
14 Set $\mathbf{U} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_P)$;
15 Compute $\mathbf{R}^k$, $\mathbf{W}$, as well as $\mathbf{H}$ by Eqs. (12), (16) and (21) respectively;
16 Set $iterCounter = 0$;
17 **end**
18 Update $\mathbf{E}$ according to Eq. (31);
19 Augment $\mathbf{E}$ to obtain $\mathbf{E}_a$ following Eq. (34);
20 Update $\mathbf{A}$ by Eq. (32);
21 $iterCounter = iterCounter + 1$;
22 **end**

(HySime) [55] and virtual dimensionality (VD) [56]. The K-means method needs to know the number of clusters in advance. The value of P estimated by HySime or VD can also be used as the number of clusters in K-means. However, in our experiments, the number of endmembers is assumed to be known *a priori* [10], [11].

The fifth issue is about how to determine the appropriate value of K, which remains a problem when the K-nearest neighbors of the mean vector in each cluster are selected. Recall that we regard the K-nearest neighbors of the mean vector as PESs of some sort of endmember. If a smaller value of K is adopted, the obtained PESs may lack accuracy owing to the noise interference in the data. However, too big a value of K will cause the selected PESs to be scattered, especially for HSIs with a high mixing degree. In order to make a better balance between these two factors, we set $K = 4$ in the experiments of this article.

The last issue concerns the termination criteria of the proposed algorithm. We use two stopping criteria in our algorithm: 1) the error tolerance of objective function $F(\mathbf{E}, \mathbf{A})$ in successive iterations, i.e.,

$$\frac{|F(\mathbf{E}^{(t)}, \mathbf{A}^{(t)}) - F(\mathbf{E}^{(t-1)}, \mathbf{A}^{(t-1)})|}{F(\mathbf{E}^{(t-1)}, \mathbf{A}^{(t-1)})} \tag{35}$$

falls below a predefined threshold 10^{-4} in ten successive iterations; 2) the maximum number of iterations reaches a prespecified value, which is set to 3000 in our experiments. With these termination criteria, our algorithm will iteratively optimize the two matrices $\mathbf{E}$ and $\mathbf{A}$ until either of the two conditions is reached.

Based on the above implementation techniques, the pseudocode of the proposed MCG-NMF algorithm is given in Algorithm 1. As can be seen from Algorithm 1, it mainly consists of two parts: the initial part and the iteration part. The initial part includes lines 1–9, by which we aim to initialize some matrices needed by the iteration part. Based on lines 1–4, our algorithm can determine parameter P, initial matrices $\mathbf{E}$ and $\mathbf{A}$, as well as construct matrix $\mathbf{Y}_a$. In line 5, Algorithm 1 runs K-means for the first time and the initial values for K-means are generated by selecting P data points uniformly at random from the range of HSI matrix Y as aforementioned. Based on the clustering results via K-means, some matrices, such as $\mathbf{U}$, $\mathbf{R}^k$, $\mathbf{W}$, and $\mathbf{H}$, are obtained by lines 5–8, respectively. To count the number of iterations of the algorithm, we initialize a count variable $iterCounter$ in line 9. Next, in the iteration part, Algorithm 1 will iteratively update matrices $\mathbf{E}$ and $\mathbf{A}$ by lines 18–20 until the number of iterations reach to a prespecified amount NI. In this case, our algorithm will run lines 12–16, by which K-means is executed once more based on the new initial values given by $\frac{1}{2}(\mathbf{E} + \mathbf{U})$ according to line 12. Then, by lines 13–15, the values for matrices $\mathbf{U}$, $\mathbf{R}^k$, $\mathbf{W}$, and $\mathbf{H}$ are updated based on the clustering results of K-means. In line 16, the algorithm reinitializes the variable $iterCounter = 0$ so that it can run lines 12–16 once again after lines 18–21 are executed NI times.

E. Complexity Analysis

Before presenting the experimental results, we briefly discuss the computational complexity of MCG-NMF. According to Algorithm 1, its computational cost mainly concerns the algorithm steps of conducting clustering analysis via K-means, computing $\mathbf{R}^k$ and $\mathbf{H}$, as well as updating $\mathbf{E}$ and $\mathbf{A}$ in the main loop. Given the dimension of pixel spectrum L, the total number of pixels N, and the endmember numbers P, the computation cost involved by the K-means method is $O(tLPN)$, where t is the iteration numbers of K-means. Considering the computation of $\mathbf{R}^k$, $k = 1, 2, ..., K$, by (12) primarily concerns finding the K-nearest neighbors of each mean pixel in the corresponding cluster, it needs the time complexity of $O(LN + KN)$. The computation of the matrix $\mathbf{H}$ primarily consists of two parts: namely the least-squares method used in (18) and the weight mean matrix calculated by (21) with matrix addition and multiplication operations, which have the time complexity of $O(LPN + P^2N)$ and $O(PN + (K + 1)LP)$, respectively. In the steps of updating $\mathbf{E}$ and $\mathbf{A}$, it mainly involves the matrix operations of addition and multiplication with the order of $O(N(LP + P^2))$. Recall that the K-means is executed once after matrix factors $\mathbf{E}$ and $\mathbf{A}$ are optimized NI times iteratively. Taking all the operations and the range of values of each variable into consideration, the overall complexity of Algorithm 1 is $O(tLPN)$ per iteration in

the worst case. In addition, an empirical runtime comparison is given in Section IV-C.

IV. Experimental Results and Analysis

To evaluate the effectiveness of our method, we conduct a series of experiments on synthetic datasets and two real-world HSIs. In the experiments on synthetic data, the proposed method is compared with four representative NMF-based unmixing methods, namely, GLNMF, NMF-PPK, SGSNMF, and MVC-NMF. Among them, GLNMF is a graph regularized NMF that jointly enforces sparseness and manifold constraints on the abundance matrix. In the NMF-PPK method, *a priori* information for part of endmembers is used as semisupervised guidance to enhance the unmixing performance. SGSNMF is a method based on spatial information, which imposes group-sparseness constraint on abundances. MVCNMF utilizes the geometrical property of HSIs to improve the unmixing performance of NMF, and it attempts to minimize the volume of the simplex enclosed by the candidate endmembers in the unmixing procedure. The algorithm parameters for each of state-of-the-art method are set to the same values as those in their original works. To compare the performance of all the algorithms, each is run twenty times and the average performance metrics are adopted. Since NMF-PPK needs the prior knowledge of partial endmembers, we randomly choose two endmembers as prior information in each run. In addition, the initial values of all the algorithms are chosen randomly in the interval [0,1], except for SGSNMF, which includes a special initial method. In the experiments on real HSIs, our method is tested on the widely used Washington DC Mall and Cuprite data. All the algorithms used in the experiments were implemented with MATLAB R2016b version and executed on a computer with 1.99 GHz Intel(R) Core(TM) i7 CPU, 16 GB main memory and Windows 10 operating system.

A. Performance Metrics

Two performance metrics, i.e., spectral angle distance (SAD) and root mean square error (RMSE) [10] are adopted to evaluate the unmixing accuracy of our algorithm. The SAD criterion is used to measure the similarity between the estimated endmember signature $\hat{\mathbf{e}}_p$ and the reference signature $\mathbf{e}_p$, which is defined as

$$\mathrm{SAD}_p = \arccos\left(\frac{\mathbf{e}_p^\top \hat{\mathbf{e}}_p}{\|\mathbf{e}_p\|\|\hat{\mathbf{e}}_p\|}\right) \tag{36}$$

where $\|\cdot\|$ denotes the Euclidean norm of the vector.

To measure the similarity between the estimated abundances and the reference abundances, we adopt the RMSE criterion, which is defined as

$$\mathrm{RMSE}_p = \sqrt{\frac{1}{N}\|\mathbf{A}_{p:} - \hat{\mathbf{A}}_{p:}\|_2^2} \tag{37}$$

where $\mathbf{A}_{p:}$ and $\hat{\mathbf{A}}_{p:}$ denote the reference abundance distributions and the estimated ones corresponding to the pth endmember, respectively. Note that in the experiments on synthetic data, the mean SAD and RMSE values over all the endmembers are used to test all the algorithms.

B. Experiments on Synthetic Data

In our experiments, the synthetic datasets are generated by five spectral signatures chosen from the United States Geological Survey (USGS) spectral library. Each of the selected spectra includes 224 spectral bands with wavelengths varying from 0.38 to 2.5 μm. Fig. 2 shows the signatures of all the selected spectra. These spectra are linearly independent and used as endmembers to create the synthetic data. By making some modifications to the procedure in [10], we generate a set of synthetic datasets as follows.

1) A $b^2 \times b^2$ synthetic image is divided into b^2 blocks, each of which has $b \times b$ pixels.
2) All pixels within each block are assigned a kind of endmember randomly selected from the endmember set.
3) A $(b+1) \times (b+1)$ averaging filter is applied to the image, by which the pixels are linearly mixed and the abundances of each pixel are obtained.
4) For each pixel of the synthetic image, if one of its fractional abundances is larger than a specified threshold θ, this pixel will be replaced by a new pixel that is mixed by randomly selected four endmembers in equal proportions. In this way, the synthetic images are highly mixed without the existence of pure pixels. In addition, the abundance fractions of each pixel are normalized to unity to satisfy with the ASC constraint.
5) The zero-mean white Gaussian noise is added to the obtained images according to the prespecified signal-to-noise ratio (SNR).

Experiment 1 (Select the Algorithm Parameters): The purpose of this experiment is to select the proper values for the algorithm parameters of MCG-NMF. The setting of NI, α and β are, respectively, tested using synthetic data with 64×64 pixels, which is generated by five endmembers with $\theta = 0.7$ and SNR= 30 dB. We first hold α and β to be fixed and choose different values for NI from the set $\{10, 20, 30, 40, 50, 60, 80, 100, 200, 300, 400, 500, 600\}$. As shown in Fig. 3(a), the algorithm has better SAD and RMSE when NI varies from 80 to 300, and $NI = 100$ is the optimal selection.

Next, the influence of parameter α on the algorithm performance is examined on the condition that $NI = 100$ and $\beta = 0.001$ whilst selecting different values for α in the collection $\{1e-3, 5e-3, 0.01, 0.05, 0.1, 0.5, 1, 1.5, 2, 2.5, 3\}$. Based on the results shown in Fig. 3(b), we can see that MCG-NMF achieves better performance w.r.t. the SAD criterion when α is changed from 0.1 to 2.5. For the RMSE criterion, MCG-NMF can perform better when α is in the interval of [0.5, 2.5]. It is not hard to verify that our algorithm can perform best when $\alpha = 1$. To find the proper value for parameter β, we fix $NI = 100$ and $\alpha = 1$, and then select different values for β from the set $\{1e-5, 5e-5, 1e-4, 5e-4, 1e-3, 5e-3, 0.01, 0.05, 0.1, 0.5, 1\}$. Fig. 3(c) shows the influence of β on the performance. Note that the values of harmonic mean (HM) w.r.t. SAD and RMSE are also plotted in Fig. 3(c) for the sake of clarity. As can be seen from HM values, our algorithm performs best in case of $\beta = 1e-3$. Based on these analysis, $NI = 100$, $\alpha = 1$ and $\beta = 1e-3$ are used in our experiments.

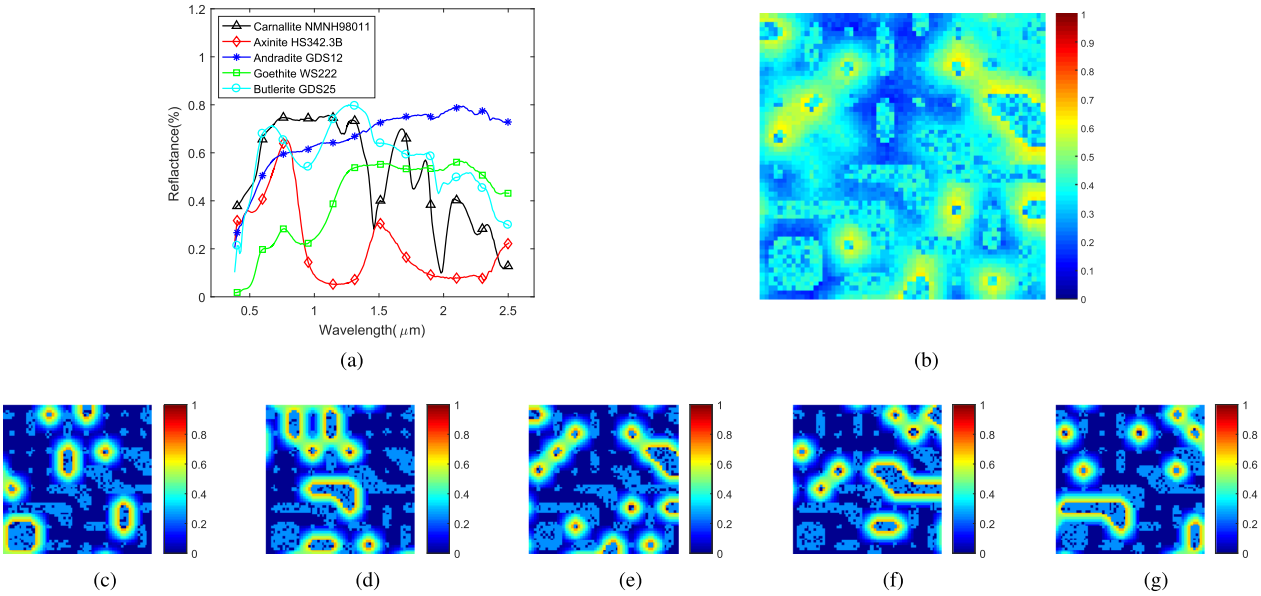


Fig. 2. Examples of the synthetic data. (a) Endmembers used to generate the synthetic data. (b) 188th band of a 64×64 synthetic data with SNR of 30 dB. (c)–(g) Abundance map corresponding to each endmember.

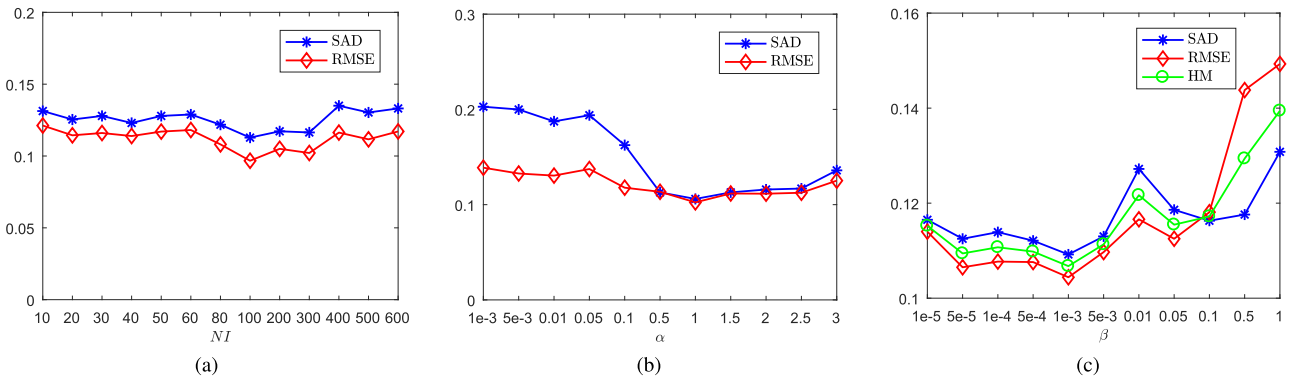


Fig. 3. Performance of MCG-NMF with respect to algorithm parameters *NI*, α, and β. (a) Parameter *NI*. (b) Parameter α. (c) Parameter β.

Experiment 2 (Robustness to Noise): In this experiment, we aim to test the robustness of MCG-NMF on the data with different levels of SNR. The datasets employed here are generated under the same conditions as the previous experiment except that their SNRs vary from 15 to 35 dB at an interval of 5 dB. Additionally, the data with SNR at infinity are also used to test MCG-NMF under the noise-free condition. Fig. 4(a) and (b) shows the performance of all the algorithms according to SAD and RMSE, respectively. We can see that MCG-NMF outperforms other baseline algorithms on all of the datasets with significant performance improvement for both endmembers and abundances. As is expected, both SAD and RMSR of all the algorithms get worse with decreasing SNR, especially for the SAD values of MVCNMF and SGSNMF. This indicates noise interference is an important factor affecting the performance. Moreover, compared with the other methods, MCG-NMF shows more stable performance by providing similar SAD and RMSE values, respectively, on datasets with different levels of noise. This demonstrates that our method has better robustness to noise disturbance.

Experiment 3 (Generalization to Mixing Degree): Our goal here is to investigate the generalization capability of MCG-NMF on the datasets with different mixing degrees. We generate the desired data by adjusting θ from 0.5 to 0.9 at an interval of 0.1 and fixing SNRs of all the data to be 30 dB. The performance of all the methods on different datasets is given in Fig. 5. Note that the term "purity" in Fig. 5 is used to indicate different mixing degrees of the data. We can see that the performance of all the methods gets better with the increase of purity. According to the SAD measures, MVCNMF shows the worst performance on all datasets, whereas MCG-NMF performs better than the others when the data have higher purity, such as 0.7 to 0.9. On the datasets with lower purity, such as 0.7, our method has comparative SAD with GLNMF; it shows slightly worse performance than GLNMF when the purity is 0.5. As for the RMSE values, it can be seen from Fig. 5(b) that MCG-NMF is

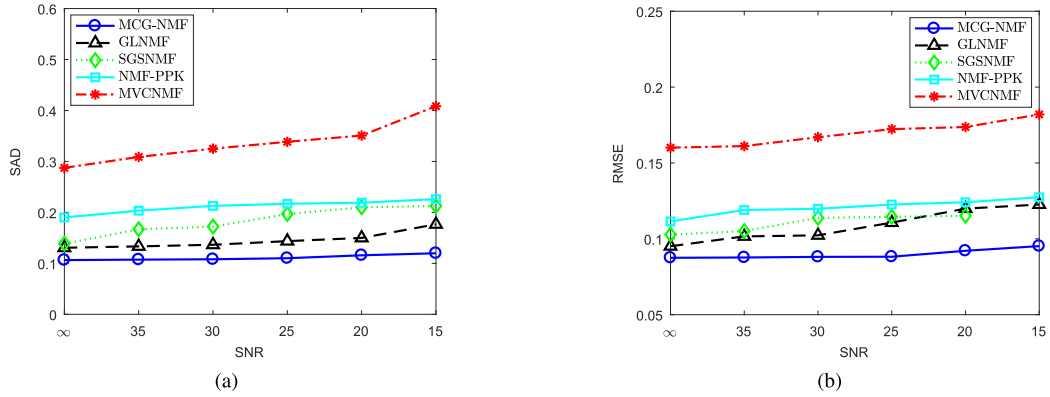


Fig. 4. Performance of MCG-NMF on synthetic datasets with different levels of SNR. (a) SAD. (b) RMSE.

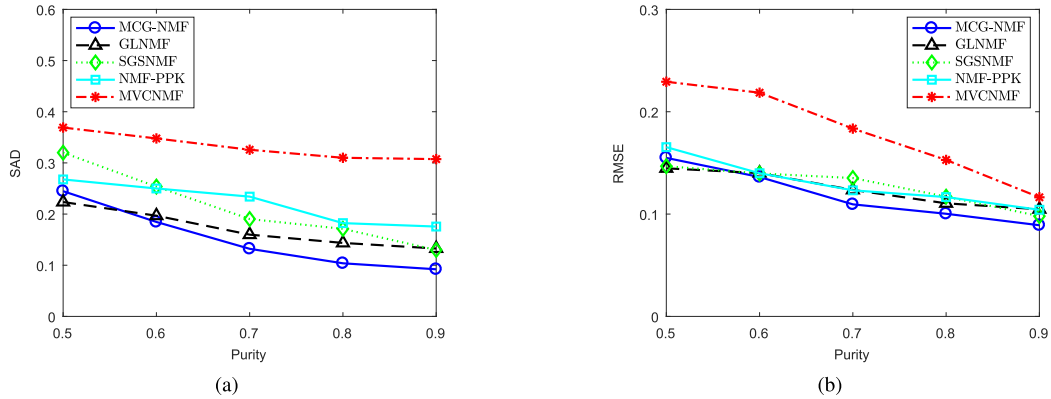


Fig. 5. Performance of MCG-NMF on synthetic datasets with different mixing degrees. (a) SAD. (b) RMSE.

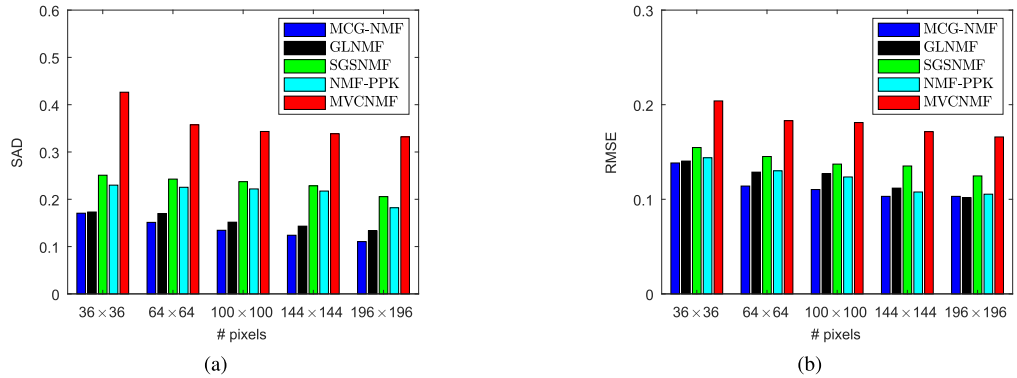
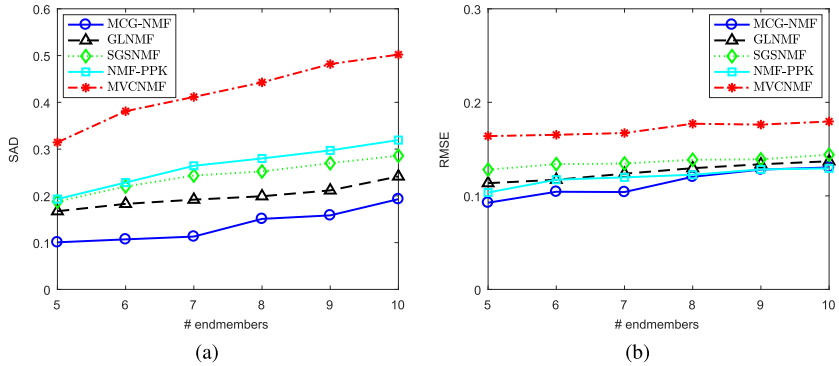


Fig. 6. Performance of MCG-NMF on synthetic datasets with different number of pixels. (a) SAD. (b) RMSE.

better than the other methods when the purity varies from 0.7 to 0.9. Meanwhile, it has comparative performance with GLNMF, SGSNMF, NMF-PPK when purity equals to 0.6. On data with purity 0.5, our method performs slightly worse than GLNMF and SGSNMF. On the whole, MCG-NMF has better generalization ability on datasets with different mixing degrees.

Experiment 4 (Sensitivity to Pixel Numbers): The purpose of this experiment was to analyze the sensitivity of MCG-NMF to the number of pixels in the data. We used datasets with the pixel numbers of 36×36, 64×64, 100×100, 144×144, and 196×196, respectively. The other generating parameters of these datasets remain the same as those employed in Experiment 1. The SAD and RMSE values of all the algorithms are shown in Fig. 6(a) and (b), respectively. As can be seen, the unmixing accuracy increases when the data have more pixels, which is rational since NMF-based methods can obtain more accurate statistical information as the sample size increases. By comparing all algorithms according to RMSE values, we



(a) (b)

Fig. 7. Performance of MCG-NMF on synthetic datasets with different number of endmembers. (a) SAD. (b) RMSE.

TABLE I
MEANS AND STANDARD DEVIATIONS (%) OF THE SAD ON WASHINGTON DC MALL DATASET

Endmembers	MCG-NMF	GLNMF	NMF-PPK	SGSNMF	MVCNMF	SDNMF-TV	MVNTF	VCA
Tree	**0.0486 ± 0.03**	0.1384 ± 1.41	0.1858 ± 4.90	0.1621 ± 2.68	0.2372 ± 0.89	0.0827 ± 4.22	0.2251 ± 0.38	0.1937 ± 1.75
Grass	**0.0824 ± 0.01**	0.3455 ± 3.80	0.2181 ± 11.15	0.1485 ± 2.50	0.2571 ± 6.64	0.2149 ± 6.43	0.2426 ± 5.94	0.2912 ± 5.64
Street	0.3760 ± 0.01	0.4094 ± 4.66	0.3454 ± 5.65	0.6139 ± 17.21	0.5859 ± 17.18	**0.3423 ± 11.38**	0.3582 ± 4.25	0.3612 ± 4.15
Roof	**0.1160 ± 0.04**	0.1409 ± 7.07	0.2252 ± 4.96	0.2520 ± 1.56	0.1663 ± 4.00	0.3455 ± 5.14	0.1505 ± 7.91	0.2087 ± 5.22
Water	0.1021 ± 0.01	0.0734 ± 0.67	0.0925 ± 4.33	0.0970 ± 1.16	0.1117 ± 1.97	0.0961 ± 3.87	0.1040 ± 0.68	**0.0438 ± 0.09**
Average	**0.1450 ± 0.02**	0.2215 ± 3.52	0.2134 ± 6.20	0.2547 ± 5.02	0.2716 ± 6.14	0.2163 ± 6.21	0.2161 ± 3.83	0.2197 ± 3.37

can see that our method has the comparative performance with GLNMF on datasets with the size of 36×36 and 196×196 pixels. However, it outperforms all the baseline algorithms on the other datasets. According to the SAD values, the performance of our method is comparable with GLNMF on data with 36×36 pixels, whereas our method shows obvious superiority with the increasing of the number of pixels. This can be attributed to the fact that the adopted multiple clustering strategy can provide more accurate guidance for the endmember estimations when the data comprise more pixels.

Experiment 5 (Effectiveness to Endmember Numbers): Here, the performance of the five algorithms is evaluated on the datasets generated by mixing a different number of endmembers from five to ten. Apart from the endmember numbers, all the synthetic data have the same size of 64×64 pixels, and are created with $\theta = 0.7$ and SNR$= 30$ dB. Fig. 7(a) and (b) shows the performance of the five algorithms w.r.t. SAD and RMSE criteria, respectively. We can see that the performance of all the algorithms will degrade when the number of endmembers increases in the data. As shown in Fig. 7(a), it is apparent that MVCNMF gives the worst performance on all the datasets and MCG-NMF performs better than the other alternative methods. As regards the RMSE values, we can see from Fig. 7(b) that MVCNMF performs worse than the other four methods, whereas MCG-NMF can perform best when the data contain a relatively small number of endmembers, such as 5–7. When the datasets have more endmembers, e.g., 8–10, MCG-NMF can show an RMSE performance comparable with NMF-PPK, GLNMF, and SGSNMF. On the whole, MCG-NMF can provide effective performance on the datasets with different numbers of endmembers.

Experiment 6 (Convergence on Different Datasets): In this experiment, we investigate the convergence of our method by testing it on datasets with different sizes. We adopt the same data as those used in the previous experiment. Fig. 8 plots the curves of reconstruction error between the original data and the reconstructed ones given by our algorithm. By comparing the curves shown in Fig. 8, we can see that MCG-NMF converges faster on datasets with smaller size. This is reasonable considering that fewer components need to be estimated in the abundance

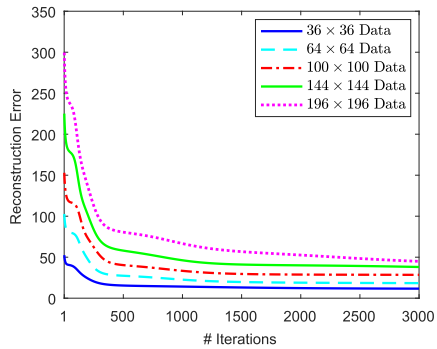


Fig. 8. Convergence curves of MCG-NMF on datasets with different sizes.

Fig. 9. Pseudocolor image of the HYDICE Washington DC Mall data.

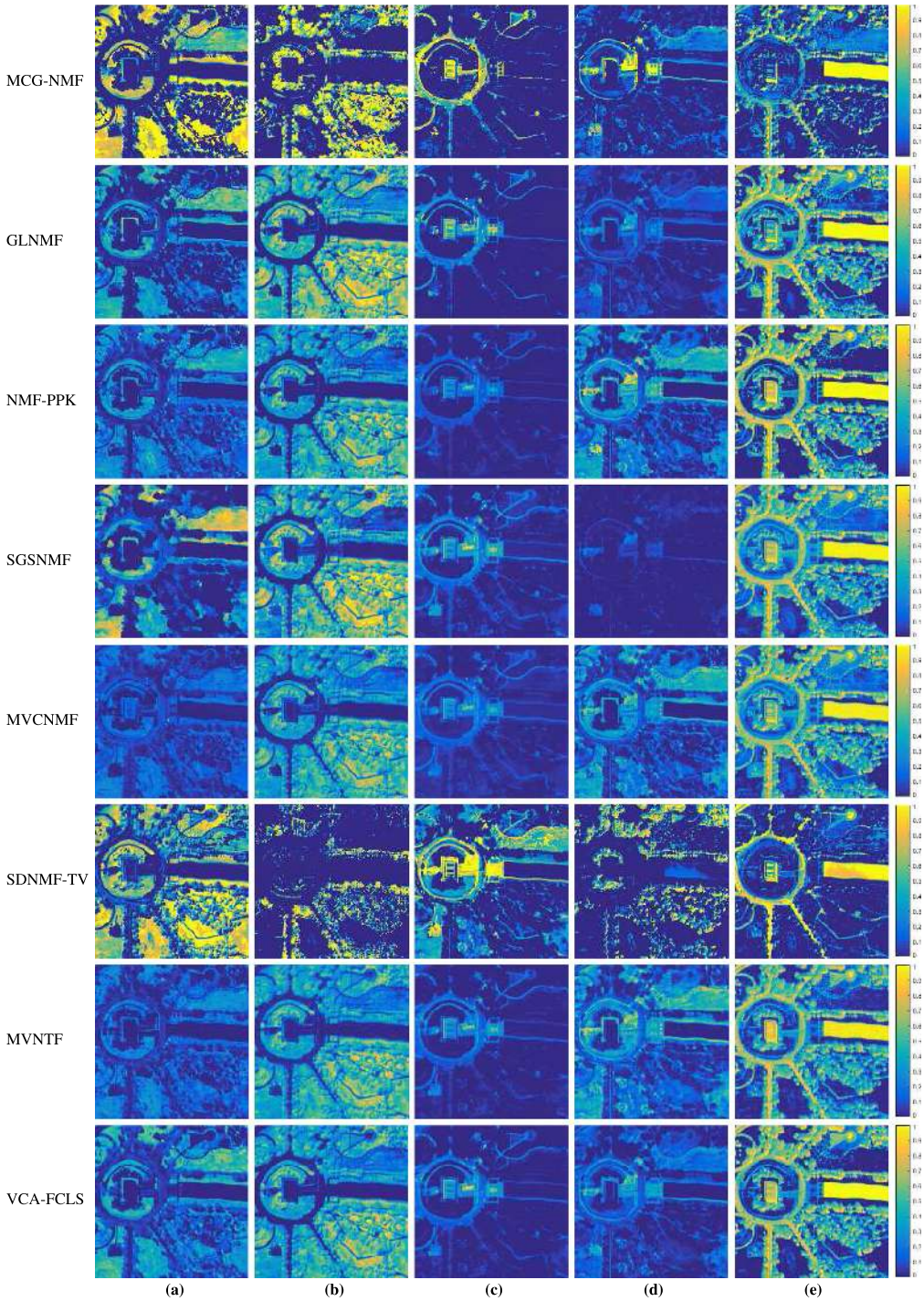


Fig. 10. Abundance maps estimated by MCG-NMF, GLNMF, NMF-PPK, SGSNMF, MVCNMF, SDNMF-TV, MVNTF, and VCA-FCLS on Washington DC Mall data. (a) Grass. (b) Tree. (c) Roof. (d) Street. (e) Water.

matrix when MCG-NMF is tested on the data with smaller size. It is easy to see that for datasets with a size of 36×36 and 64×64 pixels, respectively, MCG-NMF can converge within 1000 iterations. When it runs on the data with moderate size, such as 100×100 and 144×144 pixels, MCG-NMF will iterate about 1500 times before its convergence. On datasets of larger size, e.g., 196×196 pixels, 2500 iterations are needed to reach its convergence.

C. Experiments on Real Hyperspectral Data

In this section, the proposed MCG-NMF method is applied to two widely used real-world HSIs: the Washington DC Mall dataset and the Cuprite dataset. We compare the performance of our method with seven representative methods: GLNMF, NMF-PPK, SGSNMF, MVCNMF, sparsity-constrained deep NMF with total variation (SDNMF-TV) [57], MVNTF, and

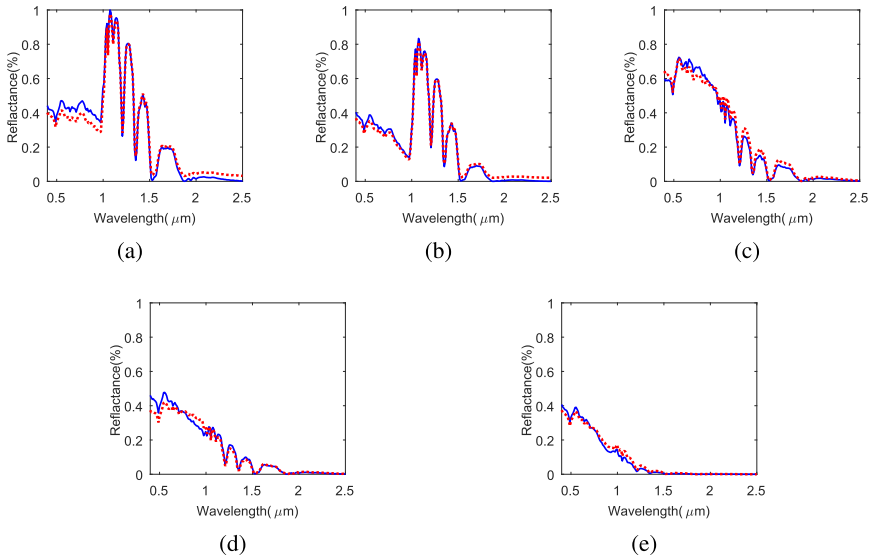


Fig. 11. Comparison between the reference signatures of endmembers (blue solid line) and the estimated ones (red dotted line) given by MCG-NMF on Washington DC Mall data. (a) Grass. (b) Tree. (c) Roof. (d) Street. (e) Water.

VCA [5] followed by FCLS [53] (VCA-FCLS). The first four are NMF-based methods and have been used in the experiments with synthetic data. SDNMF-TV is a recently proposed deep NMF unmixing method with the sparsity and piecewise smoothness constraints enforced on the abundances. MVNTF performs the unmixing tasks under the framework of nonnegative tensor factorization. VCA is a method based on the geometrical characteristic of HSIs. It can only be used to extract endmembers, and corresponding abundances are next estimated by the FCLS approach. In our experiments, average results are obtained by running each algorithm twenty times. The algorithm parameters for all baseline methods adopt the same values as those in their original works. Additionally, in each run of NMF-PPK on the Washington DC Mall data, two endmembers are randomly selected as known in advance, whereas on the Cuprite data, three endmembers are enumerated randomly as prior information.

1) Washington DC Mall Dataset: The first real HSI is the Washington DC Mall dataset collected by the airborne HYperspectral Digital Imagery Collection Experiment (HYDICE) sensor over the Washington DC Mall area. Recently, this datum has been extensively used in experiments of HU. The original data have 1280×307 pixels with 210 bands covering wavelengths from 0.4 to 2.5 μm. Their spatial and spectral resolutions are approximately 2.8 m and 10 nm, respectively. A subimage with 150×150 pixels from the original data is used in the experiments, and Fig. 9 shows its pseudocolor image. Here, we assume there are mainly five kinds of endmembers contained in the target area including tree, grass, roof, water, and street [58], [59]. Before unmixing, the bands corrupted by noise seriously, as well as the water vapor absorption bands (including bands 103–106, 138–148, and 207–210) were both removed from the data and 191 bands remained. In the experiments, the reference endmembers are generated by manually choosing some pixels from the original data cube and their positions are marked with red circles in Fig. 9.

Fig. 12. False-color image of the selected subimage from AVIRIS Cuprite data.

To evaluate the accuracy of endmember extraction quantitatively, the SAD values of all the algorithms are listed in Table I. It can be seen from Table I that MCG-NMF shows the best performance by achieving the lowest SAD for three cases, as well as the lowest average SAD value. SDNMF-TV and VCA only provide one case of the lowest SAD, respectively. We can see that MCG-NMF can achieve much lower standard deviations for both single endmember and the average value w.r.t. all the endmembers. For comparison purposes, the visible abundance maps obtained by our method and the other seven baselines are shown in Fig. 10. In addition, Fig. 11 illustrates the estimated endmember signatures by MCG-NMF and the reference signatures. It can be seen that the estimated endmembers are in good accordance with the referenced ones. It demonstrates that MCG-NMF can provide satisfactory endmember estimations.

2) Cuprite Dataset: The second real HSI used in our experiment is the Cuprite data that was captured on June 19, 1997 by

TABLE II
MEANS AND STANDARD DEVIATIONS (%) OF THE SAD ON CUPRITE DATASET

Endmembers	MCG-NMF	GLNMF	NMF-PPK	SGSNMF	MVCNMF	SDNMF-TV	MVNTF	VCA
Alunite	**0.0903 ± 0.03**	0.1560 ± 6.35	0.1269 ± 2.59	0.1500 ± 3.15	0.1832 ± 9.39	0.2588 ± 3.01	0.1407 ± 3.30	0.1367 ± 4.27
Andradite	0.0757 ± 0.27	0.0683 ± 0.83	0.0865 ± 4.18	0.0824 ± 1.08	0.1050 ± 7.75	**0.0646 ± 0.81**	0.1065 ± 6.07	0.0910 ± 1.52
Buddingtonite	**0.1081 ± 1.56**	0.1251 ± 3.10	0.1461 ± 4.72	0.1175 ± 2.32	0.1304 ± 6.49	0.1269 ± 2.01	0.1301 ± 3.97	0.1410 ± 1.95
Dumortierite	**0.0780 ± 0.06**	0.1104 ± 7.14	0.1009 ± 1.65	0.0956 ± 1.58	0.1446 ± 4.69	0.0818 ± 0.23	0.1005 ± 1.46	0.1151 ± 1.99
Kaolinite	0.2332 ± 1.21	0.2787 ± 4.24	0.2754 ± 6.68	0.2766 ± 3.35	0.2868 ± 6.11	0.3135 ± 1.13	0.2592 ± 4.35	**0.2326 ± 2.87**
Montmorillonite	**0.0931 ± 1.16**	0.0997 ± 4.33	0.1280 ± 3.09	0.1218 ± 0.62	0.1457 ± 5.39	0.1835 ± 4.40	0.1267 ± 3.72	0.1414 ± 2.37
Muscovite	0.1705 ± 0.98	0.2148 ± 3.29	0.1961 ± 4.27	0.1936 ± 2.90	0.2630 ± 6.48	**0.1548 ± 1.40**	0.2292 ± 2.76	0.1998 ± 2.96
Nontronite	0.0811 ± 0.13	0.0882 ± 0.50	**0.0807 ± 1.33**	0.0914 ± 0.78	0.1127 ± 2.64	0.0850 ± 0.15	0.0854 ± 1.11	0.1033 ± 1.32
Pyrope	0.0651 ± 0.10	0.0633 ± 0.47	0.0897 ± 4.77	0.0930 ± 2.82	0.0709 ± 1.26	**0.0618 ± 0.24**	0.0928 ± 4.97	0.1378 ± 4.70
Sphene	0.1573 ± 0.76	0.1617 ± 6.32	0.1365 ± 6.74	0.1324 ± 5.25	0.0639 ± 1.70	0.0880 ± 0.75	**0.0571 ± 0.38**	0.1087 ± 3.71
Average	**0.1152 ± 0.63**	0.1366 ± 3.66	0.1367 ± 4.00	0.1354 ± 2.39	0.1506 ± 5.19	0.1419 ± 1.41	0.1328 ± 3.21	0.1407 ± 2.77

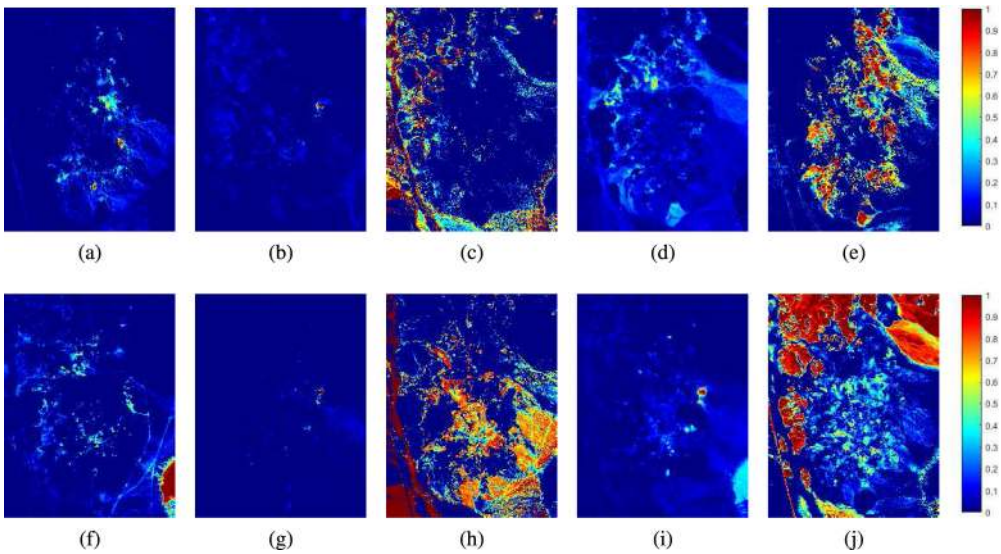


Fig. 13. Abundance maps estimated by MCG-NMF on Cuprite data. (a) Alunite. (b) Andradite. (c) Buddingtonite. (d) Dumortierite. (e) Kaolinite. vv (f) Montmorillonite. (g) Muscovite. (h) Nontronite. (i) Pyrope. (j) Sphene.

the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over Cuprite, NV, USA. These data include 224 bands with the wavelength from 0.4 to 2.5 μm. Their spectral and space resolutions are approximately 10 nm and 20 m, respectively. Since Cuprite is a mining area, rich minerals, such as alunite, nontronite, kaolinite, etc., are widely distributed in the scene, and the minerals here are commonly highly mixed. Therefore, these data are frequently used to test the HU methods. Also, ground truth and a geological summary of this site were provided in past literature [60]. We selected a subimage with 250×191 pixels from the original HSI, and the false-color image of it is shown in Fig. 12. Before the experiment, the bands with low SNR and water vapor absorption bands (including bands 1-2, 104-113, 148-167, and 221-224) are deleted, and the remaining 188 bands were used.

In our experiment, ten types of main minerals are assumed to be present in the target scene in accordance with the existing literature [5], [10]. The reference signatures used to evaluate the unmixing performance of our method are chosen from the USGS mineral spectral library [5]. Table II lists the SAD values given by all algorithms. The best performance is marked in bold font. According to the SAD values in Table II, MCG-NMF has the largest number of best performance instances. When comparing all the algorithms based on the average values w.r.t. all the endmembers, our method also shows the best performance by providing the lowest SAD. MCG-NMF shows better performance improvement by giving a significantly smaller mean standard deviation, which indicates the better robustness of the proposed method. In addition, the abundance distributions of all the endmembers estimated by MCG-NMF are shown as color abundance maps in Fig. 13, in which the higher color temperature of a pixel indicates the larger proportion of existence of corresponding endmembers. Besides, Fig. 14 illustrates the comparison between the endmember signatures estimated by MCG-NMF and the referenced ones. We can see that the estimated signatures of Alunite, Buddingtonite, Dumortierite, Montmorillonite, Nontronite, as well as Sphene, are in better accordance with the referenced ones.

3) Running Time on Two Real Datasets: To validate the efficiency of our algorithm on computation time, we tested all the algorithms on Cuprite and Washington DC Mall data. Table III gives the average running time of them. From Table III, we can see that VCA needs much less running time compared with the other methods since it is a geometric projection-based method that needs fewer iterations. Among the methods based on NMF and NTF, MVNTF needs the most running time, followed

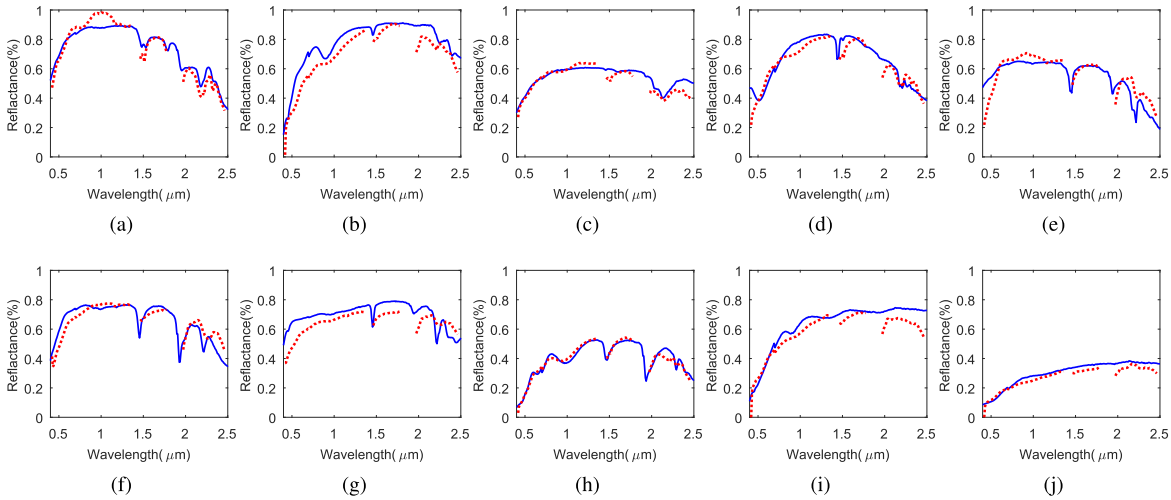


Fig. 14. Comparison between the reference endmembers (blue solid line) and the estimated ones (red dotted line) given by MCG-NMF on Cuprite data. (a) Alunite. (b) Andradite. (c) Buddingtonite. (d) Dumortierite. (e) Kaolinite. (f) Montmorillonite. (g) Muscovite. (h) Nontronite. (i) Pyrope. (j) Sphene.

TABLE III
RUNNING TIME ON THE REAL DATASETS (IN SECONDS)

Data sets	MCG-NMF	GLNMF	NMF-PPK	SGSNMF	MVCNMF	SDNMF-TV	MVNTF	VCA
Washington DC Mall	152.87	95.83	114.18	133.71	1060.70	877.26	1138.58	0.14
Cuprite	437.45	244.21	162.20	167.35	2984.20	2507.18	3972.90	0.40

by MVCNMF, on both datasets. For the Washington DC Mall data, the running time of GLNMF is the least, and our method requires a slightly longer time than SGSNMF. As to Cuprite data, NMF-PPK and SGSNMF consume less time than the other factorization-based algorithms. Although MCG-NMF demands more time than GLNMF, it performs much better than MVNTF, MVCNMF, and SDNMF-TV.

V. CONCLUSION

In this article, an MCG-NMF method for HU is proposed under a self-supervised framework. By conducting clustering analysis on the HSI, the PESs implied in the data, as well as the PADs for each endmember, are obtained as guidance information to construct effective regularization for an NMF-based unmixing model. Multiple clustering that can interact with the unmixing model is incorporated into the optimization process of NMF. In this way, more accurate PESs and abundance distributions can be learned in a self-supervised manner, and help the estimated endmembers and abundances to be more consistent with the real ones. In addition, the proposed algorithm not only has strong generalization capability for data with different mixing degree and pixel numbers but also better robustness to noise disturbance. With a series of experiments on synthetic data and real hyperspectral images, the effectiveness of the proposed algorithm has been demonstrated by comparing it with some state-of-the-art algorithms.

Two issues are worth studying in further work. First, because of the LSMM model adopted in this work, the unmixing performance of the proposed method maybe degraded when an HSI includes a number of pixels with an intimate mixture. In this circumstance, the unmixing approach with a NLSMM model is more suitable. Therefore, in a future study, we will research nonlinear unmixing methods with semisupervised information based on effective nonlinear techniques, such as the kernel method. Next, in this work, the PES and the PAD are estimates based on the results of multiple clustering, which only exploits the spectral information of an HSI without considering its spatial contexture information. In further research, both the spectral and spatial information will be jointly utilized to learn the semisupervised information of endmembers and abundances via effective techniques, e.g., superpixel segmentation. This is expected to bring more accurate guidance for NMF and thus provide superior unmixing performance.

REFERENCES

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
[3] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–275, 1999.
[4] J. W. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop*, Pasadena, CA, USA, 1993, vol. 1, pp. 11–14.
[5] J. M. P. Nascimento and J. M. 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.
[6] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
[7] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, T. K. Leen, T. G.Dietterich, and V. Tresp, Eds. Cambridge, MA, USA: MIT Press, 2001, pp. 556–562.

[8] Y. Wang and Y. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 6, pp. 1336–1353, Jun. 2013.

[9] 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.

[10] 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.

[11] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[12] X. Wang, Y. Zhong, L. Zhang, and Y. Xu, "Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6287–6304, Nov. 2017.

[13] X. Liu, W. Xia, B. Wang, and L. 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.

[14] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.

[15] Y. Yuan, M. Fu, and X. Lu, "Substance dependence constrained sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 2975–2986, Jun. 2015.

[16] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, "Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1776–1792, Mar. 2017.

[17] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2341–2357, Apr. 2019.

[18] B. Feng and J. Wang, "Constrained nonnegative tensor factorization for spectral unmixing of hyperspectral images: A case study of urban impervious surface extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 583–587, Apr. 2019.

[19] T. Imbiriba, R. A. Borsoi, and J. C. M. Bermudez, "Low-rank tensor modeling for hyperspectral unmixing accounting for spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1833–1842, Mar. 2020.

[20] L. Sun, F. Wu, T. Zhan, W. Liu, J. Wang, and B. Jeon, "Weighted nonlocal low-rank tensor decomposition method for sparse unmixing of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1174–1188, Mar. 2020.

[21] W. Tang, Z. Shi, and Z. An, "Nonnegative matrix factorization for hyperspectral unmixing using prior knowledge of spectral signatures," *Opt. Eng.*, vol. 51, no. 8, pp. 087001-1–087001-10, Aug. 2012.

[22] L. Tong, J. Zhou, Y. Qian, X. Bai, and Y. Gao, "Nonnegative-matrix-factorization-based hyperspectral unmixing with partially known endmembers," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6531–6544, Nov. 2016.

[23] N. Wang, L. Zhang, Y. Cen, and Q. Tong, "A non-negative matrix factorization approach for hyperspectral unmixing with partial known endmembers," in *Proc. 8th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Los Angeles, CA, USA, Aug.21–24, 2016, pp. 1–5.

[24] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Semi-supervised hyperspectral unmixing," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Quebec City, QC, Canada, Jul.13–18, 2014, pp. 3458–3461.

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

[26] S. Jia, J. Hu, J. Zhu, X. Jia, and Q. Li, "Three-dimensional local binary patterns for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2399–2413, Apr. 2017.

[27] F. Kizel and M. Shoshany, "Spatially adaptive hyperspectral unmixing through endmembers analytical localization based on sums of anisotropic 2D Gaussians," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 185–207, 2018.

[28] J. Song, H. Zhang, X. Li, L. Gao, M. Wang, and R. Hong, "Self-supervised video hashing with hierarchical binary auto-encoder," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3210–3221, Jul. 2018.

[29] Y. Wang *et al.*, "Self-supervised low-rank representation (SSLRR) for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5658–5672, Oct. 2018.

[30] X. Wang, K. He, and A. Gupta, "Transitive invariance for self-supervised visual representation learning," in *Proc. IEEE Int. Conf. Comput. Vision*, Venice, Italy, Oct.22–29, 2017, pp. 1338–1347.

[31] S. Lee and M. M. Crawford, "Unsupervised multistage image classification using hierarchical clustering with a Bayesian similarity measure," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 312–320, Mar. 2005.

[32] A. Plaza, D. Valencia, J. Plaza, and P. Martinez, "Commodity cluster-based parallel processing of hyperspectral imagery," *J. Parallel Distrib. Comput.*, vol. 66, no. 3, pp. 345–358, 2006.

[33] N. Dobigeon, S. Moussaoui, M. Coulon, J. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[34] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrainted least squares and active set method," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 713–730, 2008.

[35] C. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Comput.*, vol. 19, no. 10, pp. 2756–2779, 2007.

[36] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Comput. Statist. Data Anal.*, vol. 52, no. 1, pp. 155–173, 2007.

[37] N. Guan, D. Tao, Z. Luo, and B. Yuan, "NeNMF: An optimal gradient method for nonnegative matrix factorization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2882–2898, Jun. 2012.

[38] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, "Theorems on positive data: On the uniqueness of NMF," *Comput. Intell. Neurosci.*, vol. 2008, pp. 1–9, 2008.

[39] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[40] W. Wang and Y. Qian, "Adaptive $L_{1/2}$ sparsity-constrained NMF with half-thresholding algorithm for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2618–2631, Jun. 2015.

[41] Y. Lu, C. Yuan, W. Zhu, and X. Li, "Structurally incoherent low-rank nonnegative matrix factorization for image classification," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5248–5260, Nov. 2018.

[42] Z. Yang, Y. Xiang, K. Xie, and Y. Lai, "Adaptive method for nonsmooth nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 948–960, Apr. 2017.

[43] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[44] J. Choo, C. Lee, C. K. Reddy, and H. Park, "Weakly supervised nonnegative matrix factorization for user-driven clustering," *Data Mining Knowl. Discovery*, vol. 29, no. 6, pp. 1598–1621, 2015.

[45] L. Xu, J. Li, A. Wong, and J. Peng, "K-P-Means: A clustering algorithm of K 'purified' means for hyperspectral endmember estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1787–1791, Oct. 2014.

[46] W. Sun, J. Ma, G. Yang, B. Du, and L. Zhang, "A Poisson nonnegative matrix factorization method with parameter subspace clustering constraint for endmember extraction in hyperspectral imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 128, pp. 27–39, 2017.

[47] L. Xu, A. Wong, F. Li, and D. A. Clausi, "Extraction of endmembers from hyperspectral images using a weighted fuzzy purified-means clustering model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 695–707, Feb. 2016.

[48] X. Shen, W. Liu, I. W. Tsang, Q. Sun, and Y. Ong, "Multilabel prediction via cross-view search," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4324–4338, Sep. 2018.

[49] Y. Gao, R. Ji, P. Cui, Q. Dai, and G. Hua, "Hyperspectral image classification through bilayer graph-based learning," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2769–2778, Jul. 2014.

[50] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.

[51] X. Shen, F. Shen, Q. Sun, Y. Yang, Y. Yuan, and H. T. Shen, "Semi-paired discrete hashing: Learning latent hash codes for semi-paired cross-view retrieval," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4275–4288, Dec. 2017.

[52] X. Xu, X. Tong, L. Zhang, J. Hongzan, and H. Xie, "Unsupervised classification strategy utilizing an endmember extraction technique for airborne hyperspectral remotely sensed imagery," *J. Appl. Remote Sens.*, vol. 8, Oct. 2014, Art. no. 085090.

[53] 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 Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[54] X. Lu, H. Wu, and Y. Yuan, "Double constrained NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2746–2758, May 2014.
[55] 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.
[56] 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.
[57] X. Feng, H. Li, J. Li, Q. Du, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6245–6257, Oct. 2018.
[58] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.
[59] W. Wang, Y. Qian, and Y. Tang, "Hypergraph-regularized sparse NMF for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 681–694, Feb. 2016.
[60] G. A. Swayze, R. L. Clark, S. Sutley, and A. J. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, Pasadena, CA, USA, 1992, vol. 1, pp. 47–49.

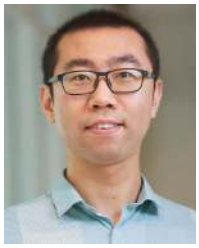
Wenhong Wang received the B.S. degree in computer software from Shandong University, Jinan, China, in 1995, the M.S. degree in computer application technology from Beijing University of Chemical Technology, Beijing, China, in 2003, and the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2016.

Since 1998, he has been with the College of Computer Science, Liaocheng University, Liaocheng, China, where he is currently an Associate Professor in computer science and technology. His research interests include hyperspectral image analysis, pattern recognition, and machine learning.

Yuntao Qian (Member, IEEE) received the B.E. and M.E. degrees in automatic control from Xi'an Jiaotong University, Xi'an, China, in 1989 and 1992, respectively, and the Ph.D. degree in signal processing from Xidian University, Xi'an, China, in 1996.

During 1996–1998, he was a Postdoctoral Fellow with the Northwestern Polytechnical University, Xi'an, China. Since 1998, he has been with the College of Computer Science, Zhejiang University, Hangzhou, China, where he became a Professor in 2002. During 1999–2001, 2006, 2010, 2013, and 2015–2016, he was a Visiting Professor with Concordia University, Hong Kong Baptist University, Carnegie Mellon University, the Canberra Research Laboratory of NICTA, Macau University, and Griffith University. His research interests include machine learning, signal and image processing, pattern recognition, and hyperspectral imaging.

Prof. Qian is an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.



Hongfu Liu (Member, IEEE) received the bachelor's and master's degrees in management information systems from the School of Economics and Management, Beihang University, Beijing, China, in 2011 and 2014, respectively, and the Ph.D. degree in computer engineering from Northeastern University, Boston, MA, USA, in 2018.

He is currently a Tenure-Track Assistant Professor with Michtom School of Computer Science, Brandeis University, Waltham, MA, USA. His research interests include data mining and machine learning, with special interests in ensemble learning.

Dr. Liu was the Reviewer for many IEEE transactions journals including the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE TRANSACTIONS ON BIG DATA. He was also on the program committee for the conferences including AAAI, International Joint Conference on Artificial Intelligence, and Conference and Workshop on Neural Information Processing Systems. He is an Associate Editor for the *IEEE Computational Intelligence Magazine*.