

Autonomous Endmember Detection via an Abundance Anomaly Guided Saliency Prior for Hyperspectral Imagery

Xinyu Wang[✉], Member, IEEE, Yanfei Zhong[✉], Senior Member, IEEE, Chunyang Cui,
Liangpei Zhang[✉], Fellow, IEEE, and Yanyan Xu[✉]

Abstract—Determining the optimal number of endmember sources, which is also called “virtual dimensionality” (VD), is a priority for hyperspectral unmixing (HU). Although the VD estimation directly affects the HU results, it is usually solved independently of the HU process. In this article, a saliency-based autonomous endmember detection (SAED) algorithm is proposed to jointly estimate the VD in the process of endmember extraction (EE). In SAED, we first demonstrate that the abundance anomaly (AA) value is an important feature of undetected endmembers since pure pixels have larger AA values than “distractors” (i.e., mixed pixels and pure pixels of detected endmembers). Then, motivated by the fact that endmembers usually gather in certain local regions (superpixels) in the scene, due to spatial correlation, a superpixel prior is introduced in SAED to distinguish endmembers from noise. Specifically, the undetected endmembers are defined as visual stimuli in the AA subspace, the EE is formulated as a salient region detection problem, and the VD is automatically determined when there are no salient objects in the AA subspace. Since the spatial-contextual information of the endmembers is exploited during the saliency analysis, the proposed method is more robust than the spectral-only methods, which was verified using both real and synthetic hyperspectral images.

Index Terms—Abundance anomaly (AA), endmember extraction (EE), hyperspectral unmixing (HU), saliency analysis, virtual dimensionality (VD).

I. INTRODUCTION

MIXED pixels are common in remotely sensed hyperspectral images (HSIs), for reasons such as the

Manuscript received January 25, 2020; revised May 16, 2020; accepted May 17, 2020. Date of publication June 23, 2020; date of current version February 25, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0504202, in part by the National Natural Science Foundation of China under Grant 41771385 and Grant 41622107, and in part by the Postdoctoral Innovative Talents Program of China. (*Corresponding author: Yanfei Zhong*).

Xinyu Wang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: wangxinyu@whu.edu.cn).

Yanfei Zhong, Chunyang Cui, and Liangpei Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China, and also with the Hubei Provincial Engineering Research Center of Natural Resources Remote Sensing Monitoring, Wuhan University, Wuhan 430072, China (e-mail: zhongyanfei@whu.edu.cn; cuichunyang@whu.edu.cn; zlp62@whu.edu.cn).

Yanyan Xu is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: xuyy@whu.edu.cn).

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2020.3001353

insufficient spatial resolution of the imaging spectrometers, atmospheric effects, multiple scattering, and the microscopic mixing of materials [1], [2]. Over the last decades, various blind HU methods have been proposed to solve the mixed pixel problem [3], by decomposing mixed pixels into spectra of pure materials, termed “endmembers,” and their proportions, termed “abundances,” with little prior knowledge.

Generally speaking, there are three basic problems to be addressed in blind HU: 1) “virtual dimensionality” (VD) estimation to determine the optimal number of endmember sources M for HU [4]; 2) endmember extraction (EE) to extract the endmember spectral signatures [5]–[9], $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_M] \in \mathbb{R}^{L \times M}$; and 3) abundance inversion to estimate the abundance of each pixel [10], $\mathbf{s} \in \mathbb{R}^M$. The VD M is crucial for blind HU since it determines the rank of the endmember and abundance matrices. However, in a narrow sense, most of the blind HU algorithms mainly focus on solving the last two problems, i.e., estimating the endmember spectral signatures and abundances sequentially [5], [10], [11] or simultaneously [12]–[18], and the VD is usually assumed to be known in advance or is estimated by a third-party VD estimation algorithm.

Determining the optimal number of endmember sources M , which is also termed VD, is a prerequisite for blind HU. According to [19] and [20], the VD estimation algorithms can be roughly divided into two categories: 1) the information-theoretic criteria-based methods, e.g., the minimum description length (MDL) and the Akaike information criterion (AIC), where $M \in (0, 1, \dots, L - 1)$ is equal to the value minimizing the MDL or AIC criterion [21] and 2) the eigenanalysis-based methods, e.g., hyperspectral signal subspace identification by minimum error (HySime) [22], the Harsanyi–Farrand–Chang (HFC) method [23], and the noise-whitened HFC (NWHFC) method [4]. HFC/NWHFC is aimed at finding a significant eigenvalue gap between the eigenvalues of the sample covariance and correlation matrices, which indicates how many eigenvectors (i.e., VD) are necessary to represent a given HSI. HySime is aimed at identifying an optimal signal subspace by minimizing the mean-square error with the original data and its projection onto the signal subspace, where the VD is equal to the dimensionality of the optimal subspace. Generally speaking, the aforementioned VD estimation algorithms are general tools for signal subspace

identification, which offer a reference number, not only for HU but also for dimensionality reduction, anomaly detection, and band selection [22]. However, because the VD estimation has no interaction or effective feedback mechanism with the follow-up HU process, if the determined VD is not ideal for blind HU, the errors of the VD will directly influence the final HU accuracy and, more seriously, the subsequent estimated endmember spectra or abundances might be meaningless.

Autonomous EE, which is aimed at estimating the number and the spectra of endmembers jointly, is a possible solution to reduce the cumulative errors caused by the VD estimation. However, among the various geometry- and statistical-based EE algorithms, there is only a limited number of algorithms that can really achieve autonomous EE, such as the sparsity-promoting iterated constrained EE (SPICE) algorithm [14], the maximum orthogonal complement algorithm (MOCA) [24], and the geometry-based estimation of number of end-members (GENE) [20], where the optimal VD is determined during EE by either monotonously decreasing or increasing the number of endmembers. SPICE is an extension of the iterated constrained endmember (ICE) algorithm [12], which introduces a row-sparsity constraint on the abundance matrix. SPICE first overestimates the number of endmembers for unmixing and then iteratively prunes the unnecessary endmembers when their maximum abundances drop below a predefined threshold until convergence. Therefore, SPICE does not require prior knowledge of the number of end-members, but the results depend on the threshold and the weights of the different constraints (i.e., the linear, volume, and sparsity constraints). Compared with SPICE, GENE and MOCA extract endmembers one after another in sequence using orthogonal subspace projection (OSP)-based approaches, and the VD is determined during the iteration via a Neyman-Pearson detector, which is utilized to test whether a new endmember lies in the convex hull or the affine hull of the previously detected endmembers [20]. It should be noted that the abovementioned autonomous EE algorithms only exploit the geometric proprieties of endmembers, while the spatial-contextual information of endmembers is not fully utilized during autonomous EE.

In this article, to fully explore the spatial and spectral properties, a saliency-based autonomous endmember detection (SAED) algorithm is proposed to simultaneously estimate the number and the spectral signatures of endmembers. SAED is motivated by the fact that if the undetected endmember target still exists, it will manifest as the maximum AA value in the abundance anomaly (AA) subspace [25] and thus will be visually salient compared with the background signals (i.e., noise and mixed pixels). If not (i.e., all the endmembers have been detected), there will be no salient objects in the AA subspace. Therefore, the AA-guided saliency prior can be utilized to identify endmembers and determine the optimal number of endmembers for unmixing.

The saliency prior was first introduced into EE in saliency-based endmember detection (SED) [25], where endmembers and backgrounds are treated as visual stimuli and visual “distractors” in the AA subspace, respectively, and endmembers are identified one after another based on their local irregularity,

i.e., the center-surround contrast in multiple features and multiple scales, until the number of endmembers reaches a predefined VD value. Since the local spatial structures are exploited during the multiple-scale analysis, SED is robust to noise. However, there are two unsolved problems for SED. One is that the local irregularity prior assumes that endmember targets are locally rare and distinctive, which is ideal for small targets but may fail when endmembers have a large size and the background is complex. This is mainly because, for large targets, the boundary pixels usually have larger center-surround contrast compared with the central pixels. The other problem is that SED only utilizes the AA-guided saliency prior to EE, and it does not offer a means for determining the VD and thus cannot achieve autonomous EE.

The proposed SAED algorithm is an improved version of SED, where the superpixel prior (i.e., oversegmented and homogeneous superpixels with a comparable size) is introduced to exploit the spatial-contextual information of end-member targets for robust and autonomous EE. The above prior is based on the fact that endmembers usually exist in the form of superpixels [26] in spaceborne and airborne HSIs with a spatial resolution of between $1 \times 1 \text{ m}^2$ and $30 \times 30 \text{ m}^2$ [27]. Specifically, instead of the center-surround contrast operation that calculates the saliency at the pixel level, SAED utilizes superpixels as the basic units for measuring saliency, which can help to avoid boundary pixels being highlighted when the endmember targets are of large sizes, since the measured saliency is independent of its surroundings. In addition, the proposed SAED algorithm introduces the superpixel-based “objectness” [28] criteria to terminate the iteration and determine the VD for unmixing when there are no salient objects (endmembers) remaining in the AA subspace. In addition, since the number and locations of endmembers are determined according to the anomalies in the abundance maps, SAED naturally establishes a connection between the VD, the endmember spectra, and the abundances, which guarantees that the autonomously estimated endmember matrix can decompose the data well. The main contributions of SAED can be summarized as follows.

- 1) A saliency-based framework for autonomous endmember detection is proposed to jointly estimate the number and the spectral signatures of endmembers, where the spatial-contextual information is exploited via saliency analysis to promote the robustness of the autonomous EE.
- 2) Since the saliency score of endmembers is measured at the superpixel level, compared with SED, SAED is also applicable for the cases where the endmembers have different sizes or the background is complex.
- 3) Since the locations and the number of endmembers are determined in the AA subspace, SAED naturally builds up the interactions between the VD, the endmembers, and the abundances, which guarantees that the estimated endmember matrix can unmix the data well, with few reconstruction errors (REs).

The rest of this article is organized as follows. Section II introduces the linear mixture model (LMM) and compares

the traditional OSP approach and the AA projection used in SAED. Section III describes the proposed SAED algorithm in detail. The experimental analysis with synthetic and real hyperspectral data is provided in Sections IV and V, respectively. Finally, our conclusions are drawn in Section V.

II. SUBSPACE PROJECTION

This section first introduces the LMM and then compares two subspace projection approaches for EE: the OSP applied in many EE algorithms, such as vertex component analysis (VCA) [5], OSP [11], and GENE [20]; and the AA subspace projection used in SED [25] and the proposed SAED algorithm for saliency analysis.

A. Linear Mixture Model

The LMM is widely used for unmixing due to its explicit physical meaning, where a mixed pixel $\mathbf{x}_j \in \mathbb{R}^L$

is considered as a linear combination of the M endmembers in the scene, weighted by their abundances

$$\mathbf{x}_j = \sum_{i=1}^M s_j[i] \mathbf{a}_i + \boldsymbol{\varepsilon}_j = \mathbf{A}\mathbf{s}_j + \boldsymbol{\varepsilon}_j \quad (1)$$

where $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_M] \in \mathbb{R}^{L \times M}$ is a collection of endmembers, $\mathbf{s}_j = [s_j[1], \dots, s_j[M]]^T \in \mathbb{R}^M$ is the abundance vector, and $\boldsymbol{\varepsilon}_j \in \mathbb{R}^L$ is the additive noise term. Since the entries refer to the proportions of the materials making up a mixed pixel, \mathbf{s}_j

satisfies the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC)

$$s_j[i] \geq 0, \quad \sum_{i=1}^M s_j[i] = 1 \quad (2)$$

According to the LMM, there are three basic problems to be addressed for blind HU: 1) the determination of the VD (i.e., M); 2) the extraction of the endmembers (i.e., $\mathbf{A} \in \mathbb{R}^{L \times M}$); and 3) the inversion of the abundances for each pixel (i.e., \mathbf{s}_j).

B. Orthogonal Subspace Projection

Since it is difficult to directly extract endmembers in the original space, the HSI is usually projected into some specific subspace, in which the undetected endmembers can be separated easily from the background. OSP is a powerful tool for identifying pure pixels in HSIs, and the basic idea has been applied in a variety of EE approaches.

Let us assume that pure pixels are present in the scene and that the HSI is noise-free. Then, according to the triangle inequality principle, each pixel in the HSI has the following property:

$$\|\mathbf{x}_j\|_p = \left\| \sum_{i=1}^M s_j[i] \mathbf{a}_i \right\|_p \leq \sum_{i=1}^M s_j[i] \cdot \|\mathbf{a}_i\|_p \leq \max_{i=1, \dots, M} \|\mathbf{a}_i\|_p \quad (3)$$

where $\|\cdot\|_p$ denotes the p -norm ($p \geq 1$), and the equality holds only when \mathbf{x}_j is a pure pixel of the endmember with the

largest p -norm value. The first endmember $\hat{\mathbf{a}}_1$ can be identified as

$$\hat{\mathbf{a}}_1 = \mathbf{x}_{\ell_1}, \quad \ell_1 = \arg \max_{j=1, \dots, N} \|\mathbf{x}_j\|_p \quad (4)$$

where \mathbf{x}_{ℓ_1} is the pure pixel with the largest p -norm value.

Once $\hat{\mathbf{a}}_1$ is determined, the other endmembers $\{\hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_M\}$ can be identified successively using the OSP algorithm. Without loss of generality, we let $\mathbf{A}_k = [\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_k] \in \mathbb{R}^{L \times k}$ denote the endmember set already determined, where $k < M$, and $\mathbf{P}_{\mathbf{A}_k}^\perp = \mathbf{I} - \mathbf{A}_k (\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T \in \mathbb{R}^{L \times L}$ denotes the projector that is orthogonal to the subspace spanned by \mathbf{A}_k . For any $j = 1, \dots, N$, we have

$$\|\mathbf{P}_{\mathbf{A}_k}^\perp \mathbf{x}_j\|_p \leq \sum_{i=1}^M s_j[i] \cdot \|\mathbf{P}_{\mathbf{A}_k}^\perp \mathbf{a}_i\|_p \leq \max_{i=1, \dots, M} \|\mathbf{P}_{\mathbf{A}_k}^\perp \mathbf{a}_i\|_p. \quad (5)$$

Since $\mathbf{P}_{\mathbf{A}_k}^\perp \hat{\mathbf{a}}_i = \mathbf{0}$ holds for any endmember $\hat{\mathbf{a}}_i \in \{\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_k\}$ already extracted, the equality in (5) holds only in the case that \mathbf{x}_j is an undetected endmember $\hat{\mathbf{a}}_i \in \{\hat{\mathbf{a}}_{k+1}, \dots, \hat{\mathbf{a}}_M\}$ with the maximum OSP, and thus, the new endmember $\hat{\mathbf{a}}_{k+1}$ can be identified as

$$\hat{\mathbf{a}}_{k+1} = \mathbf{x}_{\ell_{k+1}}, \quad \ell_{k+1} = \arg \max_{j=1, \dots, N} \|\mathbf{P}_{\mathbf{A}_k}^\perp \mathbf{x}_j\|_p. \quad (6)$$

According to the abovementioned proof, the endmembers in the scene can be identified one after another using OSP until $k = M$.

Generally speaking, OSP is an all-to-one mapping, i.e., an $L \times m \times n$ hyperspectral data cube is mapped into an $m \times n$ OSP map, in which the maximum location corresponds to an undetected endmember under the noiseless argument.

C. AA Subspace

In SAED, to enhance the robustness, the AAs [25] are used to separate pure pixels of the undetected endmembers from the background. Compared with OSP, the AA projection is an all-to- K mapping (i.e., $K = k + 2$), where an $L \times m \times n$ data cube is mapped into a $K \times m \times n$ AA subspace, where the maximum point in each AA channel corresponds to an undetected endmember. Compared with the OSP mapping, the AA subspace provides more information about the undetected endmembers, which is valuable for the subsequent saliency analysis.

First, we consider the optimal case where both the endmembers $\mathbf{A}_k = [\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_k] \in \mathbb{R}^{L \times k}$ and each abundance vector $\mathbf{s}_j = [\hat{s}_j[1], \dots, \hat{s}_j[k]]^T \in \mathbb{R}^k$ are well estimated, where $k = M$ and $\hat{\mathbf{s}}_j$ is obtained using linear least squares (LS) estimation

$$\hat{\mathbf{s}}_j = \mathbf{A}_k^+ \mathbf{x}_j \quad (7)$$

where $\mathbf{A}_k^+ = (\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T \in \mathbb{R}^{k \times L}$ is the pseudoinverse of \mathbf{A}_k . In this situation, the abundances $\hat{\mathbf{s}}_j$ should satisfy the following properties:

$$(P1) \quad 0 \leq \hat{s}_j[i] \leq 1, \quad \forall i$$

$$(P2) \quad 1 - \sum_{i=1}^M \hat{s}_j[i] \approx 0$$

$$(P3) \quad \|\mathbf{x}_j - \mathbf{A}_k \hat{\mathbf{s}}_j\|_2 \approx 0$$

where (P1) and (P2) are based on the ANC and the ASC, and (P3) is based on the LMM. It is worth noting that (P1)–(P3) hold only in the case that \mathbf{A}_k is well estimated, and when \mathbf{A}_k is not correctly estimated (i.e., $k < M$ or $\hat{\mathbf{a}}_i \notin \{\mathbf{a}_1, \dots, \mathbf{a}_M\}$), the estimated abundances in (7) may contain outliers and make properties (P1)–(P3) fail.

We now discuss the opposite case of the endmember matrix being underestimated. Without loss of generality, we let $\mathbf{A}_k \in \mathbb{R}^{L \times k}$ and $\mathbf{B} = [\hat{\mathbf{b}}_{k+1}, \dots, \hat{\mathbf{b}}_M] \in \mathbb{R}^{L \times (M-k)}$ denote the endmembers already determined and those remaining undetermined, respectively, where $k < M$, $\mathbf{A}_k \cap \mathbf{B} = \emptyset$, and $\mathbf{A}_k \cup \mathbf{B} = \mathbf{A}$. The LMM in (1) can then be represented as

$$\mathbf{x}_j = \sum_{i=1}^k s_j^a[i] \hat{\mathbf{a}}_i + \sum_{n=k+1}^M s_j^b[n] \hat{\mathbf{b}}_n = \mathbf{A}_k s_j^a + \mathbf{B} s_j^b \quad (8)$$

where $\mathbf{s}_j^a \in \mathbb{R}^k$, $\mathbf{s}_j^b \in \mathbb{R}^{M-k}$, and $\sum_{i=1}^k s_j^a[i] + \sum_{n=k+1}^M s_j^b[n] = 1$. The LS estimation in (7) can then be rewritten as

$$\hat{\mathbf{s}}_j = \mathbf{A}_k^+ (\mathbf{A}_k \mathbf{s}_j^a + \mathbf{B} \mathbf{s}_j^b) = \mathbf{s}_j^a + \Delta \mathbf{s}_j^b \quad (9)$$

where $\Delta = \mathbf{A}_k^+ \mathbf{B} = [\Delta_{k+1}, \dots, \Delta_M] \in \mathbb{R}^{k \times (M-k)}$, and for any $n = k+1, \dots, M$, $\Delta_n = (\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T \hat{\mathbf{b}}_n \in \mathbb{R}^k$. As shown in (9), when the endmembers are underestimated, the abundance vector $\hat{\mathbf{s}}_j$ consists of two parts: the original abundance \mathbf{s}_j^a and the outliers $\Delta \mathbf{s}_j^b$ caused by undetected endmembers. Due to $\Delta \mathbf{s}_j^b$, (P1)–(P3) may fail, and the AAs then appear.

The AAs of each pixel \mathbf{x}_j for \mathbf{A}_k are defined as a $k+2$ dimensional vector, $\mathbf{v}_j = [v_j[1], \dots, v_j[k], vs_j, vr_j]^T$, where

$$v_j[i] = \max(-s_j[i], 0) + \max(s_j[i] - 1, 0) \quad (10)$$

$$vs_j = \left| 1 - \sum_{i=1}^k s_j[i] \right| \quad (11)$$

$$vr_j = \left\| \mathbf{x}_j - \hat{\mathbf{A}}_k \hat{\mathbf{s}}_j \right\|_2 \quad (12)$$

where the $\max(\cdot, 0)$ operation in (10) returns the absolute value of the anomaly parts. The first k entries defined in (10), $v_j[1], \dots, v_j[k]$, correspond to the AAs that fail to meet (P1), and the other entries defined in (11) and (12) correspond to the cases that fail to satisfy (P2) and (P3), respectively.

We now consider the relationships between the abovementioned AAs and the undetected endmembers. First, according to (10), we note that when $0 \leq \hat{s}_j[i] \leq 1$, $v_j[i] = 0$, and when $\hat{s}_j[i] < 0$, we have $\min_{n=k+1, \dots, M} \Delta_n[i] < 0$ and

$$v_j[i] = \left| s_j^a[i] + \sum_{n=k+1}^M s_j^b[n] \Delta_n[i] \right| \leq \left| \min_{n=k+1, \dots, M} \Delta_n[i] \right| \quad (13)$$

and when $\hat{s}_j[i] > 1$, we have $\max_{n=k+1, \dots, M} \Delta_n[i] > 1$ and

$$v_j[i] = s_j^a[i] + \sum_{n=k+1}^M s_j^b[n] \Delta_n[i] - 1 \leq \max_{n=k+1, \dots, M} \Delta_n[i] - 1. \quad (14)$$

Note that the equalities hold only in the case that $\mathbf{x}_j = \hat{\mathbf{b}}_n$, where $n = \arg \min_{n=k+1, \dots, M} \Delta_n[i]$ for (13) and $n = \arg \max_{n=k+1, \dots, M} \Delta_n[i]$ for (14), respectively. Therefore, for any $i = 1, \dots, k$, the maximum of $v_j[i]$ in the scene

corresponds to one of the undetected endmembers $\hat{\mathbf{b}}_n \in \{\hat{\mathbf{b}}_{k+1}, \dots, \hat{\mathbf{b}}_M\}$.

The abovementioned property also holds for the AAs vs_j and vr_j since, according to (11) and (12), we have

$$\begin{aligned} vs_j &= \left| 1 - \sum_{i=1}^k \left(s_j^a[i] + \sum_{n=k+1}^M s_j^b[n] \Delta_n[i] \right) \right| \\ &= \left| \sum_{n=k+1}^M s_j^b[n] \left(1 - \sum_{i=1}^k \Delta_n[i] \right) \right| \\ &\leq \max_{n=k+1, \dots, M} \left| 1 - \sum_{i=1}^k \Delta_n[i] \right| \end{aligned} \quad (15)$$

$$\begin{aligned} vr_j &= \left\| \mathbf{A}_k s_j^a + \mathbf{B} s_j^b - \mathbf{A}_k (s_j^a + \Delta s_j^b) \right\|_2 \\ &= \left\| \sum_{n=k+1}^M s_j^b[n] (\hat{\mathbf{b}}_n - \mathbf{A}_k \Delta_n) \right\|_2 \\ &\leq \max_{n=k+1, \dots, M} \left\| \hat{\mathbf{b}}_n - \mathbf{A}_k \Delta_n \right\|_2 \end{aligned} \quad (16)$$

where the equalities in (15)–(16) hold only in the case of $\mathbf{x}_j = \hat{\mathbf{b}}_n$, where $n = \arg \max_{n=k+1, \dots, M} \left| 1 - \sum_{i=1}^k \Delta_n[i] \right|$ for (15) and $n = \arg \max_{n=k+1, \dots, M} \left\| \hat{\mathbf{b}}_n - \mathbf{A}_k \Delta_n \right\|_2$ for (16), respectively.

Since, as proved in (13)–(16), the maximum of each AA $v_j[1], \dots, v_j[k], vs_j, vr_j$ corresponds to one of the undetected endmembers, and one can conclude that the AA values, as with the OSP values, can be used as a pixel purity measure for EE. In addition, it is noteworthy that since $\Delta_n = (\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T \hat{\mathbf{b}}_n$, vr_j in (16) can be rewritten as

$$\begin{aligned} vr_j &\leq \max_{n=k+1, \dots, M} \left\| \hat{\mathbf{b}}_n - \mathbf{A}_k (\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T \hat{\mathbf{b}}_n \right\|_2 \\ &\leq \max_{n=k+1, \dots, M} \left\| \mathbf{P}_{\mathbf{A}_k}^\perp \hat{\mathbf{b}}_n \right\|_2 \end{aligned} \quad (17)$$

which is equivalent to the OSP defined in (5) (i.e., $p = 2$). Thus, in fact, OSP is a special case of AAs that fail to meet (P3).

The objectives of OSP and AA projection are consistent, i.e., projecting the data points onto a predefined subspace in which the endmembers and background can be separated more easily. The difference is that AA projection considers all the cases of AAs and thus offers a more complete and robust way of identifying pure pixels of undetected endmembers.

III. SALIENCY-BASED FRAMEWORK FOR AUTONOMOUS ENDMEMBER DETECTION

In this section, the SAED algorithm is described in detail. To begin with, the main notations used in the following discussion are summarized in Table I.

The basic framework of SAED is shown in Fig. 1, where the HSI is first iteratively projected into the AA subspace, and then, the most salient undetected endmember target is identified via object-based saliency analysis in the AA subspace, and the iteration is terminated (i.e., the VD is determined) when there are no salient objects in the scene. This approach is motivated by the fact that undetected endmembers are visually salient in the AA subspace because they tend to have larger AA

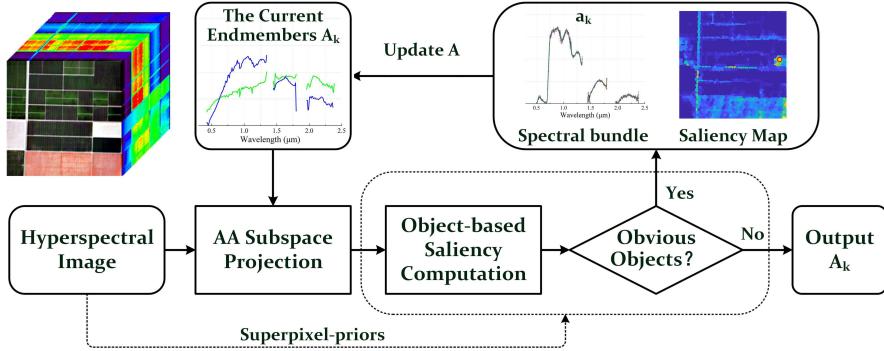


Fig. 1. Flowchart of the proposed SAED algorithm.

TABLE I
NOTATIONS USED IN THIS ARTICLE

$X \in \mathbb{R}^{L \times m \times n}$	Hyperspectral data cube
X_1, L, X_p	Superpixels obtained by SLIC
$\mathbf{A}_k \in \mathbb{R}^{L \times k}$	Current endmember matrix
$\mathbf{V}_k \in \mathbb{R}^{(k+2) \times N}$	Column-wise AA matrix corresponding to \mathbf{A}_k
$\mathbf{C} = [c_1, L, c_N] \in \mathbb{R}^{1 \times N}$	Column-wise location-based saliency map
$\mathbf{D} \in \mathbb{R}^{1 \times N}$	Column-wise superpixel-based saliency map
D_p	Saliency score of pixels inside superpixel X_p
X_{win}	Winner (most salient) superpixel in the scene
\mathbf{a}_{win}	Endmember extracted from X_{win}

values than the background (i.e., mixed pixels and pure pixels of detected endmembers), as proved in (13)–(16); In addition, the pure pixels of the same endmember tend to gather in a local area due to the spatial correlation. To highlight the endmember targets with varied sizes, SAED introduces a superpixel prior to the saliency analysis, and the EE is formulated as a salient region detection problem [29]–[31]. The VD is autonomously determined in the same framework via the superpixel-based “objectness” criteria, which are designed to determine whether there are any obvious objects in the scene. When salient superpixels still exist, SAED generates a superpixel-based saliency map, where the most salient superpixel is utilized to update the current endmember matrix for the AA subspace projection. In contrast, if there are no salient superpixels, this indicates that the HSI can be well reconstructed by the current endmember set. The iteration is then terminated, and the final number and the spectra of the endmembers are determined.

A. Superpixel Prior

Motivated by the fact that endmembers exist in the form of superpixels in remotely sensed HSIs, superpixels are used as the basic units for the saliency computation, and the EE problem is formulated as a salient region detection problem. For an HSI $X \in \mathbb{R}^{L \times m \times n}$, we first perform oversegmentation to generate superpixels $\{X_1, \dots, X_p\}$, where P is the number of superpixels in the scene. There are many means of superpixel segmentation, which can be roughly cate-

gorized into graph-based [32]–[34] and gradient ascent based approaches [35]–[37]. Here, considering the computational and memory efficiency, a typical example of the gradient ascent-based approaches—simple linear iterative clustering (SLIC) [37]—is introduced to generate superpixels. The original SLIC algorithm is a localized k -means clustering method, where the similarity between a pixel and a cluster center is measured in a 5-D spectral (color) and spatial space. There are only two input parameters required for SLIC: the desired size (or number) of superpixels and a compactness parameter used to control the weights between the spectral and spatial distances.

The original SLIC algorithm [37] cannot be applied directly to HSIs, due to the high dimensionality, and some changes are necessary for processing HSIs. One way is to transform the HSI into a 3-D base image and then apply the original SLIC algorithm, where the base image can be obtained through dimensionality reduction or band selection [38]. The other way is to measure the spectral distance in the hyperspectral space via a similarity measure, such as the Euclidean distance [39] or the spectral angle distance (SAD) [17]. In this article, in order to maintain the spectral details, the modified SLIC algorithm¹ [17] is applied for the segmentation, where the spectral similarity between the pixels and cluster centers is measured by the SAD. Note that although SAED extracts endmembers iteratively, the superpixels are only generated once at the preprocessing stage. To generate oversegmented superpixels, the average size of superpixels was set to 5 as default in the experiments.

B. Salient Object Detection

The objective of the salient object detection is to generate a superpixel-based saliency map, where pixels inside a superpixel share the same saliency score, and the most salient superpixel, which corresponds to an undetected endmember target, is highlighted.

To obtain the superpixel-based saliency map, SAED first generates a location-based saliency map, which integrates the saliency information of the undetected endmembers that are contained in the various AA channels. As proved in (13)–(16), for a current endmember matrix $\mathbf{A}_k \in \mathbb{R}^{L \times k}$, the AAs of a pixel

¹https://github.com/Xinyu-Wang/SGSNMF_TGRS

are directly proportional to the undetected endmember content, and thus, the saliency of a location can be directly quantified by its AAs. For simplicity, the AA subspace corresponding to \mathbf{A}_k is denoted as $\mathbf{V}_k = [\mathbf{v}_1, \dots, \mathbf{v}_N] \in \mathbb{R}^{(k+2) \times N}$, where $N = m \times n$ is the number of pixels in the HSI, and each column $\mathbf{v}_j = [v_j[1], \dots, v_j[k], v_s[j], v_r[j]]^T$ represents the AA vector of \mathbf{x}_j . Since different rows of \mathbf{V}_k correspond to different cases that fail to satisfy (P1)–(P3), they can have various dynamic ranges. Motivated by [40] and [41], a normalization operator $\mathcal{N}(\cdot)$ is introduced to reweight the different AA channels before feature integration, which consists of the following steps.

- 1) Normalize the AAs that correspond to (P1)–(P3), respectively, to the range $[0, 1]$, where the first k channels of \mathbf{V}_k corresponding to (P1) share the same normalization parameters.
- 2) Search for the global maximum AA value v_i in each channel of \mathbf{V}_k , and compute the average AA value \bar{v}_i of the other local maxima [40].
- 3) Set the AA values in each channel of \mathbf{V}_k that are less than \bar{v}_i to zero for background suppression.

The purpose of the first step of $\mathcal{N}(\cdot)$ is to adjust the dynamic range of the different kinds of AAs, and the last two steps are aimed at removing the visual redundancies since only the pixels with strong AA values are useful for saliency computation. Note that since the AA values in each channel are adjusted globally, the properties in (13)–(16) still hold for the normalized AAs.

We then combine the normalized AA maps into a location-based saliency map, $\mathbf{C} = [c_1, \dots, c_N] \in \mathbb{R}^{1 \times N}$, via a point-by-point addition operation

$$\mathbf{C} = \mathbf{1}^T \cdot \mathcal{N}(\mathbf{V}_k) \quad (18)$$

where $\mathbf{1} \in \mathbb{R}^{k+2}$ represents a column vector whose entries are equal to one. The location-based saliency map is a comprehensive representation of the AAs, which integrates all the pertinent information of the various AAs. Up to now, each point in \mathbf{C} has been calculated independently of its surroundings, and the spatial-contextual information of the endmembers has not been incorporated in the saliency computation. Thus, if undetected endmembers exist in the scene, the maximum of \mathbf{C} is an undetected endmember under the noiseless argument but may also be an outlier when the data are noisy.

To promote the robustness, the superpixel prior is used to distinguish undetected endmember targets from distractors, such as noise and outliers. The superpixel-based saliency map, denoted by $\mathbf{D} \in \mathbb{R}^{1 \times N}$, is generated based on the location-based saliency map \mathbf{C} and the superpixel prior, by forcing the pixels inside each superpixel X_p to share the same saliency score D_p

$$D_p = \frac{1}{n_p} \sum_{\mathbf{x}_j \in X_p} c_j \quad (19)$$

where n_p is the number of pixels inside X_p and D_p is equal to the average of the location-based saliency scores of the pixels inside. The superpixel with the highest saliency score is identified as a “winner” superpixel as follows:

$$X_{\text{win}} = X_\ell, \quad \ell = \arg \max_{p=1, \dots, P} D_p \quad (20)$$

and if undetected endmembers still exist in the scene, X_{win} is the most likely candidate.

C. Superpixel-Based “Objectness” Criteria

In SAED, two superpixel-based “objectness” criteria are introduced to terminate the iteration and determine the VD for unmixing, when there are no salient targets in the saliency maps.

Since real HSIs are inevitably contaminated with noise, the AA values will never be zero, even when all the endmembers have been accurately identified. Therefore, it is always possible to find a winner superpixel, regardless of whether an undetected endmember exists or not. The superpixel-based “objectness” criteria are utilized to determine whether the winner superpixel X_{win} is caused by an undetected endmember. If so, X_{win} is utilized to update the current endmember matrix \mathbf{A}_k and the iteration continues. Otherwise, this indicates that the HSI can be well reconstructed by the current endmember matrix, and the VD for unmixing is determined and the iteration is terminated.

Both the proposed objectness criteria are built based on the spatial distribution differences between endmembers and noise. Compared with noise and outliers with a sparse distribution, an endmember target will usually be gathered in a local area (i.e., an oversegmented and homogenous superpixel) due to the spatial correlation. Therefore, if the winner superpixel X_{win} is caused by an undetected endmember, the saliency scores of the pixels inside X_{win} will be uniform and close to the global maxima. In contrast, if X_{win} is caused by noise/outliers, the maximum AAs will be randomly scattered over the whole image, instead of being gathered in a local superpixel. Therefore, X_{win} will not be visually salient in the scene.

According to the abovementioned property, the first superpixel-based objectness criterion, τ , is introduced as follows:

$$\tau = C_{\text{max}} / D_{\text{win}} \quad (21)$$

where D_{win} is the average saliency score of X_{win} and $C_{\text{max}} = \sum_{j=1}^{n_l} \hat{c}_j / n_l$ is the average of the n_l largest scores in the whole scene, with n_l the number of pixels in X_{win} . The criterion τ is aimed at measuring the degree of spatial aggregation of the maximum saliency scores. When X_{win} is caused by an undetected endmember, the maximum saliency scores will be gathered together in a local superpixel, and thus, C_{max} and D_{win} will be approximately equal so that the ratio τ will be close to one. In contrast, if X_{win} is caused by noise, the highest saliency scores $\{\hat{c}_1, \dots, \hat{c}_{n_l}\}$ will be randomly scattered throughout the image, and thus, C_{max} will be a few times greater than D_{win} so that the ratio τ will be far greater than one. Therefore, according to the significant gaps of τ between the cases where undetected endmembers exist or not, a cutoff ratio γ can be utilized to determine whether endmember targets exist or not, and the iteration can be terminated if $\tau > \gamma$.

Moreover, to promote the robustness, a more general criterion, i.e., the mean RE, is introduced to terminate the iteration

as follows:

$$\varepsilon_{\text{win}} = \frac{1}{n_l} \sum_{\mathbf{x}_j \in X_{\text{win}}} vr_j \quad (22)$$

where it should be noted that the RE ε_{win} is calculated over the winner superpixel X_{win} , instead of over the whole HSI, since this can better reflect the reconstruction performance of the current endmember set \mathbf{A}_k . When ε_{win} is lower than a very small tolerance tol , this indicates that the most abnormal superpixel can be well reconstructed by \mathbf{A}_k , as can the other superpixels.

In SAED, the abovementioned two superpixel-based objectness criteria work together to determine whether the winner superpixel X_{win} can be attributed to an undetected endmember. Specifically, X_{win} is identified as an endmember only in the case that both $\tau < \gamma$ and $\varepsilon_{\text{win}} > tol$ hold.

D. Endmember Purification

In each iteration, SAED first locates the most salient endmember target X_{win} via the superpixel-based saliency analysis, and then, if X_{win} is identified as an undetected endmember target, the purest pixels inside X_{win} are used to generate a purified spectral signature \mathbf{a}_{win} . Specifically, we introduce a purification operation to extract a representative spectral signature \mathbf{a}_{win} from X_{win} to update the current endmember set. Specifically, we first decompose $\mathbf{X}_{\text{win}} \in \mathbb{R}^{L \times n_{\text{win}}}$ using singular value decomposition (SVD), which can be written as follows:

$$\mathbf{X}_{\text{win}} \mathbf{X}_{\text{win}}^T = \mathbf{U} \Sigma \mathbf{U}^T \quad (23)$$

where \mathbf{X}_{win} consists of the homogeneous pixel $\mathbf{x}_j \subseteq X_{\text{win}}$, for which $c_j \geq \sum_1^n c_j / n_l$, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_L)$ is the eigenvalue matrix, and $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_L] \in \mathbb{R}^{L \times L}$ is the eigenvector matrix. The eigenvector \mathbf{u}_1 corresponding to the largest eigenvalue σ_1 is then utilized to represent \mathbf{a}_{win} as follows:

$$\mathbf{a}_{\text{win}} = \mu \cdot \mathbf{u}_1 \quad (24)$$

where $\mu = \|\mathbf{X}_{\text{win}}\|_F / (n_{\text{win}})^{1/2}$ controls the scale of \mathbf{a}_{win} and \mathbf{u}_1 controls the spectrum of \mathbf{a}_{win} . Generally speaking, \mathbf{a}_{win} contains the primary statistics of the pure pixels in X_{win} and is thus more robust to noise than a single pixel in the scene.

For ease of understanding, we summarize the SAED algorithm in Algorithm 1, where endmembers are successively detected one after another, until the stopping criteria are satisfied. Note that the first endmember \mathbf{a}_1 is initialized by (4) in Step 2 but then replaced by \mathbf{a}_{win} in the first iteration, which is identified via the saliency analysis, and is thus more robust. In SAED, X_{win} is identified as an endmember only in the case that both $\tau < 2$ and $\varepsilon_{\text{win}} > 10^{-3}$ hold, which was applied for all the images in the experiments. Specifically, $\tau < 2$ means that more than 50% of the most salient points are located in the winner superpixel X_{win} , which is seldom true if X_{win} is caused by noise that is randomly distributed in the scene. The other criterion $\varepsilon_{\text{win}} \leq 10^{-3}$ works as a supplement to the first criterion. It indicates that the HSI can be well reconstructed using the current endmembers, and thus, the iteration can be

terminated. This criterion is widely used for terminating the iteration in blind HU.

In summary, SAED achieves robust autonomous EE in three steps: 1) superpixel-based saliency analysis in the AA subspace is applied to determine the accurate locations of the undetected endmembers in the scene; 2) two superpixel-based objectness criteria are applied to determine whether the saliency is really caused by an undetected endmember or not; and 3) if the winner superpixel is caused by an undetected endmember, a denoising operator is applied on the pure pixels inside the winner superpixel to obtain a new purified endmember signature. It should be noted that the locations and the number of endmembers are identified from the AAs iteratively so that SAED naturally establishes the interactions between the VD, the endmembers, and the abundances. This approach guarantees that the estimated endmember matrix can unmix the data well, with low RE, and there will be no salient undetected targets in the AA map.

Algorithm 1 SAED

Input: $X \in \mathbb{R}^{L \times m \times n}$, γ , and tol

- 1: Generate superpixels $\{X_1, \dots, X_P\}$
- 2: Initialize $\mathbf{A} \in \mathbb{R}^{L \times 1}$ using (4).
- 3: **for** $k = 1 : M_{\text{max}}$
- 4: Generate the AA matrix using (7) and (10)–(12);
- 5: Generate location-based and superpixel-based saliency maps using (18) and (19), respectively;
- 6: Calculate τ and ε_{win} according to (21) and (22);
- 7: If $\tau < \gamma$ and $\varepsilon_{\text{win}} > tol$
- 8: Update the k -th column \mathbf{a}_k of \mathbf{A} by \mathbf{a}_{win} of the winner superpixel according to (20), (23), and (24);
- 9: else
- 10: Break;
- 11: end if
- 12: **end for.**
- 13: **Output:** \mathbf{A}

IV. SYNTHETIC EXPERIMENTS

The proposed SAED algorithm was compared with the state-of-the-art VD estimation and EE methods using two synthetic data sets with different spatial patterns. These experiments were aimed at demonstrating the ability of the different methods to estimate the real number and spectra of the endmembers in various conditions (i.e., various signal-to-noise ratios (SNRs) and different ground-truth VD). As for the compared methods, HySime [22], NWHFC [4], and GENE [20] were selected for the VD estimation; OSP [11], SPP-N-FINDR [42], minimum-volume simplex analysis (MVSA) [8], and SED [25] were selected for the endmember detection; and SPICE [14] was used for both. After publication, the source code of SAED will be shared on the homepage of the author.²

As for the assessment indicators, the VD estimation was directly compared with the ground-truth VD, and the EE

²<http://rsidea.whu.edu.cn/e-resources.html>

accuracy was evaluated using the SAD

$$\text{SAD}_i = \arccos \frac{\hat{\mathbf{a}}_i^T \mathbf{a}_i}{\|\hat{\mathbf{a}}_i\|_2 \|\mathbf{a}_i\|_2} \quad (25)$$

where \mathbf{a}_i and $\hat{\mathbf{a}}_i$ are the ground truth and the estimated spectral signature of the i th endmember, respectively. Note that in the following results, the SAD results are the average of 20 random tests, and their units are in degrees.

A. Synthetic Data Sets

Two synthetic data sets—Synth-1 and Synth-2—were used for the evaluation. Both were mixed according to the LMM, with the endmember matrices consisting of mineral spectral signatures with 224 bands over the spectral range of 0.4–0.25 μm , randomly selected from the U.S. Geological Survey (USGS) spectral library. The difference between Synth-1 and Synth-2 lies in the spatial patterns of their abundances and the number of endmembers in the HSIs.

- 1) Synth-1 is a $224 \times 75 \times 75$ hyperspectral cube containing five endmembers [43]. The abundance maps of each endmember are shown in Fig. 2. As shown, in Synth-1, there are five rows of blocks in the scene, where the blocks in the first row are pure regions that contain only one of the endmembers in the scene, and the blocks in the other rows are mixed regions that consist of two-to-five kinds of endmembers.
- 2) Synth-2 consists of a set of $224 \times 128 \times 128$ hyperspectral cubes, where the VD of the data varies from 5 to 14, with an interval of 3. The abundance maps of Synth-2 were generated in the Hyperspectral Imagery Synthesis toolbox, http://www.ihu.es/ccwintco/index.php/Hyperspectral_Imagery_Synthesis_tools_for_MATLAB which provides functions for users to adjust the image size, endmember size, and the VDs of the synthetic HSIs. The abundances of Synth-2 follow a spherical Gaussian field distribution, with an adjustable parameter $\kappa \in [1, 200]$ used to control the endmember sizes. To simulate real scenarios, for each VD, $M \in \{5, 8, 11, 14\}$, Synth-2 contains 20 HSIs, for which κ varies from 10 to 200, with an interval of 10 so that the abundance maps have varied endmember sizes. The abundance maps (i.e., $M = 5$) with small and large endmember sizes are shown in Fig. 3, where parameter $\kappa = 10$ for Fig. 3(a)–(e) and $\kappa = 200$ for Fig. 3(f)–(j). In addition, it should be noted that in each abundance map, there is only one pure pixel existing for all the Synth-2 HSIs.

To simulate HSIs in a noisy environment, random white Gaussian noise or Poisson noise was added to the synthetic data sets, where the SNR of the data was defined as follows:

$$\text{SNR} = 10 \log_{10} \frac{E[\mathbf{x}^T \mathbf{x}]}{E[\varepsilon^T \varepsilon]} \quad (26)$$

where \mathbf{x} and ε denote the original signal and noise, respectively.

TABLE II

SENSITIVITY OF SAED TO THE SEGMENTATION SCALE WITH SYNTH-2, WHERE THE OPTIMAL ESTIMATION IN EACH SNR IS MARKED IN RED

SNR (dB)	Scale 3		Scale 5		Scale 7	
	SAD	VD	SAD	VD	SAD	VD
50	1.77±0.39	5.10±0.31	1.86±0.37	5.10±0.31	2.12±0.39	5.25±0.44
40	1.82±0.43	5.15±0.37	1.95±0.45	5.10±0.31	2.23±0.45	5.20±0.52
30	2.21±0.37	5.00±0.00	2.17±0.44	5.05±0.22	2.41±0.46	5.10±0.31
20	3.81±0.50	5.25±0.30	3.44±0.46	5.00±0.00	3.21±0.56	5.05±0.22

TABLE III

ROBUSTNESS OF THE EE WITH SYNTH-1 TO GAUSSIAN NOISE, WHERE THE TOP-THREE ACCURACIES ARE MARKED IN RED, GREEN, AND BLUE, RESPECTIVELY

SNR (dB)	OSP	MVSA	SED	SPICE	SPP-N-FI NDR	SAED
50	0.19±0.00	0.15±0.03	0.04±0.00	2.12 ± 0.73	0.18 ± 0.00	0.04±0.00
40	0.60±0.01	0.48±0.08	0.12±0.00	2.25 ± 1.41	0.58 ± 0.01	0.12±0.00
30	1.88 ± 0.04	1.58±0.12	0.39±0.02	2.51 ± 0.92	1.84 ± 0.04	0.37±0.01
20	6.08 ± 0.08	4.66±1.12	1.43±0.28	3.80 ± 0.64	5.88 ± 0.14	1.37±0.13

B. Sensitivity of SAED to the Segmentation Scale

This experiment was aimed at analyzing the sensitivity of SAED to the segmentation scale by using 20 different Synth-2 HSIs (i.e., $M = 5$) with different average endmember sizes and polluted by white Gaussian noise with $\text{SNR} \in \{20, 30, 40, 50\}$ dB. Since SAED assumes that the HSI is overestimated by homogenous superpixels, three oversegmentation scales—scales 3, 5, and 7—were compared in terms of the SAD and VD estimation accuracies, where the scale w denotes the mean width of the SLIC superpixels.

The experimental results are listed in Table II, where each value consists of the mean and the standard deviation of 20 times random tests, and the optimal estimation in each SNR is marked in red. As shown in Table II, in each SNR, the SAD and the VD accuracies are similar and stable, with small standard deviations. The results indicate that SAED is not sensitive to the segmentation scale when the HSI is oversegmented. This is mainly because the SLIC superpixel is a function of the target size and the segmentation scale, and the superpixels of small targets can still be identified even when the scale is large. In the experiments, scale 5 was set as the default for all the real and simulated data, for a fair comparison with the state-of-the-art algorithms.

C. Robustness of the EE to Noise

In this experiment, the Synth-1 data set was used to evaluate the EE accuracy, where the VD was equal to 5 and the data were contaminated by white Gaussian noise or Poisson noise with $\text{SNR} \in \{20, 30, 40, 50\}$ dB. The experimental results with different levels of Gaussian noise and Poisson noise are shown in Tables III and IV, respectively, where the mean SAD is shown with its standard deviation, and the top-three results are highlighted in different colors. As shown, the accuracy trends of the different algorithms with Gaussian noise and Poisson noise are almost the same, and the main factor for

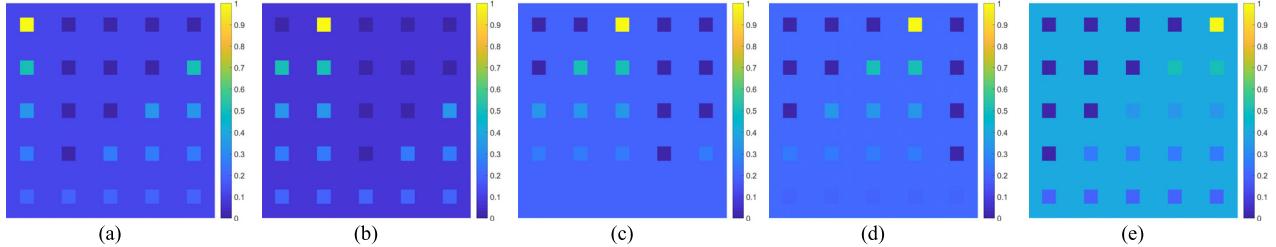


Fig. 2. Abundance maps of Synth-1. (a) #1. (b) #2. (c) #3. (d) #4. (e) #5.

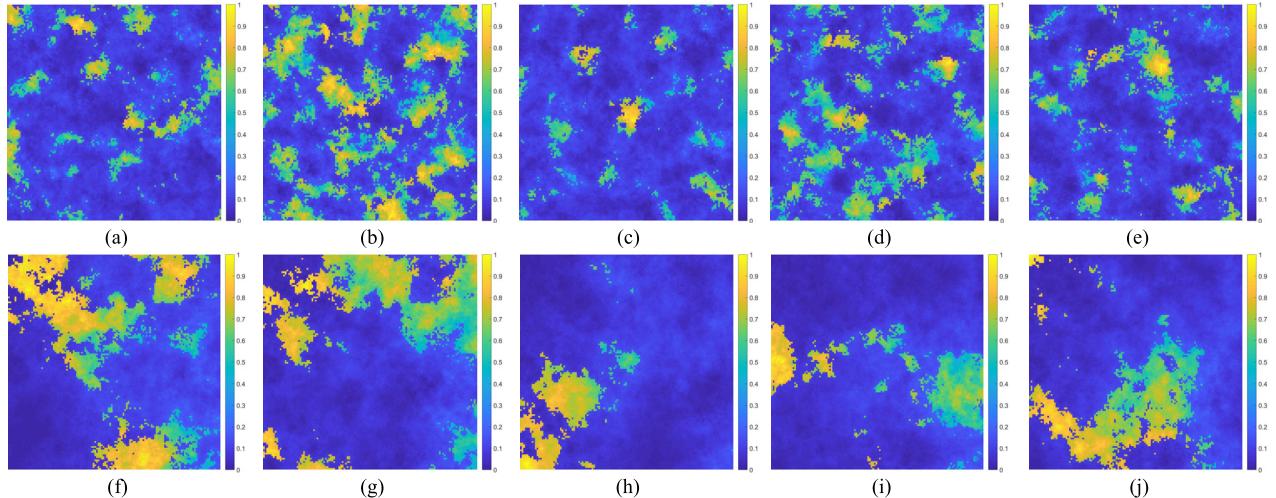


Fig. 3. Abundance maps of Synth-2 ($M = 5$) with different endmember sizes, where (a)–(e) $\kappa = 10$ and (f)–(j) $\kappa = 200$.

TABLE IV

ROBUSTNESS OF THE EE WITH SYNTH-1 TO POISSON NOISE, WHERE THE TOP-THREE ACCURACIES ARE MARKED IN RED, GREEN, AND BLUE, RESPECTIVELY

<i>SNR</i> (dB)	<i>OSP</i>	<i>MVSA</i>	<i>SED</i>	<i>SPICE</i>	<i>SPP-N-FI</i> <i>NDR</i>	<i>SAED</i>
50	0.18 ± 0.01	0.13 ± 0.03	0.04 ± 0.00	2.99 ± 0.83	0.18 ± 0.00	0.04 ± 0.00
40	0.58 ± 0.01	0.41 ± 0.04	0.12 ± 0.00	2.25 ± 0.85	0.56 ± 0.01	0.12 ± 0.00
30	1.85 ± 0.03	1.34 ± 0.12	0.41 ± 0.01	1.78 ± 0.59	1.78 ± 0.04	0.36 ± 0.02
20	6.76 ± 0.33	4.92 ± 0.27	1.37 ± 0.41	7.61 ± 0.61	5.70 ± 0.11	1.34 ± 0.18

the SAD results is the SNR of the data. SAED is more robust to noise than the other methods for two reasons: one is that the spatial patterns of the endmembers are utilized so that the endmembers can be located more precisely; and the other is that each endmember signature is obtained via a purification operation, where the statistics of a bundle of pure pixels are explored. In contrast, each endmember extracted by OSP and SPP-N-FINDR corresponds to a point in the contaminated HSI, and thus, OSP and SPP-NFINDR are more sensitive to noise. In addition, it can be seen that the results of SED and SAED in this experiment are similar. This is mainly because the endmember targets in Synth-1 are of the same size and relatively small, which satisfies the assumption of SED that endmembers are rare and distinctive in the scene.

D. Robustness of the EE to the Ground-Truth VD

In this experiment, the Synth-2 data set was applied to analyze the robustness of the EE accuracy to the number of endmembers since the ground-truth VD is tunable. Note that,

TABLE V

ROBUSTNESS OF THE EE TO THE NUMBER OF ENDMEMBERS WITH SYNTH-2, WHEN THE SNR = 30 dB, WHERE THE TOP-THREE ACCURACIES ARE MARKED IN RED, GREEN, AND BLUE, RESPECTIVELY

<i>VDS</i>	<i>OSP</i>	<i>MVSA</i>	<i>SED</i>	<i>SPICE</i>	<i>SPP-N-FI</i> <i>NDR</i>	<i>SAED</i>
5	4.41 ± 0.92	5.34 ± 1.13	3.45 ± 0.65	2.64 ± 0.98	3.29 ± 0.27	2.17 ± 0.44
8	3.70 ± 0.60	10.21 ± 1.07	2.23 ± 0.60	2.62 ± 0.78	2.29 ± 0.12	1.24 ± 0.15
11	4.01 ± 1.37	15.73 ± 2.96	2.53 ± 0.87	5.38 ± 2.25	2.27 ± 0.17	1.68 ± 0.31
14	7.05 ± 1.33	17.97 ± 2.73	3.31 ± 0.62	8.11 ± 2.52	3.63 ± 0.45	2.29 ± 0.47

as shown in Fig. 3, for each VD, the average endmember sizes of Synth-2 were varied during the 20 random tests. Table V lists the SAD results of the different EE algorithms, where the SNRs of Synth-2 were set to 30 dB, and the number of endmembers was varied from 5 to 14 with an interval of 3. Since SNR = 30 dB, the main factor that affected the EE accuracy was the number of endmembers. As shown, the EE algorithms that incorporate the spatial patterns of the endmembers, such as SPP-N-FINDR, SED, and SAED, are more robust to the ground-truth VD of the data, compared with the spectral-only EE algorithms, such as OSP, SPICE, and MVSA.

The endmember targets in Synth-2 were of various sizes and shapes for each VD so that this experiment also verified the sensitivity of the EE to the endmember sizes. As shown in Table V, compared with SED, SAED obtains better average accuracies and smaller standard deviations, which indicates

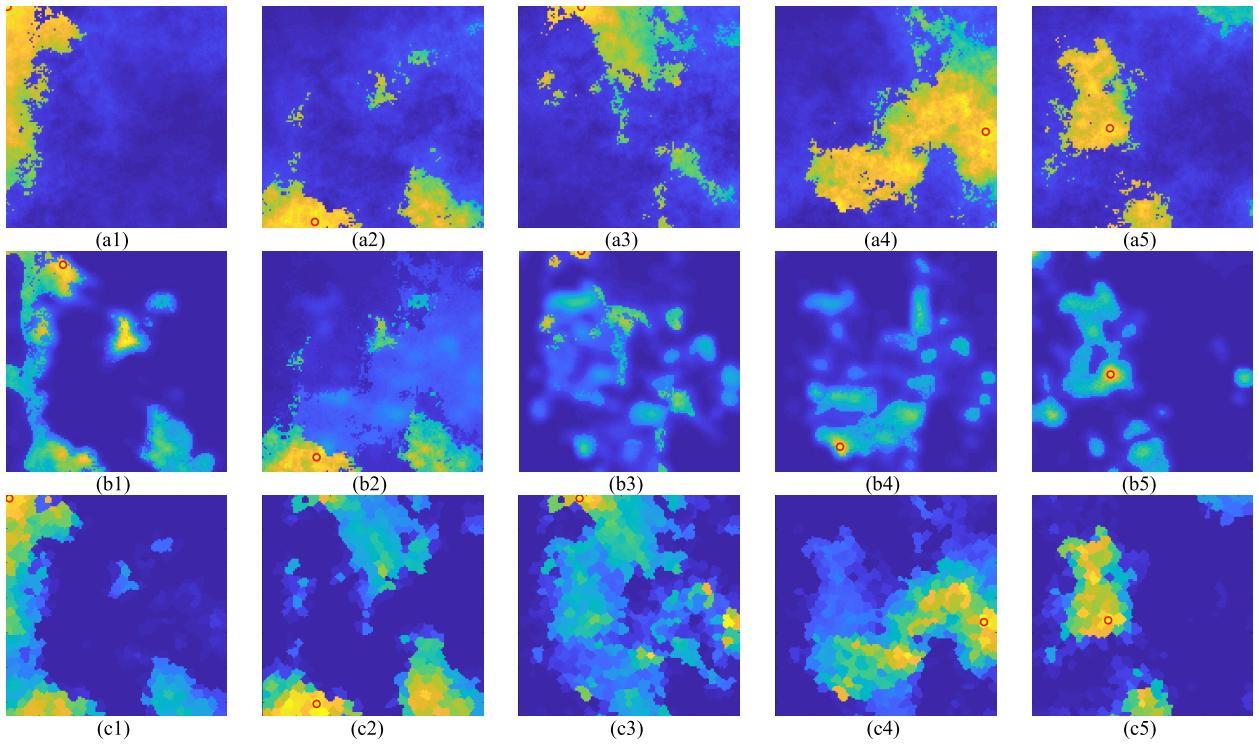


Fig. 4. Comparison of saliency maps generated by SED and SAED, where (a1)–(a5) are the ground-truth abundance maps, (b1)–(b5) are the final saliency maps of SED, and (c1)–(c5) are the superpixel-based saliency maps of SAED.

that the superpixel prior used in SAED performs better than the local irregularity prior used in SED, from the aspect of highlighting endmember targets with various sizes. In addition, to illustrate the difference between the saliency priors, the saliency maps generated by SED and SAED are compared by using a Synth-2 HSI (i.e., $M = 5$, $\kappa = 180$, and SNR = 30 dB) in Fig. 4, where (a1)–(a2), (b1)–(b5), and (c1)–(c5), respectively, represent the ground-truth abundance maps, the saliency maps of SED, and the saliency maps of SAED. The ground truth and the estimated endmembers are highlighted by red circles. In each iteration, the saliency map describes the spatial distribution of several undetected endmembers, where relatively pure pixels are highlighted. As shown in Fig. 4(b1) and (b4), when the endmember targets are of large sizes, SED tends to highlight some relatively pure pixels located at boundaries due to its local irregularity prior. In contrast, as shown in Fig. 4(c1) and (c4), the superpixel prior in SAED is not sensitive to the endmember size so that the locations of the endmembers can be determined more accurately.

E. Robustness of the VD Estimation

In this experiment, we controlled the SNR and the ground-truth VD of Synth-2 to analyze their influence on the VD estimations of the different algorithms. Specifically, the ground-truth VD for Synth-2 was set to $M \in \{5, 8, 11, 14\}$, and the HSIs (with different endmember sizes) were randomly polluted with white Gaussian noise of $\text{SNR} \in \{20, 30, 50\}$ dB. The difficulty of the VD estimation increases with the level of noise and the number of endmembers in the scene. This is

because the abovementioned factors reduce the distinguishability of the different endmembers.

The VD estimation results are shown in Table VI, where $\text{PF} \in \{10^{-3}, 10^{-4}, 10^{-5}\}$ is the false alarm probability used in NWHFC and GENE. In Table VI, each VD result consists of the mean and standard deviation values of 20 experiments, and for ease of comparison, the relatively accurate VD estimations (i.e., where the average VD is close to reality, with the standard deviation being less than 1 for 30 and 50 dB and 2 for 20 dB) are marked in red. In general, the number of accurate estimations decreases with the increase of the noise level and the ground-truth VD. Overall, GENE-AH and SAED obtain the best overall VD estimation accuracies in this experiment. SPICE determines the VD by iteratively pruning unnecessary endmembers when their maximum abundances drop below a predefined threshold. In this experiment, we fixed the threshold of SPICE for all the cases, and it can be observed that SPICE fails to prune endmembers under a low SNR (20 dB) because the serious noise contamination results in the maximum abundances always being larger than the predefined threshold. As shown, the VDs estimated by GENE-AH and NWHFC decrease with the false alarm probability PF. Although relatively accurate VDs can be estimated by GENE-AH, it is difficult to establish an optimal PF that works for all SNR and VD cases. It should be noted that the main difference between SAED and the compared VD methods is that the spatial-contextual information of the endmembers in the low-dimensional AA subspace is utilized for determining the VD so that SAED can estimate relatively accurate VDs even in the low-SNR cases where the endmembers are difficult to distinguish from noise when using spectral properties alone.

TABLE VI

COMPARISON OF THE VD ESTIMATION OF THE DIFFERENT ALGORITHMS, WITH $M \in \{5, 8, 11, 14\}$ AND $SNR \in \{20, 30, 50\}$ dB, WHERE THE RELATIVELY ACCURATE VD ESTIMATIONS ARE MARKED IN RED

SNR (dB)	<i>Method</i>	P_F	<i>The number of endmembers M</i>			
			5	8	11	14
50	<i>NWHFC</i>	10^{-3}	4.15±0.75	4.85±0.81	9.30±1.08	8.20±1.24
		10^{-4}	4.05±0.69	4.70±0.80	9.10±1.02	7.50±1.28
		10^{-5}	4.00±0.65	4.45±0.89	8.95±0.94	6.95±1.36
	<i>GENE-AH</i>	10^{-3}	7.15±0.99	9.30±0.98	11.60±0.68	14.80±0.89
		10^{-4}	5.65±0.67	8.45±0.60	11.15±0.37	14.15±0.37
		10^{-5}	5.15±0.49	8.05±0.22	11.00±0.00	14.05±0.22
	<i>SPICE</i>	-	5.45±0.60	8.10±0.79	10.85±0.59	12.65±1.14
	<i>HySime</i>	-	5.00±0.00	8.00±0.00	11.00±0.00	13.95±0.22
	<i>SAED</i>	-	5.10±0.31	8.25±0.44	11.55±0.69	14.65±0.81
30	<i>NWHFC</i>	10^{-3}	4.15±0.74	4.80±0.89	9.20±1.11	6.70±1.56
		10^{-4}	4.05±0.69	4.65±0.82	9.00±1.12	6.00±1.49
		10^{-5}	4.00±0.65	4.40±0.94	8.85±1.04	5.60±1.35
	<i>GENE-AH</i>	10^{-3}	7.60±1.14	9.50±0.88	11.95±0.94	13.90±0.97
		10^{-4}	5.95±1.10	8.65±0.81	11.15±0.37	13.10±0.55
		10^{-5}	5.35±0.48	8.10±0.44	11.00±0.32	12.75±0.55
	<i>SPICE</i>	-	5.50±0.68	9.15±1.18	11.90±0.97	13.20±1.32
	<i>HySime</i>	-	5.00±0.00	7.85±0.36	10.85±0.37	11.40±0.60
	<i>SAED</i>	-	5.05±0.22	8.10±0.31	11.20±0.41	13.75±0.95
20	<i>NWHFC</i>	10^{-3}	4.15±0.75	4.05±0.83	8.35±1.18	3.30±0.86
		10^{-4}	4.05±0.69	3.85±0.67	7.85±1.27	2.80±1.01
		10^{-5}	4.00±0.65	3.60±0.68	7.35±1.27	2.45±1.00
	<i>GENE-AH</i>	10^{-3}	7.35±1.04	9.85±2.16	10.35±1.31	11.25±1.71
		10^{-4}	6.40±0.75	8.15±1.39	9.55±0.83	9.60±1.64
		10^{-5}	5.60±0.68	7.05±0.51	9.15±0.49	8.45±0.64
	<i>SPICE</i>	-	19.80±0.62	20.00±0.00	20.00±0.00	19.95±0.22
	<i>HySime</i>	-	4.15±0.44	6.10±0.45	7.95±0.22	7.45±0.69
	<i>SAED</i>	-	5.00±0.00	7.40±0.78	10.35±0.78	10.75±1.25

V. REAL-DATA EXPERIMENTS

In the real-data experiments, two real data sets captured by spaceborne and airborne hyperspectral sensors were applied to evaluate the VD estimation and EE performance of the different algorithms.

A. AVIRIS Cuprite Data Set

The Cuprite data set is a benchmark data set for spectral unmixing [44] that was captured by NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the Cuprite mining site, Nevada, USA, in 1997. The original Cuprite data set contains 224 spectral bands with a 20-m spatial resolution and a 0.10- μ m spectral resolution over the spectral range of 0.4–2.5 μ m. As shown in Fig. 5(a), a subset of the Cuprite data set was selected in this experiment, containing 250 × 190 pixels, and 188 bands were selected after removing the low-SNR bands (1–3, 106–115, 151–169, and 221–224).

The VD estimations for the Cuprite data set and the second real data set are listed in Table VII, where the VD estimation of SAED is 13 for the Cuprite data set. It can be seen in Table VII that the VD estimated by the different methods varies greatly from 9 to 34. Since the ground-truth VD in the Cuprite data set is not known, it is difficult to comment directly on the accuracy of SAED. In previous studies, the VD for unmixing of the

TABLE VII

COMPARISON OF THE VD ESTIMATION FOR THE DIFFERENT REAL DATA SETS

<i>Method</i>	P_F	<i>Dataset</i>	
		<i>Cuprite</i>	<i>CA-Cropland</i>
<i>NWHFC</i>	10^{-3}	11.00±0.00	9.00±0.00
	10^{-4}	10.00±0.00	9.00±0.00
	10^{-5}	9.00±0.00	7.00±0.00
<i>GENE-AH</i>	10^{-4}	34.00±0.00	45.00±0.00
	10^{-6}	31.00±0.00	41.00±0.00
	10^{-8}	27.00±0.00	37.00±0.00
<i>SPICE</i>	-	9.60±1.37	5.05±0.22
<i>HySime</i>	-	19.00±0.00	29.00±0.00
<i>SAED</i>	-	13.00±0.00	5.00±0.00

TABLE VIII

SAD COMPARISON USING THE CUPRITE DATA SET, WHERE THE TOP-THREE ACCURACIES ARE MARKED IN RED, GREEN, AND BLUE, RESPECTIVELY

<i>Endmember</i>	<i>OSP</i>	<i>MVSA</i>	<i>SED</i>	<i>SPICE</i>	<i>SPP-NFINDR</i>	<i>SAED</i>
<i>Alunite</i>	4.599	5.259	5.144	7.317	5.220	5.400
<i>Andradite</i>	4.758	3.612	3.830	4.767	7.048	6.776
<i>Buddingtonite</i>	6.530	4.291	6.929	10.840	6.530	4.894
<i>Dumortierite</i>	4.139	6.166	3.756	4.733	3.909	3.611
<i>Kaolinite #1</i>	4.714	5.089	5.635	3.879	4.713	4.348
<i>Kaolinite #2</i>	3.790	3.508	3.530	7.683	3.641	3.640
<i>Montmorillonite</i>	4.857	5.134	4.702	7.500	5.139	8.181
<i>Muscovite</i>	3.214	3.913	3.385	3.335	3.429	3.360
<i>Nontronite</i>	3.986	4.094	3.798	5.271	3.986	3.938
<i>Pyrope</i>	13.616	9.428	6.221	3.019	4.398	4.186
<i>Sphene</i>	5.281	4.404	3.670	3.621	14.089	3.029
<i>Chalcedony</i>	3.909	14.706	7.650	3.197	4.593	3.747
<i>Average</i>	5.283	5.800	4.854	5.432	5.558	4.593

Cuprite data set has not been fixed, and 14 [5], [8], 10 [9], 9 [13], and 12 [17] have been the frequently used VD values. The Cuprite data set can be well represented with a very small RE by 9–14 endmembers. Hence, the VD estimation of SAED can be considered as reasons for this data set.

To evaluate the EE performance, 12 mineral spectral signatures from the USGS library, as shown in Fig. 5(b), were selected as references, according to [17] and [45], i.e., Alunite, Andradite, Buddingtonite, Muscovite, Chalcedony, Dumortierite, Kaolinite_1, Kaolinite_2, Montmorillonite, Nontronite, Pyrope, and Sphene. Note that since the number of endmembers identified by SAED is 13, for a fair comparison, we extracted the same number of endmembers using the compared methods and then selected the 12 most closely matched ones to calculate the SAD values. The SAD comparison is provided in Table VIII, where the top-three accuracies are marked in red, green, and blue, respectively. As shown, SAED, SED, and OSP, respectively, obtain the top three average SAD accuracies.

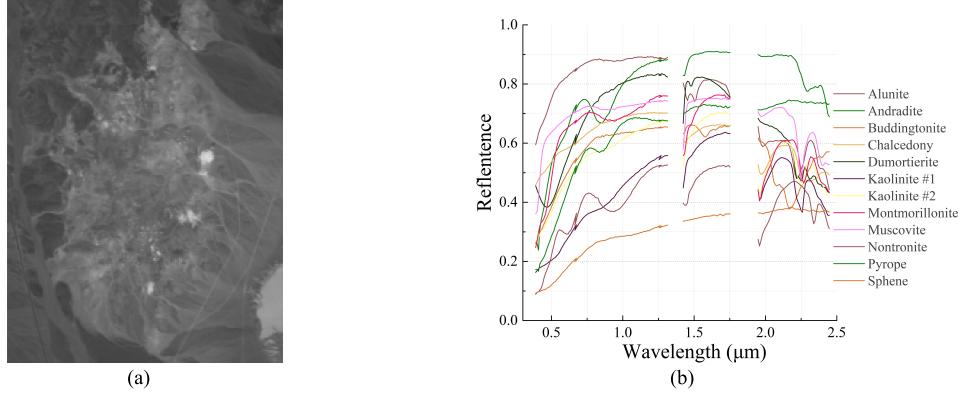


Fig. 5. AVIRIS Cuprite data set. (a) Band 66 (998 nm). (b) 12 reference spectra.

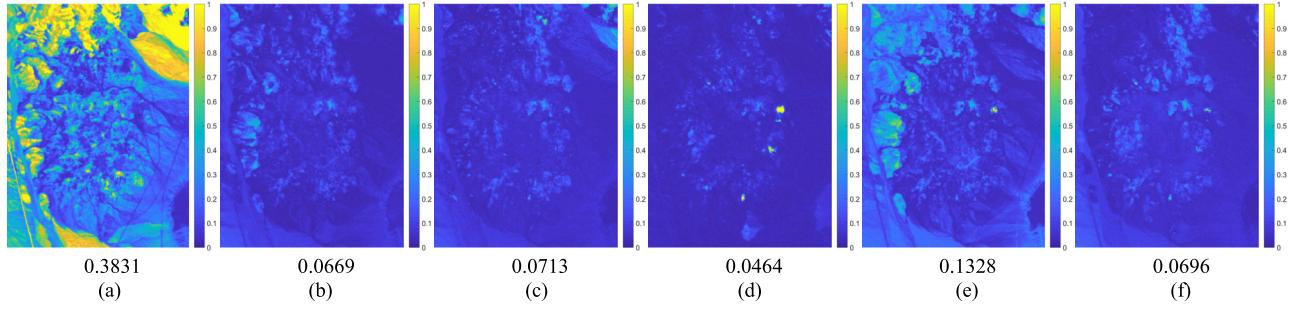


Fig. 6. Reconstruction error maps for the Cuprite data set, where the abundances were estimated by the FCLS method, and the average RMSE values are listed below each map. (a) OSP. (b) MVSA. (c) SED. (d) SPICE. (e) SPP-NFINDR. (f) SAED.

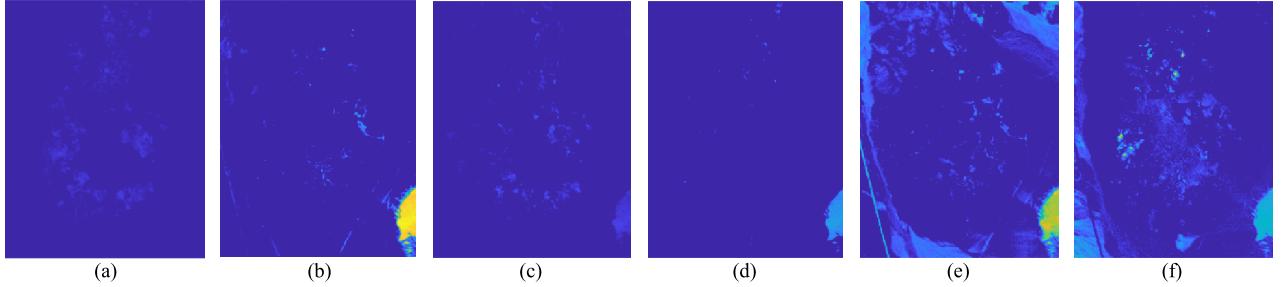


Fig. 7. Normalized AA maps of SAED, where (a)–(d), (e), and (f) correspond to the AAs failing to satisfy (P1), (P2), and (P3), respectively. (a) AA#1. (b) AA#2. (c) AA #3. (d) AA #4. (e) AA #5. (f) AA #6.

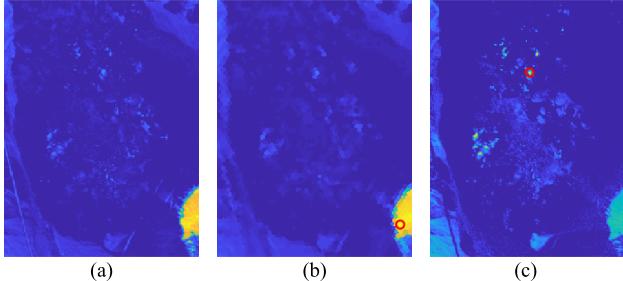


Fig. 8. Saliency maps and OSP map obtained by the same endmembers. (a) Location-based saliency map. (b) Superpixel-based saliency map. (c) OSP map (AA#6).

To further compare the performances, the RE maps generated by the different EE algorithms are compared in Fig. 6, where the fully constrained LS (FCLS) method [10] was used to estimate the abundances, for a fair comparison, and the average root-mean-square errors (RMSEs) are listed below

each map. As discussed in Section II, if all the endmembers are well estimated, there will be no obvious target in the RE map. As shown in Fig. 6(a), OSP fails to identify the endmember in the upper-right corner so that it has the maximum RMSE. SPICE obtains the best average RMSE because the average RMSE is minimized during its optimization but, as shown in Fig. 6(d), SPICE fails to identify a small endmember target (Muscovite), which is visually salient in the RE map. Overall, the endmembers extracted by MVSA, SED, and SAED show the best reconstruction performances.

In addition, the intermediate results of SAED, i.e., the normalized AA maps, in the case of the current endmember matrix containing four endmembers, are shown in Fig. 7, where (a)–(d), (e), and (f) correspond to the AAs failing to satisfy (P1), (P2), and (P3), respectively. The AA maps, in practice, illustrate the rough distributions of some of the undetected endmembers. Fig. 8 compares the differences between the location-based saliency map, the superpixel-based saliency map, and the OSP map obtained by the same endmember matrix, where the winner superpixel and the maximum OSP

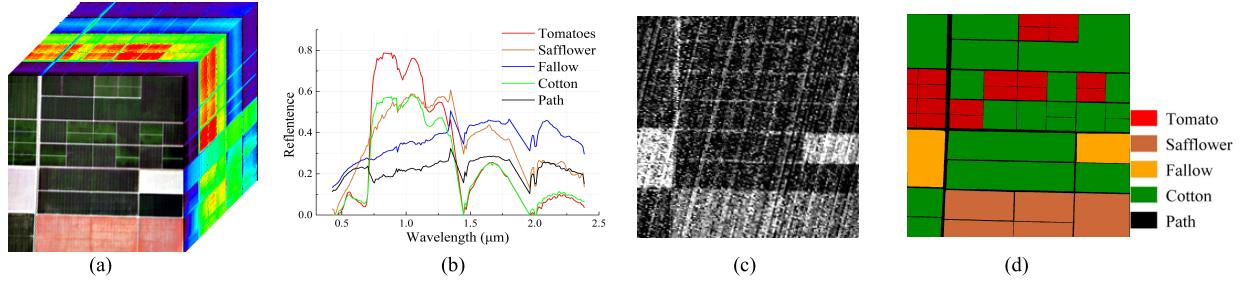


Fig. 9. EO-1 Hyperion CA-Cropland data set. (a) Hyperspectral cube. (b) Endmember spectra. (c) Band 171 (2385 nm). (d) ROIs.

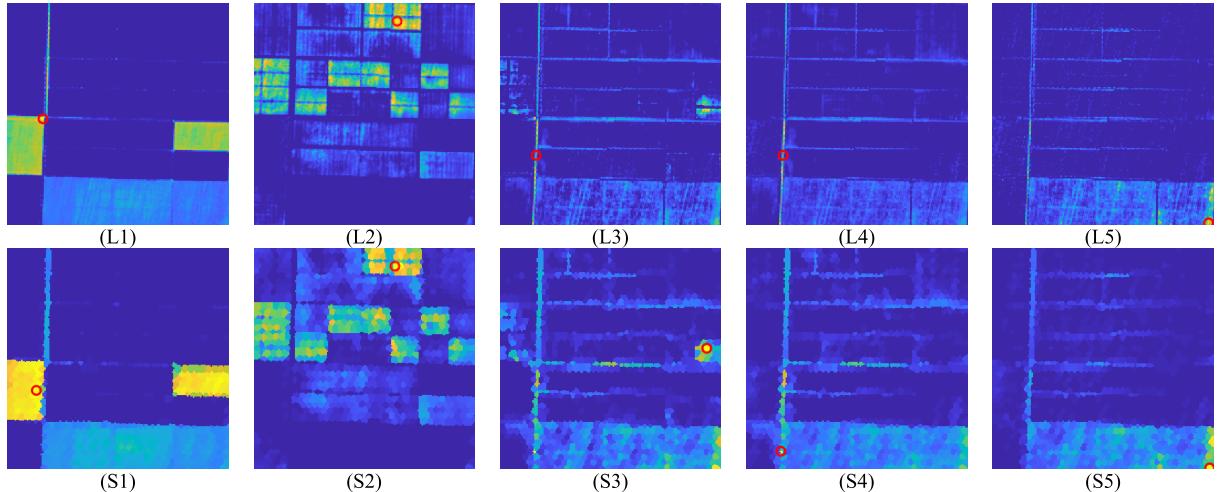


Fig. 10. Location- and superpixel-based saliency maps of SAED in different iterations, where (Li) and (Si) denote the location-based and superpixel-based saliency maps obtained in the i th iteration, respectively.

TABLE IX

SAD COMPARISON USING THE CA-CROPLAND DATA SET, WHERE THE TOP-THREE ACCURACIES ARE MARKED IN RED, GREEN, AND BLUE, RESPECTIVELY

Endmember	OSP	MVSA	SED	SPICE	SPP-N-F INDR	SAED
Tomato	3.078	12.209	2.789	2.300	3.078	2.577
Safflower	7.503	9.253	2.363	3.265	3.437	1.670
Fallow	2.675	5.425	2.147	5.556	2.785	2.257
Cotton	11.889	28.878	17.710	9.128	30.256	7.307
Path	17.437	13.403	10.600	20.889	10.361	7.831
Average	8.516	13.834	7.122	8.228	9.983	4.329

pixel are marked by the red circle. It can be seen from Fig. 8 that OSP and SAED detect different endmembers for the same \mathbf{A}_k . This is mainly because the saliency map, which integrates important information from various AAs, is a more comprehensive representation of the undetected endmembers.

B. CA-Cropland Data Set

The CA-Cropland data set was captured by the Hyperion sensor onboard the Earth Observing-1 (EO-1) satellite over an agricultural region in Kings, CA, USA, on July 29, 2009. Hyperion is a line-scan sensor that collects 242 bands over the spectral range of $0.4\text{--}2.5 \mu\text{m}$, with a 30-m spatial resolution.

The original CA-Cropland data set is the Level 1T product downloaded through the USGS EarthExplorer³ website. Before unmixing, the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) [46] software was applied for the atmospheric correction, through which most of the bad bands (1–7, 58–78, 121–129, 166–180, and 224–242) of the original data were identified and removed. The final CA-Cropland data set used for the HU contained 171 bands and 250×190 pixels.

The data cube and the regions of interest (ROIs) of the CA-Cropland data set are shown in Fig. 9(a) and (d), respectively, where the distribution of the vegetation shown in Fig. 8(d) was drawn by the authors according to the cropland cover data provided by the United States Department of Agriculture (USDA⁴) and high spatial resolution data provided by Google Earth for the same region and time period. As shown, the CA-Cropland data contain five kinds of endmembers, i.e., cotton, safflower, tomato, fallow, and path, and the reference spectrum of each in Fig. 8(b) is the mean spectrum of 20 pure pixels randomly selected from the image. Note that the FLAASH software does not detect and remove all the low-SNR bands so that there are still a small number of bands corrupted by noise. A typical example of the low-SNR bands is shown in Fig. 8(c).

³<https://earthexplorer.usgs.gov/>

⁴<https://nassgeodata.gmu.edu/CropScape/>

TABLE X
RUN TIME COMPARISON USING THE CUPRITE AND CA-CROPLAND DATA SETS (SECONDS)

<i>Method</i>	<i>VD estimation</i>			<i>EE</i>				<i>Autonomous EE</i>	
	<i>NWHFC</i>	<i>GENE-AH</i>	<i>HySime</i>	<i>OSP</i>	<i>MVSA</i>	<i>SED</i>	<i>SPP-NFINDR</i>	<i>SPICE</i>	<i>SAED</i>
<i>Cuprite</i>	0.12	24.08	1.14	0.96	107.85	10.53	43.42	21.32	5.54
<i>CA-Cropland</i>	0.13	50.42	1.18	0.25	6.96	2.16	15.65	33.02	5.67

It should be noted that the existence of noise increased the difficulty of both the VD estimation and EE.

The VD estimation results for the CA-Cropland data are provided in Table VII, where the VD estimations of SAED, SPICE, and NWHFC are the closest to the reality. In contrast, the eigenanalysis-based methods, such as HySime (i.e., 29) and GENE-AH (i.e., 37–49), tend to overestimate the VD of the CA-Cropland data set, and the estimated VD might be applied to dimensionality reduction or other applications, but not blind HU.

As for the EE, an SAD comparison of the different EE algorithms is shown in Table IX, where the number of endmembers was set to five for all the EE algorithms. It can be seen from the table that SAED obtains the best SAD results. It is also interesting to note that MVSA and SPP-N-FINDR fail to detect all the vegetation spectra, i.e., tomato and cotton, when $M = 5$, due to the relatively high spectral correlation of the different vegetation types, but the small spectral differences can be identified by SAED through the saliency analysis.

In addition, to better illustrate the effect of the superpixel prior in SAED, the location- and superpixel-based saliency maps in each iteration are shown and compared in Fig. 10, where the most salient locations and superpixels are marked by red circles. It can be seen from Fig. 10(L1) and (S1) that compared with the location-based saliency map, the superpixel-based saliency map highlights certain regions. In addition, some maximum points scattered at the boundaries, which might be noise, are restrained.

C. Efficiency Analysis Using the Real Data Sets

In this experiment, the run time of the VD estimation algorithms (i.e., NWHFC, GENE-AH, and HySime), the EE algorithms (i.e., OSP, MVSA, and SED), and the autonomous EE algorithms (i.e., SPICE and SAED) were compared using MATLAB 2017a on a personal computer equipped with an Intel Core i7-8565U CPU at 1.80 GHz and 16-GB RAM. Table X lists the run times of the different algorithms in seconds. As shown, the geometry-based algorithms, such as GENE-AH, MVSA, SPP-NFINDR, and SPICE, have a relatively high computational complexity due to their complex optimization and large iteration numbers. As shown in Table X, for the autonomous EE methods, SAED is superior to SPICE in terms of run time in both data sets. SAED is also superior to SED in the Cuprite data set where the number of endmembers is large. This is mainly because the local irregularity prior used in SED is calculated per iteration, while the superpixel prior used in SAED is calculated only once. As for the proposed SAED algorithm, most of the run time

is for generating the superpixel prior, and the other operations have a very low computational complexity.

VI. CONCLUSION

In this article, to fully explore the spatial and spectral properties, an SAED algorithm has been proposed to simultaneously estimate the number and the spectral signatures of endmembers. This article was motivated by the fact that if an undetected endmember target exists, it will manifest as the maximum AA value and will be visually obvious in the scene. If not (i.e., all the endmembers have been detected), there will be no salient objects in the AA subspace. In SAED, the EE is defined as a superpixel-based saliency detection problem in the AA subspace, and the VD is determined autonomously via superpixel-based objectness criteria when there are no salient targets present in the AA subspace. Since both the spectral and spatial properties of the endmembers are exploited during the autonomous EE, SAED can provide a relatively stable and accurate endmember matrix for blind HU, which was verified by using both real and simulated data. In addition, as the number and locations of the endmembers are determined according to the AAs, SAED naturally establishes a connection between the VD, the endmember spectra, and the abundances, which guarantees that the autonomously estimated endmember matrix can decompose the data well.

In real applications, SAED can realize autonomous detection of endmembers without prior knowledge of the number of endmembers and is applicable to most of the current spaceborne and airborne HSIs whose spatial resolution varies from $1 \times 1 \text{ m}^2$ to $30 \times 30 \text{ m}^2$. It is also noteworthy that as a spatial–spectral EE algorithm, SAED performs well when pure pixels are present in the scene, with a relatively high spatial correlation, but it may fail in the opposite case, due to the lack of pure pixels.

REFERENCES

- [1] 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.
- [2] N. Keshava and J. F. Mustard, “Spectral unmixing,” *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [3] W.-K. Ma *et al.*, “A signal processing perspective on hyperspectral unmixing: Insights from remote sensing,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.
- [4] 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.
- [5] J. M. P. Nascimento and J. M. B. 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] M. E. Winter, "N-FINR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–275, Oct. 1999.
- [7] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.
- [8] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.
- [9] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1946–1961, Apr. 2016.
- [10] 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.
- [11] H. Ren and C. I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.
- [12] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.
- [13] 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.
- [14] A. Zare and P. Gader, "Sparsity promoting iterated constrained end-member detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] X. Wang, Y. Zhong, L. Zhang, and Y. Xu, "Blind hyperspectral unmixing considering the adjacency effect," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6633–6649, Sep. 2019.
- [19] C.-I. Chang, *Real-Time Recursive Hyperspectral Sample and Band Processing*. Cham, Switzerland: Springer, 2017.
- [20] A. Ambikapathi, T.-H. Chan, C.-Y. Chi, and K. Keizer, "Hyperspectral data geometry-based estimation of number of endmembers using p-norm-based pure pixel identification algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2753–2769, May 2013.
- [21] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.
- [22] 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.
- [23] J. Harsanyi, W. Farrand, and C.-I. Chang, "Determining the number and identity of spectral endmembers: An integrated approach using Neyman–Pearson eigentresholding and iterative constrained RMS error minimization," in *Proc. Thematic Conf. Geol. Remote Sens.*, 1993, p. 395.
- [24] C.-I. Chang, W. Xiong, H.-M. Chen, and J.-W. Chai, "Maximum orthogonal subspace projection approach to estimating the number of spectral signal sources in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 504–520, Jun. 2011.
- [25] X. Wang, Y. Zhong, Y. Xu, L. Zhang, and Y. Xu, "Saliency-based endmember detection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3667–3680, Jul. 2018.
- [26] D. R. Thompson, L. Mandrake, M. S. Gilmore, and R. Castaño, "Superpixel endmember detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4023–4033, Nov. 2010.
- [27] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.
- [28] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 73–80.
- [29] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 883–890.
- [30] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 2083–2090.
- [31] D. Li, S. Li, and H. Li, "Hyperspectral image unmixing based on sparse and minimum volume constrained nonnegative matrix factorization," in *Pattern Recognition*. Berlin, Germany: Springer, 2014, pp. 44–52.
- [32] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [33] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
- [34] A. P. Moore, S. J. D. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
- [35] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [36] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [37] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [38] X. Zhang, S. E. Chew, Z. Xu, and N. D. Cahill, "SLIC superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery," *Proc. SPIE*, vol. 9472, May 2015, Art. no. 947209.
- [39] X. Cui, Y. Tian, L. Weng, and Y. Yang, "Anomaly detection in hyperspectral imagery based on low-rank and sparse decomposition," in *Proc. 5th Int. Conf. Graph. Image Process. (ICGIP)*, Jan. 2014, Art. no. 90690R.
- [40] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [41] M. W. Cannon and S. C. Fullenkamp, "A model for inhibitory lateral interaction effects in perceived contrast," *Vis. Res.*, vol. 36, no. 8, p. 1115, 1996.
- [42] M. Zortea and A. Plaza, "Spatial preprocessing for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2679–2693, Aug. 2009.
- [43] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
- [44] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *Proc. Summaries 3rd JPL Airborne Geosci. Workshop*, 1992, pp. 1–3.
- [45] 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.
- [46] G. P. Anderson *et al.*, "FLAASH and MODTRAN4: State-of-the-art atmospheric correction for hyperspectral data," in *Proc. IEEE Aerosp. Conf.*, Mar. 1999, pp. 177–181.



Xinyu Wang (Member, IEEE) received the B.S. degree in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2014, and the Ph.D. degree in communication and information systems from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, in 2019.

He is an Associate Research Fellow with the School of Remote Sensing and Information Engineering, Wuhan University. His major research interests range over blind source separation, hyperspectral unmixing, and target detection.



Yanfei Zhong (Senior Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

Since 2010, he has been a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. He organized Intelligent Data Extraction, Analysis and Applications of Remote Sensing (RSIDEA) research group. He has authored more than 100 research articles in international journals, such as *Remote Sensing of Environment*, *ISPRS Journal of Photogrammetry and Remote Sensing*, and *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*. His research interests include hyperspectral remote sensing information processing, high-resolution remote sensing image understanding, and geoscience interpretation for multisource remote sensing data and applications.

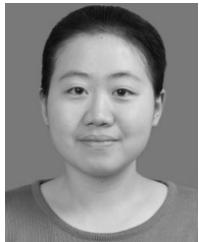
Dr. Zhong is a fellow of the Institution of Engineering and Technology (IET). He was a recipient of the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). He won the Second-Place Prize in 2013 IEEE GRSS Data Fusion Contest and the Single-view Semantic 3-D Challenge of the 2019 IEEE GRSS Data Fusion Contest, respectively. He is serving as an Associate Editor for the *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*, and the *International Journal of Remote Sensing*.



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

He was a Principal Scientist for the China state key basic research project from 2011 to 2016 appointed by the ministry of national science and technology of China to lead the remote sensing program in China. He is a Chair Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University. He has authored more than 700 research articles and five books. He is the Institute for Scientific Information (ISI) highly cited author. He holds 30 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a fellow of the the Institution of Engineering and Technology (IET). He was a recipient of the 2010 Best Paper Boeing Award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS) and the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) student paper contest in recent years. He also serves as an Associate Editor or an Editor of more than ten international journals. He is serving as an Associate Editor of the *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*. He is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter.



Chunyang Cui received the B.S. degree in surveying and mapping engineering from Northeastern University, Shenyang, China, in 2018. She is pursuing the M.S. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China.

Her major research interests include the hyperspectral unmixing and blind source separation.



Yanyan Xu received the B.S. degree in mechanics from the Southwest University of Science and Technology, Mianyang, China, in 1995, the M.S. degree in electronics from the HuBei University of Technology, Wuhan, China, in 2000, and the Ph.D. degree in communication and information system from Wuhan University, Wuhan, in 2007.

She is a Professor of the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. She has authored more than 40 research articles and one book. Her research interests include multimedia communication systems, information security.