

Article

Hyperspectral Unmixing via Double Abundance Characteristics Constraints Based NMF

Rong Liu ¹, Bo Du ^{2,*} and Liangpei Zhang ¹

¹ The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China; lr@whu.edu.cn (R.L.); zlp62@whu.edu.cn (L.Z.)

² School of Computer, Wuhan University, Wuhan 430079, China

* Correspondence: gunspace@163.com; Tel.: +86-138-7146-1059

Academic Editors: András Jung, Lenio Soares Galvao and Prasad S. Thenkabail

Received: 11 March 2016; Accepted: 24 May 2016; Published: 31 May 2016

Abstract: Hyperspectral unmixing aims to obtain the hidden constituent materials and the corresponding fractional abundances from mixed pixels, and is an important technique for hyperspectral image (HSI) analysis. In this paper, two characteristics of the abundance variables, namely, the local spatial structural feature and the statistical distribution, are incorporated into nonnegative matrix factorization (NMF) to alleviate the non-convex problem of NMF and enhance the hyperspectral unmixing accuracy. An adaptive local neighborhood weight constraint is proposed for the abundance matrix by taking advantage of the spatial-spectral information of the HSI. The spectral information is utilized to calculate the similarities between pixels, which are taken as the measurement of the smoothness levels. Furthermore, because abrupt changes may appear in transition areas or outliers may exist in spatially neighboring regions, any inappropriate smoothness constraint on these pixels is removed, which can better express the local smoothness characteristic of the abundance variables. In addition, a separation constraint is used to prevent the result from over-smoothing, preserving the inner diversity of the same kind of material. Extensive experiments were carried out on both simulated and real HSIs, confirming the effectiveness of the proposed approach.

Keywords: hyperspectral unmixing; mixed pixels; abundance smoothness; selected local neighborhood; nonnegative matrix factorization (NMF)

1. Introduction

Past decades have witnessed the great success of hyperspectral imaging in a wide range of applications, due to its capacity to synchronously acquire both spatial and spectral information [1,2]. In hyperspectral images (HSIs), the spectral vector of each pixel contains hundreds or even thousands of elements, which provides rich spectral information to efficiently identify and distinguish different types of land cover [3]. However, due to the limited spatial resolution and the complexity of surface features [4], mixed pixels are common in HSIs. The existence of numerous mixed pixels conflicts with the demands for accurate recognition and interpretation of the material properties of the pixels. Hyperspectral unmixing (HU), which decomposes the mixed pixels into a set of constituent materials called “endmembers”, as well as the corresponding mixture coefficients called “abundances”, was developed to alleviate the mixed pixels problem [5]. HU makes it possible to reveal the material properties of pixels, so that the recognition and interpretation of pixels can be carried out at the sub-pixel scale, such as sub-pixel mapping [6] and sub-pixel target detection [7].

The linear mixing model (LMM) and the nonlinear mixing model (NLMM) are the two basic models used in HU, depending on the mixing degree of each type of material [4]. The LMM assumes that the mixed pixel spectrum is a linear combination of the pure material signatures weighted by the corresponding abundance fractions, and has been more widely applied than the NLMM in the

past decade, on account of its simplicity and suitability, as well as its clear physical meaning. It has therefore attracted a lot of attention in remote sensing fields [8–11]. Based on the LMM, three general approaches for unmixing have been developed: (1) geometric theory based approaches; (2) sparse regression based approaches; and (3) statistical theory based approaches [2].

Considering the relationship between geometric theory and the LMM, it can be found that a hyperspectral dataset actually lies in a simplex whose vertices correspond to the endmembers. Typical geometric theory based methods include the pixel purity index (PPI) [12], N-FINDR [13], the simplex growing algorithm (SGA) [14], vertex component analysis (VCA) [15], and so on. These methods all assume that there is at least one pure pixel per endmember. The practicability of the above methods is limited if the pure pixel assumption is violated, since they would not be able to find the exact endmembers in the image. The minimum volume based methods work whether the pure-pixel assumption is fulfilled or not. They aim at determining the simplex to enclose the observed data with the minimum volume. The endmembers are generated instead of selected in the image. The minimum volume enclosing simplex (MVES) [16], minimum volume simplex analysis (MVSA) [11], and the simplex identification via split augmented Lagrangian (SISAL) algorithm [17] are typical methods in this class. Sparse regression based approaches are relatively new developments, which conduct HU in a semi-supervised fashion by assuming that a spectral library of materials in the scene is available and that the observed data can be expressed in the form of linear combinations of a certain amount of signatures from the spectral library [18,19]. Sparse theory based methods have been developed to find the optimal combinations from the library since the number of materials in one pixel is far less than the number of signatures in the library [20–22], and the low-rank constraint based on the utilization of spatial correlation has been used to enhance the unmixing result [23,24]. The statistical theory based approaches can work under a highly mixed situation [8], and they generally obtain the endmembers and the corresponding abundances simultaneously. Independent component analysis (ICA) [25] and nonnegative matrix factorization (NMF) [26] are typical statistical approaches, both of which are blind source separation (BSS) approaches since they estimate the signals from the observations in the absence of prior knowledge. ICA is able to separate the original source signals under the assumption that the sources are statistically independent. However, this assumption is invalid under the LMM of HU since the sum of abundance fractions within each pixel is constant, which compromises the performance of the ICA algorithm in hyperspectral unmixing [27]. A number of methods [28–30] have also been proposed to improve the performance by means of adding auxiliary constraints to the ICA.

NMF approximately factorizes a nonnegative matrix into the product of two nonnegative matrices by adopting a multiplicative algorithm [26]. NMF has shown great potential for solving HU since it can obtain nonnegative results with physical significance [31]. Unfortunately, the algorithm may fall into many local minima due to the non-convexity of the objective function, and may not produce an accurate result [32]. Additional constraints on the NMF model according to the properties of the HSIs are needed. A variety of methods based on constrained NMF have been developed in either (or both) of two ways: imposing constraints on the spectral matrix or on the abundance matrix. In terms of the geometric features of HSIs, the minimum volume constrained NMF (MVCNMF) [33] method incorporates a minimum volume constraint into the NMF formulation to force the endmembers to enclose the data cloud. Considering the properties of the spectra, Wang *et al.* [34] proposed the endmember dissimilarity constrained NMF (EDCNMF) method by assuming the endmember spectra should be smooth and different from each other. Another method named abundance separation and smoothness constrained NMF (ASSNMF) [35] introduces two constraints, namely, the abundance separation constraint and abundance smoothness constraint, into the basic NMF. The ASSNMF method is based on two properties of HSIs: the correlation between different endmembers is weak, and ground objects usually vary slowly. The abundance matrix is generally supposed to be sparse since most pixels are mixed by only a few endmembers [9]. Accordingly, this feature has been widely exploited in HU methods. Among the various sparsity-constrained methods, $L_{1/2}$ sparsity-constrained NMF ($L_{1/2} - NMF$) [36] is a very popular approach. The objective of the L_q ($0 \leq q \leq 1$) regularizer is to

minimize the L_q ($0 \leq q \leq 1$) norm of the abundance matrix and the definition of the L_q ($0 \leq q \leq 1$) norm can refer to [37]. $L_{1/2}$ -NMF is superior to the L_1 regularization method since the L_1 regularizer cannot enforce further sparsity when the full additivity constraint of the material abundance is used [38]. Furthermore, the sparsity-constrained method has been generalized to L_q -NMF for $0 < q < 1$, and the sparsity imposed by the regularizers upon the unmixing task has been investigated [36,37], which demonstrated the superiority of the $L_{1/2}$ regularizer over the L_1 regularizer.

In recent years, researchers have made attempts to use the spatial information between different pixels as prior knowledge to enhance HU [39–42]. To further improve the performance of the sparse NMF algorithm, the graph-regularized $L_{1/2}$ -NMF (GLNMF) [42] method was proposed. This method utilizes the latent manifold structure of the data during the decomposition by incorporating an additional manifold regularization term into $L_{1/2}$ -NMF, which can keep the close link between the original image and the material abundance maps. Although the existing abundance characteristic based NMF methods have achieved good performances, we still believe that there is room for improvement.

Firstly, the smoothness levels may not be correctly described. Take Figure 1 as an example. Figure 1a shows that the ground objects between adjacent pixels in homogeneous areas vary slowly; that is to say, HSIs are spatially smooth. The abundances also possess the same smoothness feature since each abundance map characterizes the distribution of a certain kind of ground object. The method in [35] exploits the smoothness feature by assuming that two pixels are more similar if they are spatially closer, and assigns the same smoothness weight to the pixels which are the same spatial distance from the observation pixel. However, in some cases, as shown in the close-ups of Figure 1a, it is not always true that the smoothness levels of two-pixel pairs (a pixel pair is composed of the observation pixel and one of its spatially neighboring pixels in the surrounding local window) are the same if the spatial distances between them are the same. A more reliable measurement of smoothness levels should be used to reflect this difference.

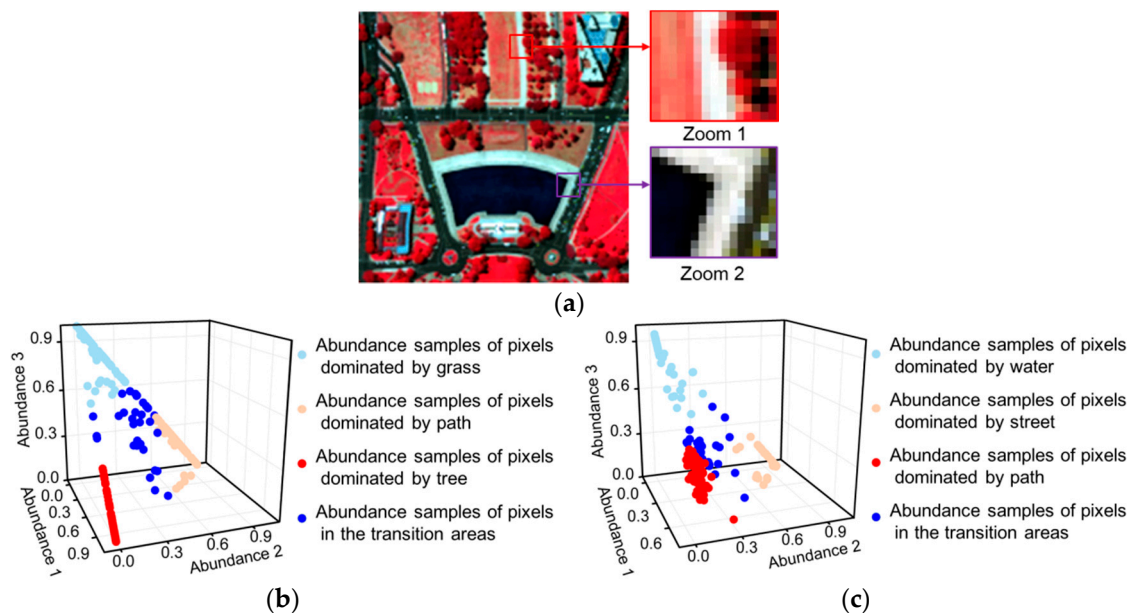


Figure 1. Three observations from the figure: the adjacent pixels vary slowly; pixels of a closer spatial distance are not always more similar; the abundance variables dominated by different ground objects have their own dominant regions and they are dispersed in a convex. (a) An HSI and close-ups of two patches in it; (b) 3-D scatter plot of the abundance samples of zoom 1 in (a); (c) 3-D scatter plot of the abundance samples of zoom 2 in (a).

Secondly, the spatial structure information in a local area may not be fully explored. The use of the smoothness constraint is based on the precondition that the pixels being constrained are similar.

However, this precondition is violated when abrupt changes appear in transition areas or outliers exist in the spatially neighboring regions, as shown in the close-ups of Figure 2a. Methods such as those proposed in [35,39] impose a smoothness constraint on all the spatially neighboring pixels in the local window. The spatial structure information is not fully explored by these methods since they lose sight of the pixels that are inappropriate for the smoothness constraint. In this condition, the smoothness constraint may be imposed on pixels which are actually dissimilar to the observation pixel, leading to extra error in the unmixing result.

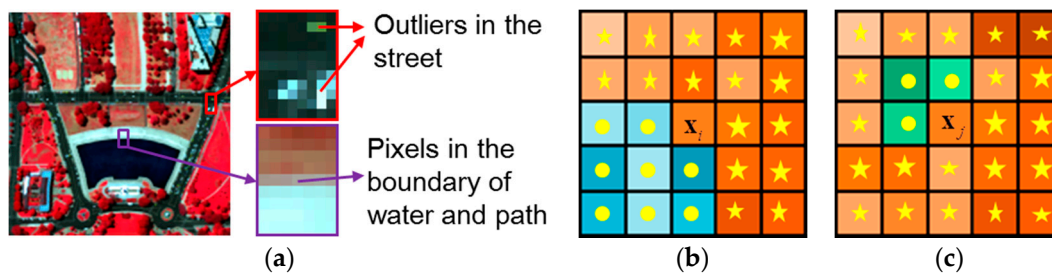


Figure 2. The concept of the selected local neighborhood weight method. (a) A hyperspectral image and close-ups of two patches; (b) An example of local neighborhood weights for x_i , which locates in the boundary of two classes; (c). An example of local neighborhood weights for x_j , which is spatially adjacent to outliers; In (b) and (c), the different colors represent different materials, and different shades of the same color represent the diversity of the same material; the dots denote a zero value of the assigned weight; the stars denote an assigned weight that is greater than zero, and a larger star represents a larger weight for more similar pixels.

Finally, the dispersed characteristic of the abundance variables is not fully taken into consideration. Actually, the dispersed characteristic of the abundance variables describes the statistical characteristics of the abundances, as reflected in Figure 1b,c. The 3-D scatter plots of the abundance samples show that the abundance variables dominated by different ground objects (an abundance variable dominated by a kind of ground object means that the pixel corresponding to this abundance variable is mainly composed of the ground object) have their own dominant regions and they are dispersed in a convex, which indicates that the abundance variables corresponding to different kinds of materials should be separate and only have a faint correlation. The method proposed in [39] only adds the smoothness constraint to NMF, which may lead to over-smooth results with the increase in iteration times. A constraint to exploit the dispersed characteristic can prevent these undesirable results by pulling the variables toward their own dominant region through minimizing the correlation between any two of the abundance variables.

Based on the above problems, we propose a novel double abundance characteristics constrained NMF (DAC²NMF) method, taking both the spatial structure information and the statistical distribution of the abundances into consideration. Our contributions can be summarized as:

- (1) The smoothness levels of each pixel pair are measured according to the similarities between them by taking advantage of the spectral information of the HSIs. In this way, more similar pixels are given a higher smoothness weight, as shown in Figure 2b,c, which is closer to the reality than a smoothness level determined by spatial distance.
- (2) Incorrect smoothness constraints are avoided by assigning a zero smoothness level to the pixels that are dissimilar to the observation pixel. The dissimilar pixels are excluded from the neighborhood pixels in the local window. The schematic diagrams in Figure 2b,c express this idea.
- (3) A separation constraint is used to prevent an over-smooth result by utilizing the dispersed characteristic of the abundance variables. A more stable and desirable result can be obtained in the interaction of these two constraints.

The remainder of this paper is organized as follows. Section 2 briefly presents the LMM and the basic NMF model. A detailed description of the proposed method is given in Section 3. Sections 4 and 5 evaluate the proposed algorithm using experiments on both synthetic and real hyperspectral data. Section 6 concludes the paper.

2. Related Works

2.1. The Linear Mixing Model (LMM)

As mentioned in Section 1, the LMM assumes that a pixel in an HSI can be expressed as a linear combination of a set of endmembers and the corresponding abundance fractions. Assuming that the image scene is dominated by P kinds of distinct materials with L bands, mathematically, an observation $\mathbf{x} \in \mathbb{R}^{L \times 1}$ can be written as:

$$\mathbf{x} = \sum_{i=1}^P s_i \mathbf{a}_i + \mathbf{n} = \mathbf{A}\mathbf{s} + \mathbf{e} \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_L]^T$ is the $L \times 1$ obtained pixel vector, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_P]$ is an $L \times P$ matrix, with each column being an endmember signature vector. $\mathbf{s} = (s_1, s_2, \dots, s_P)^T$ is a P -dimensional column vector composed of abundance coefficients of each endmember at the observation pixel, and \mathbf{e} represents the $L \times 1$ additive observation noise and error vector. Assuming that there are altogether N observations in the image, the LMM for all the pixels can be expressed by the matrix notation:

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \quad (2)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N]$, and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N]$. Clearly, the l_{th} column of matrix \mathbf{S} is the abundance coefficients of the l_{th} column of matrix \mathbf{X} . To be physically meaningful, the LMM is subject to two constraints on the entries of \mathbf{S} , namely, the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC), which can be explicitly given by $s_i \geq 0$, $i = 1, 2, \dots, P$ and $\mathbf{1}^T \mathbf{s} = 1$.

2.2. Nonnegative Matrix Factorization (NMF)

NMF approximates a high-dimensional nonnegative matrix with the product of two low-dimensional nonnegative matrices. The non-negativity constraints of NMF lead to a part-based representation because they allow only additive combinations [43]. This part-based property makes the NMF method well suited to many applications, such as face recognition [44] and document clustering [45]. Using the preceding notations, given a nonnegative matrix $\mathbf{X} \in \mathbb{R}^{L \times N}$, NMF aims to find the nonnegative matrix factors $\mathbf{A} \in \mathbb{R}^{L \times P}$ and $\mathbf{S} \in \mathbb{R}^{P \times N}$, such that:

$$\mathbf{X} \approx \mathbf{A}\mathbf{S} \quad (3)$$

Comparing the basic NMF model with the LMM model under a noise-free scenario, it can be found that they can both be seen as seeking linear combinations of a set of basis vectors, and their combination coefficients are both nonnegative. These similarities between the two models make NMF a suitable algorithm for HU. Lee and Seung [26] developed two simple multiplicative algorithms to solve the factorization problem of Equation (3), for which the square of the Euclidean distance between \mathbf{X} and $\mathbf{A}\mathbf{S}$ is a commonly used cost function. The objective function can be written as:

$$\begin{aligned} \text{minimize } f(\mathbf{A}, \mathbf{S}) &= \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 = \sum_{i,j=1}^{L,N} (\mathbf{X}_{ij} - (\mathbf{A}\mathbf{S})_{ij})^2 \\ \text{s.t. } &\mathbf{A} \geq 0, \mathbf{S} \geq 0 \end{aligned} \quad (4)$$

where the operator $\|\cdot\|_F$ represents the Frobenius norm. Although the minimization problem Equation (4) is separately convex in \mathbf{A} and \mathbf{S} , it is not simultaneously convex in both matrices. The widely used multiplicative algorithm presented in [26] is simple to implement and performs well, and can be generated from the traditional gradient descent algorithm. The gradient of Equation (4) can be written as:

$$\frac{\partial f(\mathbf{A}, \mathbf{S})}{\partial \mathbf{A}} = \mathbf{A}\mathbf{S}\mathbf{S}^T - \mathbf{X}\mathbf{S}^T \quad (5)$$

$$\frac{\partial f(\mathbf{A}, \mathbf{S})}{\partial \mathbf{S}} = \mathbf{A}^T\mathbf{A}\mathbf{S} - \mathbf{A}^T\mathbf{X} \quad (6)$$

where $(\cdot)^T$ denotes the transpose of the matrix. The update rules can be given by:

$$\mathbf{A} \leftarrow \mathbf{A} - u_A * (\mathbf{A}\mathbf{S} - \mathbf{X})\mathbf{S}^T \quad (7)$$

$$\mathbf{S} \leftarrow \mathbf{S} - u_S * \mathbf{A}^T(\mathbf{A}\mathbf{S} - \mathbf{X}) \quad (8)$$

where u_A and u_S are the step sizes. They are set as $u_A = \mathbf{A}/(\mathbf{A}\mathbf{S}\mathbf{S})$ and $u_S = \mathbf{S}/(\mathbf{A}\mathbf{A}\mathbf{S})$ to meet the nonnegative constraints. Substituting them into Equations (7) and (8), the update rules can be obtained:

$$\mathbf{A} \leftarrow \mathbf{A} * \mathbf{X}\mathbf{S}^T ./ \mathbf{A}\mathbf{S}\mathbf{S}^T \quad (9)$$

$$\mathbf{S} \leftarrow \mathbf{S} * \mathbf{A}^T\mathbf{X} ./ \mathbf{A}^T\mathbf{A}\mathbf{S} \quad (10)$$

where $*$ and $./$ represent the element-wise multiplication and division, respectively. The initialization of \mathbf{A} and \mathbf{S} should be nonnegative to ensure their non-negativity during the iteration under rules presented by Equations (5) and (6). The cost function Equation (4) is non-increasing after each iteration under the update rules, and it will be convergent to a stationary point.

3. The Double Abundance Characteristics Constrained NMF Method

Although NMF is quite appealing for HU, there are some challenges to be faced. One of the challenges is the lack of a unique solution to Equation (4) due to the aforementioned non-convexity in both \mathbf{A} and \mathbf{S} [34]. If one solution of \mathbf{A} and \mathbf{S} is obtained, then for any nonnegative invertible matrix \mathbf{D} whose inverse \mathbf{D}^{-1} is also nonnegative, $\mathbf{A}\mathbf{D}$ and $\mathbf{D}^{-1}\mathbf{S}$ are also a pair of solutions. In order to narrow the solution space and draw the decomposition toward the correct result, constraints or penalty terms are generally used to provide desirable results by considering the different properties of the HSIs. In this paper, we propose a constrained NMF method by considering two characteristics of the abundance variables, which is described in the following parts.

3.1. Smoothness Feature of the Abundances

The low spatial resolution of HSIs means that they lack tiny details; that is to say, the ground objects vary smoothly and abrupt changes rarely occur. It can be seen from the close-ups in Figure 1a that in the homogeneous areas, the spatially neighboring pixels are similar to each other and change very little, and sudden changes are found only in transition areas or where anomalies exist. This is an important spatial structure property of HSIs, which can be introduced to guide the unmixing. Remove the noise and error item in Equation (1), and each column of \mathbf{X} can be written as:

$$\mathbf{x} \approx \sum_{i=1}^P s_i \mathbf{a}_i \approx \mathbf{A}\mathbf{s} \quad (11)$$

It is easy to see from Equation (7) that each pixel in the image can be described as a combination of endmember set \mathbf{A} and the corresponding abundance coefficients. Taking \mathbf{A} as a set of basis vectors in a space, then \mathbf{S} can be regarded as the presentation of the L -dimension vector \mathbf{x} in the P -dimension space. Therefore, a kind of projection link can be established between the original hyperspectral dataset

and the abundance vectors. More specifically, similar pixels in the original image are expected to have similar abundance fractions after unmixing under this linear projection mode. In this sense, the smoothness feature of pixels is also appropriate for describing the relationship of the corresponding abundances. Based on the above analysis, the smoothness characteristic of the abundances between neighboring pixels in a local window is introduced as a constraint to the objective function of the basic NMF model.

The following two factors are taken into consideration when we design the smoothness constraint: (1) different ground objects may have diverse smoothness levels; and (2) the smoothness feature is violated in some places, such as transition areas. In the proposed method, the similarities between pixel pairs (the central pixel and each of its surrounding pixels in the local window) are utilized as prior knowledge to measure the smoothness levels of their abundances. The diverse smoothness levels are expressed by the assigned smoothness weights according to the different similarities between different pixel pairs. As mentioned before, some spatially neighboring pixel pairs may show sudden changes. In fact, these pairs should be abandoned when imposing the smoothness constraint on the abundances, because this kind of constraint on these pairs is not consistent with reality and will produce extra error instead of generating better results. As a result, we must identify these pixels and exclude them from the smoothness constraint to better describe the local spatial structure information of the image. The high spectral resolution of HSIs provides the observations with hundreds of spectral bands, which makes it possible for each pixel to characterize a certain kind of ground object with an almost continuous spectral curve. This valuable spectral information contained in each pixel vector can be used to account for pixel variability, similarity, and discrimination [46]. Inspired by the spectral angle mapper [47] the similarity between \mathbf{x}_i and \mathbf{x}_j is calculated by:

$$\beta_{ij} = \frac{\mathbf{x}_i^T \mathbf{x}_j}{\|\mathbf{x}_i\| \cdot \|\mathbf{x}_j\|} \quad (12)$$

where a larger β_{ij} stands for a higher similarity degree between \mathbf{x}_i and \mathbf{x}_j . The size of the local window is empirically chosen as 5×5 . After the calculation of the similarity values of all the pixel pairs (the central pixel and each of its surrounding pixels in the local window), an ascending order operator is conducted on them, and only the corresponding pixel pairs of the first 45% of the values participate in the abundance smoothness constraint, based on experimental investigation.

Based on the above process, the neighboring candidate pixels of each central pixel in the local window are determined. Given a pixel vector \mathbf{x}_i , if pixel vector \mathbf{x}_j is in the neighboring candidate pixel set of \mathbf{x}_i , then we define $j \in \mathcal{N}(i)$. It is clear that different pixel pairs may have different smoothness levels, and one may naturally hope that there should be a strategy to describe this kind of spatial structure feature. Here, the diverse smoothness levels are reflected with local neighborhood weights, which are defined as:

$$W_{ij} = \begin{cases} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}}, & \text{if } j \in \mathcal{N}(i) \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where Equation (13) is known as the heat kernel [42], and σ is a scaling parameter of the heat kernel weighting. The selection of σ is usually done manually. Considering that the same value of σ may fail to capture the data structure when the data contain multiple scales [48], the adaptive value of σ can be defined as:

$$\sigma = \frac{1}{k-1} \sum_{j \in \mathcal{N}(i)} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad (14)$$

where k is the number of selected neighboring pixels of pixel \mathbf{x}_i .

It is apparent that the more similar \mathbf{x}_i and \mathbf{x}_j are, the bigger W_{ij} is. Through the previous analysis, it is known that the abundance of a given pixel \mathbf{x}_i is similar to the abundance of the pixel in its neighbor $\mathcal{N}(i)$. To achieve the abundance smoothness constraint with consideration of the selection of

neighboring pixels and the calculation of different similarity levels, the selected local neighborhood weight regularization term for the abundance smoothness is defined as:

$$\begin{aligned} \text{minimize } J_1(\mathbf{S}) &= \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} W_{ij} \|\mathbf{s}_i - \mathbf{s}_j\|^2 \\ \text{s.t. } \mathbf{S} &\geq 0 \end{aligned} \quad (15)$$

3.2. Dispersed Characteristic of the Abundance Variables

We can roughly distinguish different classes of objects from an image scene such as Figure 1 by visual judgment because, in the real world, every kind of ground object has its own dominant region, which leads to the dispersed distribution of the abundance variables in HSIs. This distribution feature reveals that the abundance variables corresponding to each endmember seem to be independent of each other. However, the independent assumption is violated by the ASC in the LMM, which indicates that there is some correlation between the abundances of different objects due to mixed pixels. “Weak correlation” is more suitable for the relationship between different abundance variables. Mutual information is appropriate to express the above statistical information of the abundance variables, which is a commonly used measurement to describe the independence degree of variables. The mutual information function is given by the K-L divergence of the probability density of the abundance distributions [49]. By minimizing the mutual information function to a proper level, the dispersed characteristic of the abundance variables can be properly expressed. However, the distributions of the signals need to be estimated because they are usually not known in advance. One way used in [30] expresses the probability density function by a Gaussian distribution, and has achieved favorable unmixing results. In fact, by minimizing the abundance information divergence (AID) [33] of the abundance variables, the minimization of the mutual information can also be achieved, and, what is more, it needs no prior knowledge about the abundance distributions. AID is derived from the K-L divergence. Given two probability distributions of two discrete random signals $\mathbf{P} = [P_1, \dots, P_n, \dots, P_N]^T$ and $\mathbf{Q} = [Q_1, \dots, Q_n, \dots, Q_N]^T$, the definition of K-L divergence of \mathbf{Q} from \mathbf{P} is:

$$D_{KL}(\mathbf{P}||\mathbf{Q}) = \sum_n P_n \log \frac{P_n}{Q_n} \quad (16)$$

where $\sum_n P_n = \sum_n Q_n = 1$. Based on the K-L divergence, the AID of two abundance distributions $\mathbf{s}_i = (S_{i1}, S_{i2}, \dots, S_{iN})^T$ and $\mathbf{s}_j = (S_{j1}, S_{j2}, \dots, S_{jN})^T$ is defined as:

$$\text{AID}(\mathbf{s}_i, \mathbf{s}_j) = D_{KL}(\mathbf{p}||\mathbf{q}) + D_{KL}(\mathbf{q}||\mathbf{p}) \quad (17)$$

where $\mathbf{p} = \mathbf{s}_i / \sum_{n=1}^N S_{in}$ and $\mathbf{q} = \mathbf{s}_j / \sum_{n=1}^N S_{jn}$ are the normalizations of \mathbf{s}_i and \mathbf{s}_j , respectively. Clearly, AID is symmetric and always nonnegative. Due to the fact that the value range of its gradient is $(-\infty, +\infty)$, which may cause divergence during iteration of the gradient descent algorithm, Liu *et al.* [35] improved the AID to be more suitable and stable for the iterative algorithm, and they named the improved version the “separation function”. Attracted by its suitability and effectiveness, we chose this function to describe the dispersed characteristic of the abundance variables and to minimize the mutual information of the abundance variables. This regularization term is defined as:

$$\text{maximize } J_2(\mathbf{S}) = \frac{1}{2P^2} \sum_{i=1}^P \sum_{j=1}^P \sum_{n=1}^N \left[Q_{in} f\left(\frac{Q_{in}}{Q_{jn}}\right) + Q_{jn} f\left(\frac{Q_{jn}}{Q_{in}}\right) \right] \quad (18)$$

where $Q_{jk} = S_{jk} / \sum_{n=1}^N S_{jn}$; that is to say, \mathbf{Q} is the normalized \mathbf{S} . $f(x)$ is a replacement for the logarithm function used in the K-L divergence and AID, which is defined as:

$$f(x) = 1 - 2^{1-x^2} \quad (19)$$

The objective function J_2 is always nonnegative, and a larger value represents a lower correlation between abundances, so maximization of J_2 is equal to minimization of the mutual information of the abundances.

3.3. Abundance Sum-to-One Constraint

To satisfy the physical meaning of the LMM, the abundance should be subject to the ANC and ASC constraints. The ANC constraint is naturally satisfied by NMF. To satisfy the ASC, we adopt the widely used approach in [50]. In the iteration, the original dataset \mathbf{X} and the endmember matrix \mathbf{A} are augmented as:

$$\mathbf{X}_C \leftarrow \begin{bmatrix} \mathbf{X} \\ \delta \mathbf{1}^T \end{bmatrix} \quad \mathbf{A}_C \leftarrow \begin{bmatrix} \mathbf{A} \\ \delta \mathbf{1}^T \end{bmatrix} \quad (20)$$

where $\mathbf{1}^T$ is a vector with all the elements being 1, and δ is a parameter to adjust the impact of the ASC. In the implementation, a larger δ will result in a better performance, but the convergence rate is decreased. A appropriate value of $\delta = 20$ is selected to balance the tradeoff between accuracy and efficiency in this paper.

3.4. Objective Function and Update Rules of the Proposed Method

By integrating the auxiliary constraints presented in part A and part B into the basic NMF model, the objective function of the constrained NMF method is formed as follows:

$$\text{minimize } f(\mathbf{A}, \mathbf{S}) = \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + u_1 J_1(\mathbf{S}) - u_2 J_2(\mathbf{S}) \quad (21)$$

where u_1 and u_2 are regularization parameters, which balance the tradeoff among the three terms. To minimize Equation (21), a gradient descent algorithm based on the optimization of the basic NMF is developed. According to Equations (7) and (8), the update rules for DAC²NMF can be formulated as follows:

$$\mathbf{A} \leftarrow \mathbf{A} - u_A * (\mathbf{ASS}^T - \mathbf{XS}^T) \quad (22)$$

$$\mathbf{S} \leftarrow \mathbf{S} - u_S * \left(\mathbf{A}^T \mathbf{AS} - \mathbf{A}^T \mathbf{X} + u_1 \frac{\partial J_1(\mathbf{S})}{\partial \mathbf{S}} - u_2 \frac{\partial J_2(\mathbf{S})}{\partial \mathbf{S}} \right) \quad (23)$$

where the step sizes u_A and u_S are defined as $u_A = \mathbf{A} ./ (\mathbf{ASS})$ and $u_S = \mathbf{S} ./ (\mathbf{AAS})$, respectively. Substituting u_A and u_S into Equations (22) and (23), the multiplicative update rules can be written as:

$$\mathbf{A} \leftarrow \mathbf{A} * \mathbf{XS}^T ./ \mathbf{ASS}^T \quad (24)$$

$$\mathbf{S} \leftarrow \mathbf{S} * \left(\mathbf{A}^T \mathbf{X} - u_1 \frac{\partial J_1(\mathbf{S})}{\partial \mathbf{S}} + u_2 \frac{\partial J_2(\mathbf{S})}{\partial \mathbf{S}} \right) ./ (\mathbf{A}^T \mathbf{AS}) \quad (25)$$

The derivatives of $J_1(\mathbf{S})$ and $J_2(\mathbf{S})$ with respect to each element in \mathbf{S} are given as follows:

$$\frac{\partial J_1(\mathbf{S})}{\partial S_{pn}} = \sum_{j \in \mathcal{N}(n)} S_{pj} W_{nj} + \sum_{n \in \mathcal{N}(i)} S_{pn} W_{in} - \sum_{j \in \mathcal{N}(n)} S_{pj} W_{nj} - \sum_{n \in \mathcal{N}(i)} S_{pi} W_{in} \quad (26)$$

$$\frac{\partial J_2(\mathbf{S})}{\partial S_{pn}} = \frac{4 \ln 2}{P^2 \sum_{n=1}^N S_{pn}} \left\{ \sum_{j=1}^P \sum_{k=1}^N Q_{pk} \left[\frac{Q_{jk}^3}{Q_{pk}^3} 2^{-\frac{Q_{jk}^2}{Q_{pk}^2}} + \left(\frac{1}{2 \ln 2} - \frac{Q_{pk}^2}{Q_{jk}^2} \right) 2^{-\frac{Q_{pk}^3}{Q_{jk}^3}} - \frac{1}{4 \ln 2} \right] \right. \\ \left. - \sum_{j=1}^P \left[\frac{Q_{jn}^3}{Q_{pn}^3} 2^{-\frac{Q_{jn}^2}{Q_{pn}^2}} + \left(\frac{1}{2 \ln 2} - \frac{Q_{pn}^2}{Q_{jn}^2} \right) 2^{-\frac{Q_{pn}^3}{Q_{jn}^3}} - \frac{1}{4 \ln 2} \right] \right\} \quad (27)$$

3.5. Implementation Issues

3.5.1. Initialization

The number of endmembers is an essential parameter for most of the unmixing methods when no *a priori* knowledge is offered, and the accuracy of the number is crucial for the following unmixing result. The determination of the endmember number is a challenging task due to the interference of noise and anomalies. Among the numerous related studies, virtual dimensionality (VD) [51] and hyperspectral signal subspace identification by minimum error (HySime) [52] are two commonly used approaches for estimating the endmember number. Although these methods are effective, they cannot guarantee 100% accuracy. Since our work is focused on the unmixing stage, and the endmember number estimation is another independent topic, further discussion about the endmember number estimation is not included in this paper. Instead of relying on the endmember number estimation methods, we determine the endmember number by comprehensive visual interpretation, VD estimation, and by referring to the endmember numbers used in previous research.

Apart from the endmember number, the initializations of the endmember matrix \mathbf{A} and abundance matrix \mathbf{S} are also significant with regard to the NMF-based methods. It is impractical to obtain a global minima of the non-convex objective function through iterative optimization, since different initializations will result in various results for the same method. In this study, we used two different initialization methods for the synthetic experiments and the real data experiments, respectively. The simple approach adopted in [34], which utilizes the spectral information divergence (SID) [53], was employed in the synthetic experiments to initialize \mathbf{A} . Since the real image scene was more complex than the synthetic data, the VCA approach was used to initialize \mathbf{A} for the real data.

The initialization of abundance matrix \mathbf{S} is achieved by the maximum likelihood estimation from \mathbf{A} , which is defined as:

$$\mathbf{S} = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{X} \quad (28)$$

Considering that the initializations of \mathbf{A} and \mathbf{S} should both be nonnegative to ensure their non-negativity during the optimization, all of the elements in the initialized \mathbf{S} are checked and the negative values are forced to be 0 through the following operator:

$$\mathbf{S} = \max(\mathbf{S}, 0) \quad (29)$$

3.5.2. Stopping Condition

The stopping criterion, which can terminate the procedure when a stationary point is reached, is essential for NMF-based methods. The algorithm is considered to be converged if:

$$\frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{L} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2} \leq \tau \quad (30)$$

where τ is a specified error tolerance. In addition, the maximum number of iterations is also predefined. The procedure should be stopped when either of the stopping criteria is reached.

3.5.3. The Procedure of DAC²NMF

The flowchart of DAC²NMF is summarized as follows:

1. Determine the endmember number P ; initialize the endmember matrix \mathbf{A} by the SID-based algorithm for the synthetic experiments, and VCA algorithm for the real experiments; initialize the abundance matrix \mathbf{S} according to Equations (28) and (29);
2. update \mathbf{A} by Equation (24);
3. replace matrices \mathbf{A} and \mathbf{X} with matrices \mathbf{A}_C and \mathbf{X}_C according to Equation (20);

4. update \mathbf{S} by Equations (25)–(27);
5. replace matrices \mathbf{A}_C and \mathbf{X}_C with matrices \mathbf{A} and \mathbf{X} ;
6. repeat step 2–step 5 until reaching the maximum number of iterations or Equation (30) is satisfied;

3.5.4. Computational Complexity Analysis

Here, the computational complexity of the proposed DAC²NMF method is analyzed. There are two additional auxiliary constraints based on NMF. For the smoothness constraint, it needs $(24NL)$ to build the local neighborhood weights. The floating-point calculation times needed for calculating the gradients of the smoothness and separation constraints are $4PNk$ and $10P^2N$, respectively, where k is the number of selected neighboring pixels. In addition, the multiplicative update of the matrices needs $2[P^2(L+N) + PLN]$ times. If the procedure is stopped after m times of iterations, then the total cost is $O[m(P^2(2L+12N) + PN(2L+4k)) + 24NL]$.

4. Synthetic Image Experiments

We conducted a series of experiments to test the performance of the proposed DAC²NMF method with synthetic images. To verify the performance, the proposed method was compared with four related methods: ASSNMF, $L_{1/2}$ -NMF, GLNMF, and MVCNMF.

4.1. Performance Metrics

Two metrics were used to assess the quantitative accuracy of the estimated endmembers and their abundances: the spectral angle distance (SAD) [34] and the root-mean-square error (RMSE) [42]. The SAD was used to evaluate the shape similarity between the estimated endmember signature \mathbf{a} and the true endmember signature $\hat{\mathbf{a}}$, and is defined as:

$$\text{SAD}(\mathbf{a}, \hat{\mathbf{a}}) = \arccos\left(\frac{\mathbf{a}^T \hat{\mathbf{a}}}{\|\mathbf{a}\| \|\hat{\mathbf{a}}\|}\right) \quad (31)$$

Since the SAD value describes the spectral angle distance between two endmember signatures, a smaller value indicates a better estimation result. The similarity between the estimated abundance \mathbf{d} and the corresponding reference abundance $\hat{\mathbf{d}}$ was measured by the RMSE, which is defined as:

$$\text{RMSE}(\mathbf{d}, \hat{\mathbf{d}}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i - \hat{d}_i)^2} \quad (32)$$

where \mathbf{d} stands for a row vector of the estimated abundance matrix \mathbf{S} . Similar to SAD, a smaller value of RMSE represents a better estimation result for the abundance map.

4.2. Generation of Synthetic Images

The simulated synthetic images were of a size of 64×64 , with 182 bands, except for the image used to investigate the algorithm sensitivity to the image size. Ten spectral signatures were chosen from the U.S. Geological Survey (USGS) digital spectral library to create the synthetic data. Apart from the image used for testing the algorithm sensitivity to the number of endmembers in the image, seven spectra were used to generate the synthetic images. The seven spectral signatures are shown in Figure 3a. The generation of the abundance fractions was similar to the method used in [33], and can be described as follows: (1) an image of size 64×64 was divided into units of 8×8 blocks; (2) each block was randomly covered by one of the endmember classes; (3) a spatial low-pass filter of size 9×9 was utilized to generate mixed pixels; (4) according to the required mixing degree, some unsatisfactory abundance fractions were replaced by $1/P$. To be more explicit, if the desired mixing degree was 0.8, then we replaced all the pixels whose abundance was larger than 0.8 with a mixture composed of all

the endmembers, with abundances of $1/P$. One example of abundance maps with a mixing degree of 0.8 is shown in Figure 3b, as well as the first band of the synthetic image.

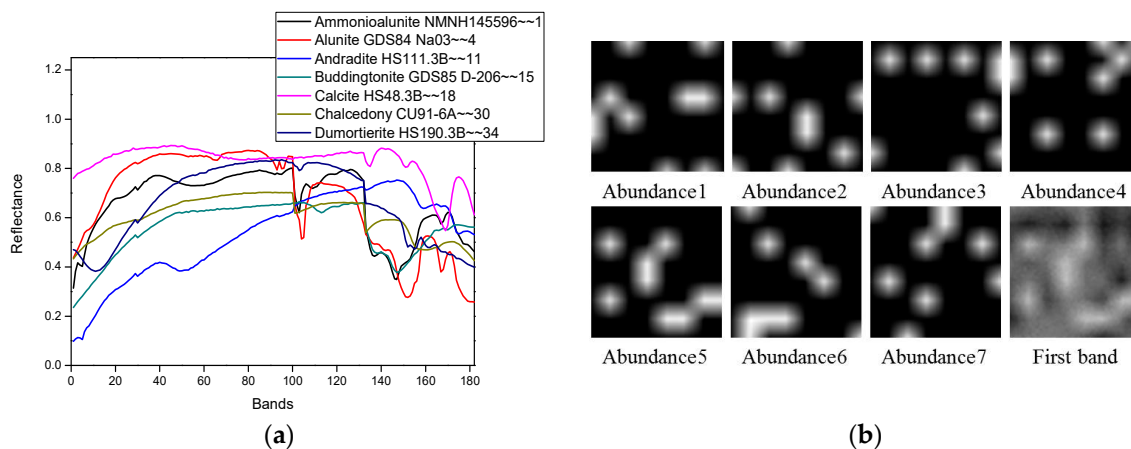


Figure 3. Example of the synthetic data. (a) Endmember spectra; (b) Abundance maps and the first band of the synthetic data.

By the above procedure, synthetic images without pure pixels were generated. In reality, images are generally corrupted by noise or possible errors. To evaluate the algorithm performance under the existence of noise, zero-mean white Gaussian noise was added into the synthetic data, and different signal-to-noise ratio (SNR) values were used to obtain the noisy synthetic data. The SNR is defined as [33]:

$$\text{SNR} = 10 \log_{10} \frac{E[\mathbf{x}\mathbf{x}^T]}{E[\mathbf{e}\mathbf{e}^T]} \quad (33)$$

where \mathbf{x} and \mathbf{e} represent a pixel vector and the noise on it, respectively. $E[\cdot]$ denotes the expectation operator.

4.3. Performance Evaluation

To evaluate the performance of the proposed method, five synthetic image experiments were designed. In the first experiment, the selections of the two regularization parameters and the convergence of the proposed method are analyzed. In the second experiment, the algorithms' robustness to noise corruption was studied by adding different levels of noise to the image. The third experiment was aimed at performing a sensitivity analysis of the proposed method to different mixing degrees. The fourth experiment was designed to investigate the algorithm's performance under different numbers of endmembers. In the final experiment, the algorithms were evaluated by different numbers of pixels in the HSI dataset in order to illustrate the impact of data quantity on the algorithm performance. To fairly evaluate the performance of the algorithms, for all the experiments, the algorithms used the same initial values and the same maximum number of iterations, which was set to 1000 in all the simulated experiments. The SAD values and RMSE values were obtained by an average of 10 individual implementations for each method. Furthermore, the number of endmembers was regarded as being known *a priori* for all the synthetic experiments.

4.3.1. Parameters Selection and Convergence Analysis

In this experiment, the selections of the two regularization parameters u_1 and u_2 are considered when $\text{SNR} = 30$ dB, and $\text{purity} = 0.8$. We change parameter $u_2 \in [1, 1 \times 10^3]$ with $u_1 \in [0, 0.4]$ and plot the performance of DAC^2NMF in terms of SAD and RMSE values, setting u_1 as the x coordinate as shown in Figure 4. We can see that, when we fix parameter u_2 , the algorithm achieves the best

performance with $u_1 = 0.1$ in terms of the SAD values, and with u_1 falling in the range $[0.1, 0.2]$ in terms of the RMSE values. When we fix parameter u_1 , a larger u_2 will lead to better results in terms of the SAD values; while a larger u_2 will lead to worse results in terms of RMSE when u_1 is set in the range $[0, 0.12]$. From the parameter analysis, the parameters of DAC²NMF were set as: $u_1 = 0.1$, $u_2 = 600$ for both synthetic and real experiments. Furthermore, the error tolerance τ is set to 0.01. The parameters for the other methods were chosen according to the experiments and analyses in the corresponding references for both synthetic and real experiments.

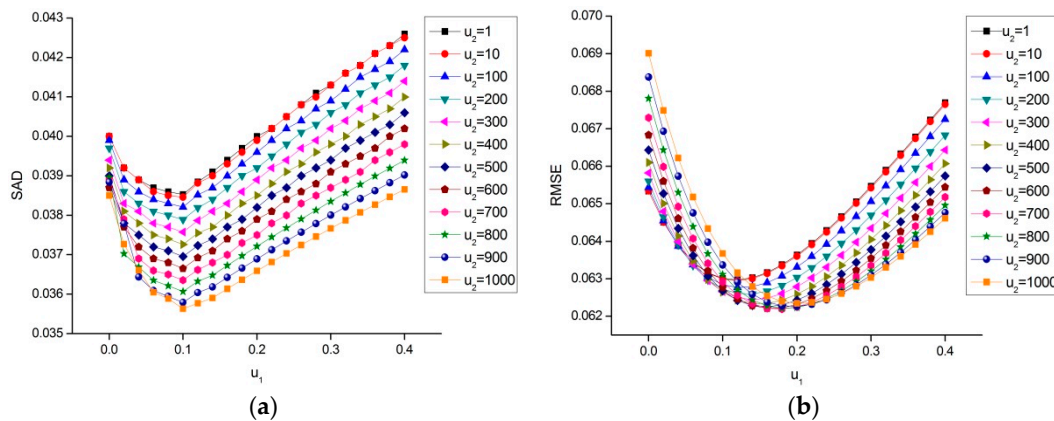


Figure 4. Parameter analysis of DAC²NMF in terms of (a) SAD and (b) RMSE.

Under the condition of the former parameter analysis, the approximation error of DAC²NMF with iterations is plotted in Figure 5. The approximation error is calculated according to Equation (4), and the red solid line represents the real approximation error. We can see that the approximation error curve of DAC²NMF can converge to the reality for about 800 iterations.

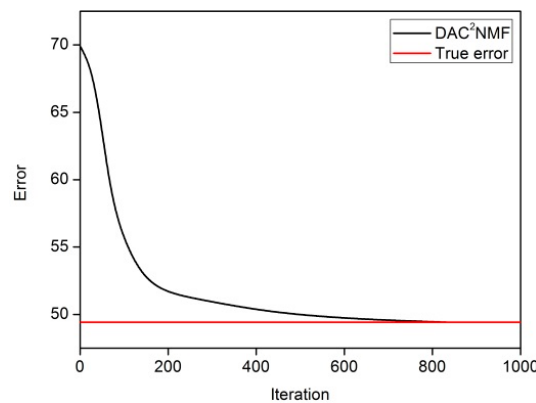


Figure 5. The approximation error of DAC²NMF with iterations.

4.3.2. Noise Robustness Analysis

In this experiment, the SNR was changed from 15 dB to 35 dB with a step size of 5 dB. The purity of the image was fixed at 0.8 and the image size was 64×64 . There were seven endmembers in the image. Figure 6 shows the results, where the bar and error bar represent the mean and standard deviation, respectively. It is clear that DAC²NMF leads to the best performance in terms of both SAD values and RMSE values. The superiority of DAC²NMF to the other algorithms is particularly evident when the SNR is equal to 15 dB. In terms of SAD, ASSNMF gives the second-best performance, followed by GLNMF. GLNMF outperforms $L_{1/2} - NMF$ with regard to SAD. Although MVCNMF gives the worst performance in terms of SAD, it performs better in the estimation of abundances.

Overall, it can be concluded that the proposed algorithm is robust with respect to different levels of noise.

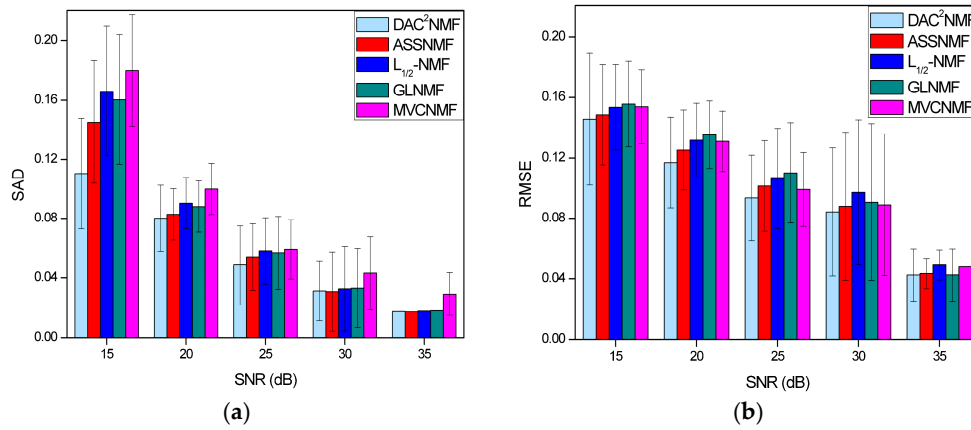


Figure 6. Experimental results with different SNR. (a) SAD; (b) RMSE.

4.3.3. Robustness Analysis to Degree of Mixing

The mixing degree of different real hyperspectral data may change with different spatial resolutions and the diverse complexity of ground objects. This experiment was aimed at evaluating the unmixing performance of the algorithm under various degree of mixing. The purity of the synthetic image was changed from 0.6 to 1 with a step size of 0.1. The image size, number of endmembers, and noise level were assigned as 64×64 , 7, and 20 dB, respectively. It is clear in Figure 7a that ASSNMF slightly outperforms $L_{1/2} - NMF$, GLNMF, and MVCNMF for the SAD values, and DAC²NMF still results in the smallest SAD values under different degrees of mixing. Particularly when the purity equals 0.9 and 1, the SAD values of DAC²NMF are clearly superior to the other methods. GLNMF obtains a smaller SAD value but a larger RMSE value than $L_{1/2} - NMF$ when the purity varies from 0.8 to 1. In terms of RMSE, DAC²NMF and ASSNMF are comparative when the purity equals 0.6 and 0.8, as shown in Figure 7b, and DAC²NMF outperforms ASSNMF under the other conditions. MVCNMF gives the smallest RMSE when the purity equals 0.6, followed by $L_{1/2} - NMF$ and GLNMF; DAC²NMF and ASSNMF can be seen to have larger RMSE values than the other methods under this condition, while they give the best performance with regard to SAD. Overall, the results reveal that DAC²NMF can contribute to a better estimation of endmember signatures, and the abundance estimation is better when the purity ≥ 0.7 .

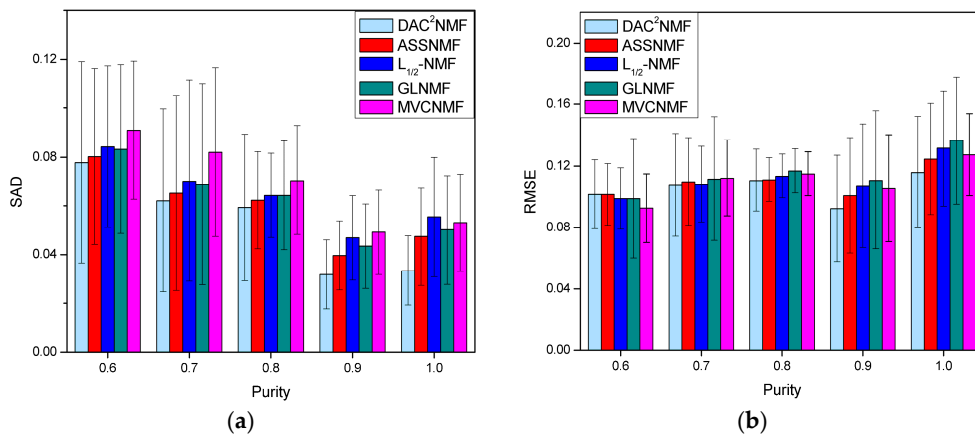


Figure 7. Experimental results with different degrees of mixing. (a) SAD; (b) RMSE.

4.3.4. Robustness Analysis to the Number of Endmembers

This experiment evaluated the performance of the algorithm when the synthetic image contained different number of endmembers, and the number of endmembers was changed from 3 to 10 with a step size of 1. The other conditions for the image were assigned as: SNR equal to 20 dB, purity equal to 0.8, and the image size was 64×64 . The average values and standard deviations of SAD and RMSE are shown in Figure 8a,b, respectively. It can be seen from the figure that ASSNMF results in the smallest SAD value but the largest RMSE value when there are three endmembers in the image, and DAC²NMF performs the best in terms of both SAD and RMSE, except for the next-best SAD value when $P = 3$. The performance of ASSNMF is slightly better than $L_{1/2} - NMF$, GLNMF, and MVCNMF when P varies from 5 to 10. MVCNMF obtains the largest SAD values but smaller RMSE values than $L_{1/2} - NMF$. Overall, the performances are weakened when the endmember number increases.

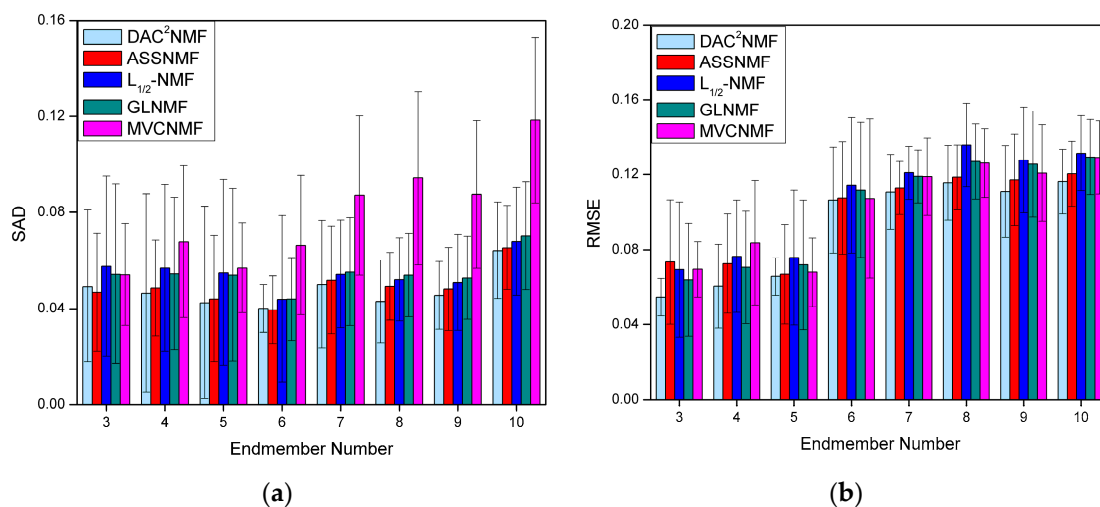


Figure 8. Experimental results with different numbers of endmembers. (a) SAD; (b) RMSE.

4.3.5. Robustness Analysis to the Image Size

In this experiment, the issue of algorithm sensitivity to different image sizes was studied. The synthetic images were designed with three different sizes: 64×64 , 96×96 , and 128×128 . The other conditions were the same: the images were corrupted by an SNR of 20 dB; there were seven endmembers in the images; and the purity was equal to 0.8. The results are shown in Figure 9. We can again see that the proposed DAC²NMF outperforms the other algorithms for the three image scenes, and ASSNMF is the next best. $L_{1/2} - NMF$ and GLNMF are comparative, and MVCNMF results in the largest SAD values, while it produces a better estimation of the abundances than $L_{1/2} - NMF$ and GLNMF. As the number of pixels increases, the performances of all the algorithms slightly decrease in terms of SAD. As for RMSE, the performances are comparative under the size of 96×96 and 128×128 . Overall, it is concluded that the proposed algorithm is suitable for different sizes of images.

Finally, we analyze the computational complexity of the algorithms. Table 1 shows the computational complexity of the algorithms calculated with the updated rules in one iteration, where N denotes the number of pixels in the image, L denotes the number of bands, P denotes the number of endmembers, k denotes the number of selected neighboring pixels of each pixel, and r and c denote the number of rows and columns of the image, respectively. From Table 1, we can see that the computational complexity of all the NMF-based algorithms increases rapidly with the increasing number of pixels, especially for $L_{1/2} - NMF$ and GLNMF. The computational complexity of MVCNMF increases faster than other algorithms with the increasing of endmember number.

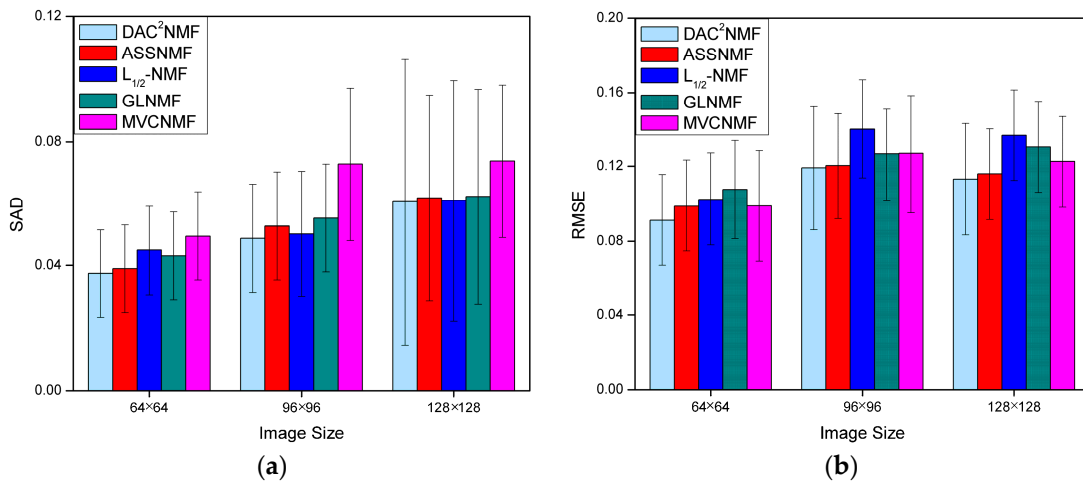


Figure 9. Experimental results with different image sizes. (a) SAD; (b) RMSE.

Table 1. The computational complexity of the algorithms.

Algorithms	Computational Complexity
DAC ² NMF	$O(p^2(2L + 12N) + PN(2L + 4k) + 24NL)$
ASSNMF	$O(p^2(2L + 12N) + PN(2L + 3r + 3c))$
$L_{1/2}$ -NMF	$O(LPN + (PN)^2)$
GLNMF	$O(NP(L + NP) + N^2L)$
MVCNMF	$O[2(p^2(L + N) + PLN) + p^2(L + P) + P!]$

5. Real Data Experiments

5.1. HYDICE Dataset

We also applied the DAC²NMF method to two real hyperspectral datasets. The first image was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor, covering an area in Washington DC. The image has 210 spectral channels, with a spectral resolution of 10 nm, covering the wavelength range of 0.4–2.4 μm . A subset of 150×150 was extracted from the original image for use in the experiment. The false-color image is shown in Figure 10.

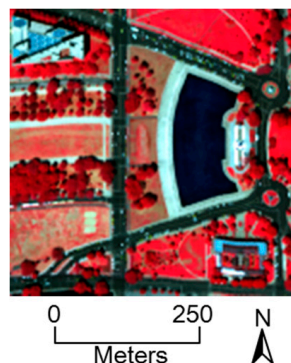


Figure 10. Sub-scene extracted from the Washington DC dataset.

The image in Figure 10 is composed of six main kinds of ground objects: roof, grass, water, tree, path, and street, so the number of endmembers was defined as 6. The parameters were the same as those in the synthetic experiments, and VCA was utilized for initialization of the endmember matrix.

The SAD values for DAC²NMF, ASSNMF, $L_{1/2}$ -NMF, GLNMF, and MVCNMF are provided in Table 2, and the best performance of each endmember is denoted in bold. It can be seen from Table 2 that DAC²NMF has the maximum number of best-performance cases, and ASSNMF, GLNMF, and MVCNMF each have one best-performance case, respectively. Figure 11 plots the mean SAD values for all the algorithms, and it shows that DAC²NMF performs the best, successively followed by GLNMF, $L_{1/2}$ -NMF, and ASSNMF. MVCNMF has the highest average SAD value. Figure 12 displays the abundance maps of the six types of materials. The abundance maps are grayscale maps, where a brighter pixel indicates a larger value of abundance fraction. We can clearly see the distribution of the six materials in the abundance maps. The abundance maps show that most of the materials cluster together or present a pattern of distribution in their own dominant area, which is consistent with the strategies designed in DAC²NMF. The smoothness constraint in DAC²NMF can facilitate the abundances of the same kind of material to be similar, and the separation constraint is used to prevent the abundance variables from over-smoothing and to ensure that each kind of material is distributed in its own dominant area. ASSNMF has no consideration for the pixels that are inappropriate for the smoothness constraint, which may bring error into the results. The performance of GLNMF is better than $L_{1/2}$ -NMF due to the strategy of integrating the latent manifold structure regularization term into $L_{1/2}$ -NMF.

Table 2. SAD Values of the Estimated Endmembers and Reference Endmembers for the Washington DC Dataset, USA.

	DAC ² NMF	ASSNMF	$L_{1/2}$ -NMF	GLNMF	MVCNMF
Roof	0.0458	0.0904	0.0846	0.0946	0.0613
Grass	0.2153	0.214	0.2246	0.2209	0.2803
Water	0.1082	0.145	0.1372	0.1378	0.1464
Tree	0.0223	0.0327	0.0325	0.0239	0.076
Path	0.1398	0.1437	0.1323	0.1302	0.1079
Street	0.0882	0.079	0.0518	0.0477	0.0576

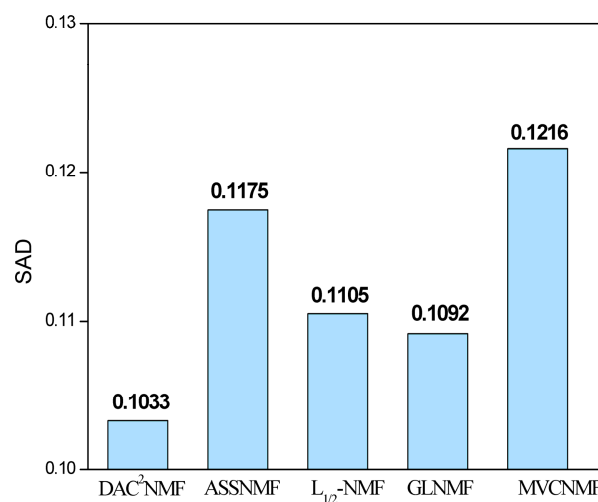


Figure 11. Average SAD values for the Washington DC dataset.

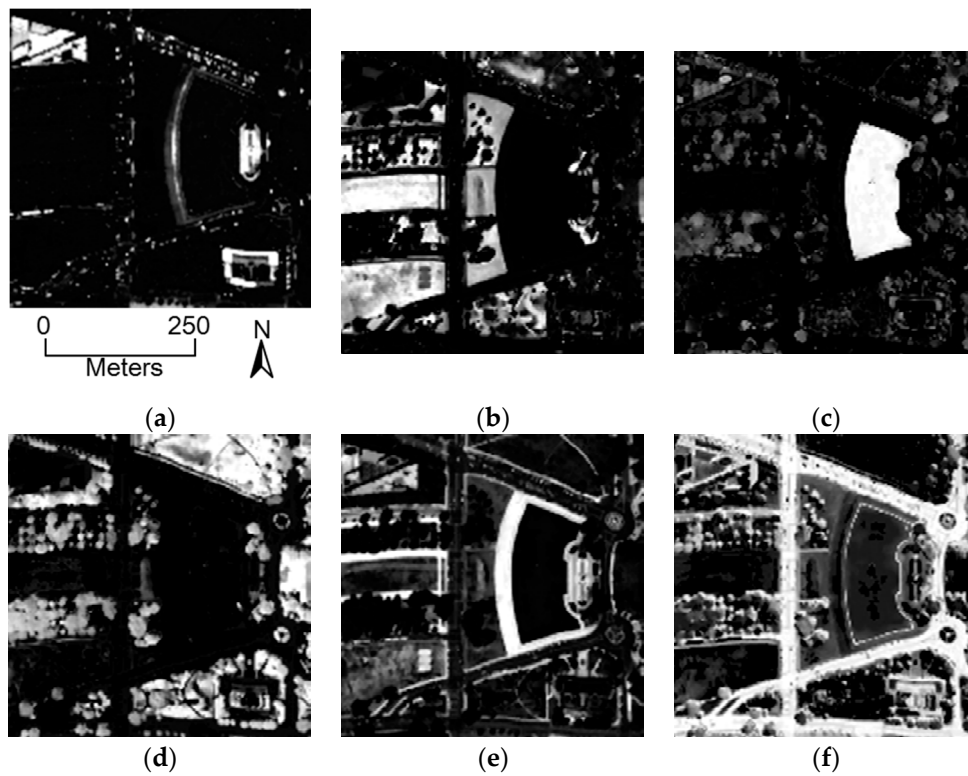


Figure 12. Abundance maps of the different endmembers using DAC²NMF with the Washington DC dataset. (a) Roof; (b) Grass; (c) Water; (d) Tree; (e) Path; (f) Street.

5.2. AVIRIS Dataset

The second image was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Cuprite, Nevada. There are 224 bands in the image, covering the wavelength range of 0.37–2.48 μm , with a spectral resolution of 10 nm. For our experiment, a block with the size of 250×190 was cut from the original data. The noisy bands (1–3 and 221–224) and water absorption bands (104–115 and 148–170) were removed, leaving 182 bands. The false-color image is shown in Figure 13.

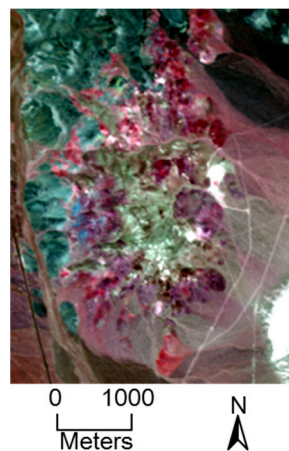


Figure 13. Sub-scene extracted from the Cuprite dataset.

The number of endmembers of this image is difficult to determine due to the fact that: (1) the existence of abundant rare minerals in this area means that the minerals are highly mixed, which leads to difficulty in interpreting the dataset; (2) the same kind of mineral may have different spectra with different chemical compositions. Rogge *et al.* [54] revealed that there are about 12 endmembers in this image, and this number was adopted in our experiment. The SAD values between the reference endmembers and the estimated endmembers by the algorithms are shown in Table 3, in which the bold font represents the best result. The corresponding abundance maps are displayed in Figure 14. In the results, nontronite, montmorillonite, and kaolinite are divided into two endmembers, respectively, which is a result of the signature variability. Figure 15 plots the average SAD values. The proposed DAC²NMF method outperforms the other methods in terms of both numbers of best-performance cases and average SAD values. The other methods all obtain two best-performance cases, while MVCNMF shows the worst performance in terms of average SAD value. It is clear that the distributions of the ground objects in this dataset are more complicated than for the Washington DC dataset. There are more materials in the dataset, and the distributions are more scattered, leading to an increase in the number of pixels located in transition areas. DAC²NMF again achieves the best results for this dataset due to the strategy of removing the dissimilar pixels from the smoothness constraint, and the separation constraint ensures the results are not over-smooth. ASSNMF imposes a smoothness constraint on all of the neighboring pixels, which is not appropriate for the pixels in transition areas. The performance of GLNMF is better than $L_{1/2} - NMF$ for this dataset due to the use of the manifold structure information.

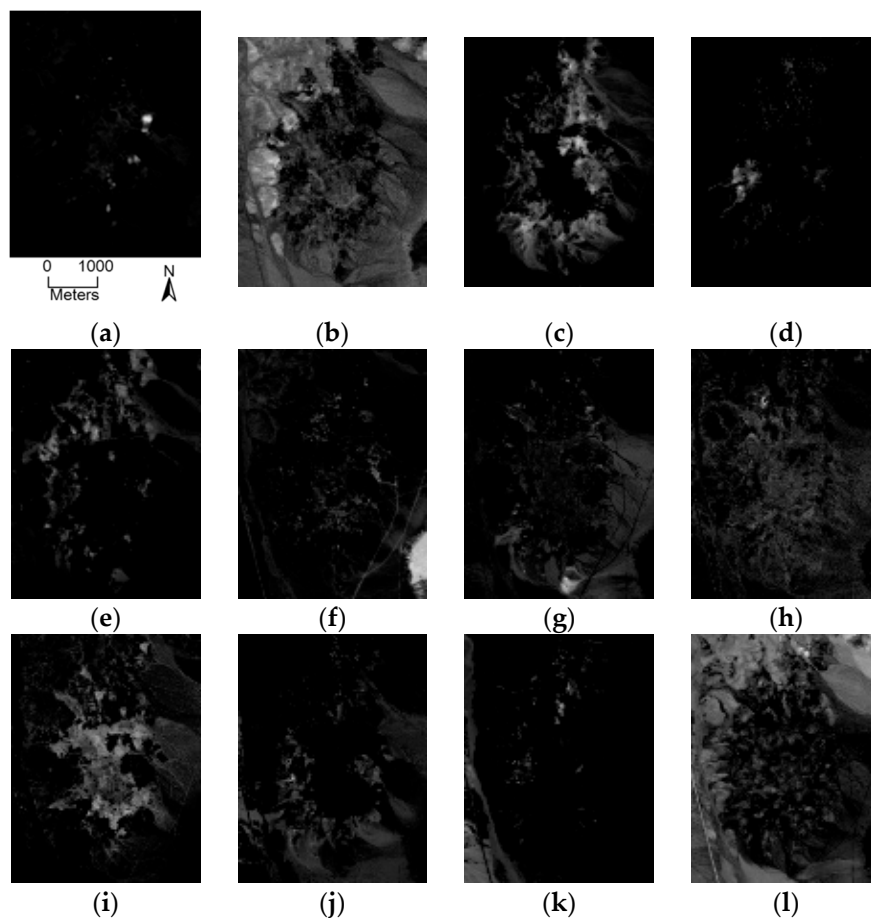


Figure 14. Abundance maps of the different endmembers using DAC²NMF with the Cuprite dataset. (a) Muscovite; (b) Sphene; (c) Alunite; (d) Buddingtonite; (e) Nontronite#1; (f) Montmorillonite#1; (g) Dumortierite; (h) Nontronite#2; (i) Chalcedony; (j) Kaolinite#1; (k) Kaolinite#2; (l) Montmorillonite#2.

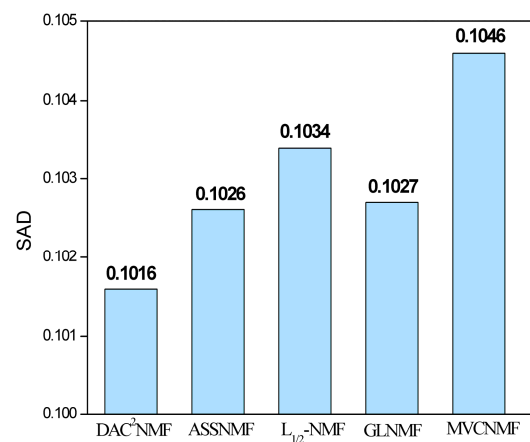


Figure 15. Average SAD values for the Cuprite dataset.

Table 3. SAD Values of the Estimated Endmembers and Reference Endmembers for the Cuprite dataset.

	DAC ² NMF	ASSNMF	L _{1/2} - NMF	GLNMF	MVCNMF
Muscovite	0.0663	0.0674	0.069	0.0684	0.0685
Sphene	0.0523	0.0508	0.0538	0.0527	0.0581
Alunite	0.0905	0.0909	0.0996	0.09	0.1067
Buddingtonite	0.1087	0.1032	0.1032	0.1016	0.1012
Nontronite#1	0.1018	0.1052	0.1072	0.104	0.1023
Montmorillonite#1	0.0794	0.082	0.0829	0.0831	0.0833
Dumortierite	0.0761	0.0781	0.0785	0.0755	0.0768
Nontronite#2	0.0715	0.0684	0.0692	0.0694	0.0713
Chalcedony	0.1231	0.1257	0.1257	0.1233	0.1212
Kaolinite#1	0.1828	0.1857	0.1823	0.1866	0.187
Kaolinite#2	0.2218	0.2257	0.2209	0.2273	0.2278
Montmorillonite#2	0.0454	0.0476	0.048	0.05	0.0508

6. Discussion

In this study, we investigated the validity of the abundance smoothness and dispersed characteristics for the NMF-based hyperspectral unmixing method. The endmember spectra and abundances estimated by the proposed DAC²NMF method fit well with the references. However, some issues still need to be resolved or improved for further research.

First, all of the NMF-based unmixing methods have regularization parameters to be set before the unmixing task. This is inevitable because different images may have different feature distribution characteristics, and the optimal parameter may differ. Although the NMF methods with constraints are more suitable and effective for unmixing than the original NMF method, the complexity is also increased. Therefore, how to make a trade-off between efficiency and performance is a problem that can be further considered.

Second, the number of endmembers is a necessary parameter for the unmixing method, but the NMF-based method is not able to adaptively determine the number, which cannot meet the real-time unmixing task. In the future, it is preferred to establish an unmixing system that also contains the estimation of the number of endmembers.

Third, the NMF-based methods did not take the intraclass spectral variability into consideration, which determines that they are not applicable to high-demand tasks that need to distinguish endmembers within class. The advantage of the multiple endmember analysis lies on that it uses various spectra of each endmember class to unmix each pixel. It is more accurate when there are plenty of spectral differences within each class. While the advantage of the NMF-based methods is that they can estimate the endmember spectral even if the hyperspectral is highly mixed and the endmember

spectral is not present in the image. Therefore, finding a way to integrate the two advantages may be a direction for further research.

7. Conclusions

In this paper, a new algorithm, double abundance characteristics constrained NMF (DAC²NMF), which utilizes both the abundance smoothness and the dispersed characteristic of the abundance variables, has been proposed. The local spatial smoothness structure of the abundances can be more accurately expressed by virtue of the adaptively selected local neighborhood weight regularization strategy. By removing the dissimilar pixel pairs from the abundance smoothness constraint, extra errors can be avoided, thus enhancing the unmixing result. Synthetic and real HSI experiments were analyzed to study the performance of the DAC²NMF method and other state-of-the-art algorithms. Experimental results show that in most cases, DAC²NMF method is superior to the other competing algorithms, which indicates that the adopted abundance characteristics in DAC²NMF are effective.

Acknowledgments: This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB719905, the National Natural Science Foundation of China under Grants 61471274, 41431175, U1536204, 60473023, and 61302111, the Natural Science Foundation of Hubei Province under Grants 2014CFB193 and the Fundamental Research Funds for the Central Universities.

Author Contributions: All authors contributed equally to this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tong, Q.; Xue, Y.; Zhang, L. Progress in Hyperspectral Remote Sensing Science and Technology in China Over the Past Three Decades. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 70–91. [[CrossRef](#)]
2. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379. [[CrossRef](#)]
3. Du, B.; Zhang, L.; Zhang, L.; Chen, T.; Wu, K. A discriminative manifold learning based dimension reduction method for hyperspectral classification. *Int. J. Fuzzy Syst.* **2012**, *14*, 272–277.
4. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57. [[CrossRef](#)]
5. Parra, L.C.; Spence, C.; Sajda, P.; Ziehe, A.; Müller, K.-R. Unmixing Hyperspectral Data. In Proceedings of the NIPS, Denver, CO, USA, 29 November–4 December 1999; pp. 942–948.
6. Roberts, D.A.; Gardner, M.; Church, R.; Ustin, S.; Scheer, G.; Green, R.O. Mapping Chaparral in the Santa Monica Mountains Using Multiple Endmember Spectral Mixture Models. *Remote Sens. Environ.* **1998**, *65*, 267–279. [[CrossRef](#)]
7. Zhang, L.; Du, B.; Zhong, Y. Hybrid detectors based on selective endmembers. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2633–2646. [[CrossRef](#)]
8. Bioucas-Dias, J.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [[CrossRef](#)]
9. Zhu, F.; Wang, Y.; Fan, B.; Xiang, S.; Meng, G.; Pan, C. Spectral Unmixing via Data-Guided Sparsity. *IEEE Trans. Image Process.* **2014**, *23*, 5412–5427. [[CrossRef](#)] [[PubMed](#)]
10. Cui, J.; Li, X.; Zhao, L. Linear Mixture Analysis for Hyperspectral Imagery in the Presence of Less Prevalent Materials. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4019–4031. [[CrossRef](#)]
11. Li, J.; Agathos, A.; Zaharie, D.; Bioucas-Dias, J.M.; Plaza, A.; Li, X. Minimum Volume Simplex Analysis: A Fast Algorithm for Linear Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5067–5082.
12. Boardman, J.W. Automating spectral unmixing of AVIRIS data using convex geometry concepts. In Proceedings of the Summaries 4th Annu. JPL Airborne Geoscience Workshop, Washington, DC, USA, 25–29 October 1993; Volume 1, pp. 11–14.
13. Winter, M.E. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. *Proc. SPIE* **1999**, 3753. [[CrossRef](#)]
14. Chang, C.-I.; Wu, C.-C.; Liu, W.-M.; Ouyang, Y.-C. A new growing method for simplex-based endmember extraction algorithm. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2804–2819. [[CrossRef](#)]

15. Nascimento, J.M.; Dias, J.M.B. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910. [[CrossRef](#)]
16. Chan, T.-H.; Chi, C.-Y.; Huang, Y.-M.; Ma, W.-K. A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Trans. Signal Process.* **2009**, *57*, 4418–4432. [[CrossRef](#)]
17. Bioucas-Dias, J.M. A variable splitting augmented Lagrangian approach to linear spectral unmixing. In Proceedings of the IEEE GRSS Workshop Hyperspectral Image Signal Processing: Evolution in Remote Sensing (WHISPERS), Grenoble, France, 26–28 August 2009; pp. 1–4.
18. Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Sparse Unmixing of Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2014–2039. [[CrossRef](#)]
19. Ma, W.; Bioucas-Dias, J.; Chan, T.; Gillis, N.; Gader, P.; Plaza, A.; Ambikapathi, A.; Chi, C.-Y. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Process. Mag.* **2014**, *31*, 67–81. [[CrossRef](#)]
20. Bioucas-Dias, J.M.; Figueiredo, M.A. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In Proceedings of the IEEE GRSS Workshop Hyperspectral Image Signal Processing: Evolution in Remote Sensing (WHISPERS), Reykjavik, Iceland, 14–16 June 2010; pp. 1–4.
21. Iordache, M.-D.; Bioucas-Dias, J.M.; Plaza, A. Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4484–4502. [[CrossRef](#)]
22. Iordache, M.-D.; Bioucas-Dias, J.M.; Plaza, A. Collaborative sparse regression for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 341–354. [[CrossRef](#)]
23. Qu, Q.; Nasrabadi, N.M.; Tran, T.D. Abundance Estimation for Bilinear Mixture Models via Joint Sparse and Low-Rank Representation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4404–4423.
24. Giampouras, P.V.; Themelis, K.E.; Rontogiannis, A.A.; Koutroumbas, K.D. Simultaneously Sparse and Low-Rank Abundance Matrix Estimation for Hyperspectral Image Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2016**, *PP*, 1–15. [[CrossRef](#)]
25. Hyvärinen, A.; Oja, E. Independent component analysis: Algorithms and applications. *Neural Netw.* **2000**, *13*, 411–430. [[CrossRef](#)]
26. Lee, D.D.; Seung, H.S. Algorithms for non-negative matrix factorization. In Proceedings of the NIPS 2000, Breckenridge, CO, USA, 2 December 2000; pp. 556–562.
27. Nascimento, J.M.; Dias, J.M. Does independent component analysis play a role in unmixing hyperspectral data? *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 175–187. [[CrossRef](#)]
28. Nascimento, J.M.; Bioucas-Dias, J.M. Hyperspectral unmixing based on mixtures of Dirichlet components. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 863–878. [[CrossRef](#)]
29. Xia, W.; Liu, X.; Wang, B.; Zhang, L. Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2165–2179. [[CrossRef](#)]
30. Wang, N.; Du, B.; Liangpei, Z.; Zhang, L. An Abundance Characteristic-Based Independent Component Analysis for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 416–428. [[CrossRef](#)]
31. Yang, Z.; Guoxu, Z.; Xie, S.; Ding, S.; Yang, J.-M.; Zhang, J. Blind Spectral Unmixing Based on Sparse Nonnegative Matrix Factorization. *IEEE Trans. Image Process.* **2011**, *20*, 1112–1125. [[CrossRef](#)] [[PubMed](#)]
32. Jia, S.; Qian, Y. Constrained nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 161–173. [[CrossRef](#)]
33. Miao, L.; Qi, H. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 765–777. [[CrossRef](#)]
34. Wang, N.; Du, B.; Zhang, L. An Endmember Dissimilarity Constrained Non-Negative Matrix Factorization Method for Hyperspectral Unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 554–569. [[CrossRef](#)]
35. Liu, X.; Xia, W.; Wang, B.; Zhang, L. An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 757–772. [[CrossRef](#)]
36. Qian, Y.; Jia, S.; Zhou, J.; Robles-Kelly, A. Hyperspectral Unmixing via $L_{1/2}$ Sparsity-Constrained Nonnegative Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4282–4297. [[CrossRef](#)]
37. Sigurdsson, J.; Ulfarsson, M.O.; Sveinsson, J.R. Hyperspectral Unmixing With l_q Regularization. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6793–6806. [[CrossRef](#)]
38. He, W.; Zhang, H.; Zhang, L. Sparsity-Regularized Robust Non-Negative Matrix Factorization for Hyperspectral Unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *PP*, 1–13. [[CrossRef](#)]

39. Liu, J.; Zhang, J.; Gao, Y.; Zhang, C.; Li, Z. Enhancing Spectral Unmixing by Local Neighborhood Weights. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1545–1552.
40. Martin, G.; Plaza, A. Spatial-Spectral Preprocessing Prior to Endmember Identification and Unmixing of Remotely Sensed Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 380–395. [[CrossRef](#)]
41. Yuan, Y.; Min, F.; Xiaoqiang, L. Substance Dependence Constrained Sparse NMF for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2975–2986. [[CrossRef](#)]
42. Lu, X.; Wu, H.; Yuan, Y.; Yan, P.; Li, X. Manifold regularized sparse NMF for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2815–2826. [[CrossRef](#)]
43. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. [[PubMed](#)]
44. Huang, S.; Elhoseiny, M.; Elgammal, A.; Yang, D. Improving non-negative matrix factorization via ranking its bases. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 5951–5955.
45. Li, L.; Yang, J.; Xu, Y.; Qin, Z.; Zhang, H. Documents clustering based on max-correntropy nonnegative matrix factorization. In Proceedings of the 2014 International Conference on Machine Learning and Cybernetics (ICMLC), Lanzhou, China, 13–16 July 2014; pp. 850–855.
46. Chang, C.-I. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Trans. Inf. Theory*. **2000**, *46*, 1927–1932. [[CrossRef](#)]
47. Dennison, P.E.; Halligan, K.Q.; Roberts, D.A. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sens. Environ.* **2004**, *93*, 359–367. [[CrossRef](#)]
48. Zelnik-manor, L.; Perona, P. Self-tuning spectral clustering. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 13–18 December 2004; pp. 1601–1608.
49. Chen, C.H.; Zhang, X. Independent component analysis for remote sensing study. *Proc. SPIE* **1999**, *3871*. [[CrossRef](#)]
50. Heinz, D.C.; Chang, C.-I. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 529–545. [[CrossRef](#)]
51. Chang, C.-I.; Du, Q. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 608–619. [[CrossRef](#)]
52. Bioucas-Dias, J.M.; Nascimento, J.M. Hyperspectral subspace identification. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2435–2445. [[CrossRef](#)]
53. Wang, S.; Wang, N.; Tao, D.; Zhang, L.; Du, B. A K-L divergence constrained sparse NMF for hyperspectral signal unmixing. In Proceedings of the International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Wuhan, China, 18–19 October 2014; pp. 223–228.
54. Rogge, D.; Rivard, B.; Zhang, J.; Sanchez, A.; Harris, J.; Feng, J. Integration of spatial–spectral information for the improved extraction of endmembers. *Remote Sens. Environ.* **2007**, *110*, 287–303. [[CrossRef](#)]

