

# Total Variation and Sparsity Regularized Decomposition Model With Union Dictionary for Hyperspectral Anomaly Detection

Tongkai Cheng<sup>ID</sup> and Bin Wang<sup>ID</sup>, *Senior Member, IEEE*

**Abstract**—Anomaly detection in hyperspectral imagery has been an active topic among the remote sensing applications. It aims at identifying anomalous targets with different spectra from their surrounding background. Therefore, an effective detector should be able to distinguish the anomalies, especially for the weak ones, from the background and noise. In this article, we propose a novel method for hyperspectral anomaly detection based on total variation (TV) and sparsity regularized decomposition model. This model decomposes the hyperspectral imagery into three components: background, anomaly, and noise. In order to distinguish effectively these components, a union dictionary consisting of both background and potential anomalous atoms is utilized to represent the background and anomalies, respectively. Moreover, the TV and the sparsity-inducing regularizations are incorporated to facilitate the separation. Besides, we present a new strategy for constructing the union dictionary with the density peak-based clustering. The proposed detector is evaluated on both simulated and real hyperspectral data sets and the experimental results demonstrate its superiority compared with several traditional and state-of-the-art anomaly detectors.

**Index Terms**—Anomaly detection, density peak-based clustering, hyperspectral imagery, sparsity-inducing regularization, total variation (TV).

## I. INTRODUCTION

HYPERSPECTRAL images with hundreds of narrow spectral bands can provide a wealth of spectral information of the ground surface [1]. Compared with multispectral images, the spectral curve for each pixel in hyperspectral images is nearly continuous, which makes possible the identification of different materials based on their distinct spectral signatures [2]. This discrimination power has led to the development of several applications, such as classification and target detection [1]–[4].

The main objective of target detection is to search for the presence of specific pixels. Based on the availability of a prior information of target, two types of target detection have

Manuscript received January 20, 2020; revised May 21, 2020 and June 17, 2020; accepted June 19, 2020. Date of publication July 3, 2020; date of current version January 21, 2021. This work was supported by the National Natural Science Foundation of China under Grant 61971141 and Grant 61731021. (Corresponding author: Bin Wang.)

The authors are with the Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China, and also with the Research Center of Smart Networks and Systems, School of Information Science and Technology, Fudan University, Shanghai 200433, China (e-mail: wangbin@fudan.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.2020.3004478

been extensively researched, i.e., supervised and unsupervised. When the spectral signature of target is known in advance, the target detection is generally performed using a matching strategy. When the spectral signature of target is unknown, the target detection is processed in an unsupervised fashion, namely anomaly detection. The aim of anomaly detection is to distinguish the objects that are anomalous with regard to the natural background. Usually, the anomalies occur as few pixels or mixed pixels and have spectral signatures different from those of background pixels, such as man-made objects. Since the spectral variability and the availability of target information are two main difficulties that need to be addressed in supervised target detection, anomaly detection is more consistent with real scenarios and has been proven valuable in many fields including search and rescue, mine detection, and military surveillance [4], [5]. In this article, we focus on anomaly detection.

In recent years, many methods for hyperspectral anomaly detection have been proposed [6]. Among them, the Reed-Xiaoli (RX) detector [7] is considered as the benchmark, which assumes that the background follows a multivariate Gaussian distribution. Two statistics, i.e., mean vector and covariance matrix, are computed to represent the background. Then, the Mahalanobis distance between the pixel under test and the background is utilized to measure the probability of the test pixel to be part of background. The global-RX (GRX) and local-RX (LRX) are two widely used versions of RX. The GRX estimates the background statistics with all pixels in the scene, while LRX performs estimation using neighboring pixels of each test pixel. However, in real hyperspectral images, the Gaussian assumption made in RX is hard to be satisfied. Moreover, the computed statistics are susceptible to the contamination of the anomalies and noise. Several improved detectors based on RX were developed to address these issues. The kernel-RX (KRX) [8], [9] projects the observed data to a high-dimensional feature space through a nonlinear mapping and thus can model a much more complex distribution of the original data. The cluster KRX (CKRX) [10] which groups background pixels into clusters and applies a fast eigendecomposition algorithm is proposed to alleviate high computational complexity of KRX and can achieve comparable detection accuracy. The segmented-RX [11] also divides the hyperspectral imagery into several clusters; thus, the background in each cluster is more homogeneous.

The weighted-RX [12] and the blocked adaptive computationally efficient outlier nominator (BACON) [13] were proposed in order to improve the robustness of background estimation to the outliers. Nevertheless, the performance of these methods is limited because the specific distribution assumption of background is generally not valid.

In addition to the aforementioned methods, the representation-based detectors without any explicit assumptions on the background distribution have gained much attention recently. The main principle of these methods is that the background pixel can be represented by a dictionary, whereas the anomalies cannot. The sparse-representation-based (SR) anomaly detector proposed in [14] is based on the concept that a background pixel can be represented as a sparse linear combination of its spatial neighbors. Likewise, the collaborative-representation-based detector (CRD) [15] also assumes that background pixels can be represented by its spatial neighbors, while the collaboration is reinforced by  $\ell_2$ -norm minimization of the representation vector. The spatial neighbors of each test pixel are usually obtained with the sliding windows. Unfortunately, there is no general rule for choosing appropriate window sizes. Besides, these detectors deal with pixel vectors in an independent manner, and thus the correlations of all the pixels are ignored.

More recently, robust principal component analysis (RPCA) [16] as a subspace recovery technique has been employed in hyperspectral anomaly detection [17], in which the hyperspectral imagery is decomposed into low-rank and sparse components. To better characterize the multisubspace structure of hyperspectral data, the low-rank representation (LRR) [18] reconstructs the background with a dictionary and the residual is used to indicate the anomalies [19]. In [20] and [21], the local structure of background is also considered, which leads to a more accurate representation. However, in hyperspectral images, the spectra are always corrupted with noise. Moreover, some of the anomalous targets occupy a portion of the pixel area. The spectra of these subpixel anomalies slightly deviate from the background. Therefore, the above detectors based on the decomposition of background and anomalies are prone to false alarms caused by noise [22], [23]. Some approaches consider to model the anomalies and noise separately, such as the low-rank and sparse matrix decomposition (LRaSMD) technique [24], [25], but they still cannot effectively distinguish the weak anomalies from noise. In [26], the anomalies are represented by a potential anomaly dictionary. Some anomalies with significant difference from the background can be extracted and chosen to represent other anomalies hidden in the scene. The low-rank and sparse representation strategy is utilized to accomplish the decomposition. In this way, the anomalous part can be well separated from the noise, because the noise cannot be represented by the potential anomaly dictionary. The extended  $k$ -means clustering algorithm is used to group all the pixels into several clusters, and some representative pixels in each cluster are selected to construct the background and potential anomaly dictionaries, respectively. However, making the assumption of known cluster number is often unrealistic.

Besides, this detector uses only the spectral information, while the spatial-contexual information which concerns the relationship between each pixel and its spatial neighbors has been ignored. Previous works [27]–[29] have shown that incorporating spatial information has a positive effect on the background estimation.

In this article, we propose a total variation and sparsity regularized decomposition model (TVSDM) for hyperspectral anomaly detection. In the proposed model, the hyperspectral imagery is decomposed into three components: background, anomaly, and noise. In order to distinguish effectively between these components, we utilize a union dictionary that consists of both background and potential anomalous atoms to represent the background and anomalies, respectively. Moreover, the total variation (TV) and the sparsity regularizers that characterize the spatial smoothness of background and the spatial scarcity of anomalous targets, respectively, are incorporated to facilitate the separation. To construct the union dictionary, we present a new strategy based on a simple yet effective clustering algorithm called the density peak-based clustering [30]. Unlike other clustering approaches in which some involved parameters should be properly specified, the density peak-based clustering has its basis only in the similarity between data points. After all pixels have been grouped into several classes, some representative pixels in each cluster can be selected to form the background dictionary. On the other hand, some strong anomalous pixels can also be picked out easily to constitute the potential anomaly dictionary.

The main contributions in our proposed method can be concisely summarized as follows.

- 1) A decomposition model with union dictionary is proposed to effectively separate the anomalies from background and noise. The TV and sparsity regularizers are incorporated to characterize the properties of background and anomalies, respectively, leading to a more accurate separation. The experimental results demonstrate that the proposed method can evidently improve the detection accuracy compared with several traditional and state-of-the-art anomaly detectors.
- 2) In view of the difficulty of specifying appropriate parameters in most of clustering approaches, we propose a method for the construction of union dictionary based on the density peak-based clustering. By means of the density peak-based clustering, the representative background pixels in each class and some strong anomalies can be easily picked out to construct the background and potential anomaly dictionaries, respectively, without any preset parameter in clustering.

The rest of this article is organized as follows. Section II describes the related works, which lay a foundation of our work. The proposed method for anomaly detection is presented in Section III, including the construction of union dictionary, the decomposition model, and the optimization procedure. In Section IV, the experimental results on both simulated and real hyperspectral data sets are reported and analyzed. Finally, the conclusions are drawn in Section V.

## II. RELATED WORKS

### A. Density Peak-Based Clustering

The density peak-based clustering algorithm has its basis only in the similarity between data points [30]. The cluster centers are defined as local maxima in the density of data points and have the following two properties.

- 1) The cluster centers are surrounded by neighboring samples with lower local densities.
- 2) The cluster centers are at a relatively large distance from any other samples with higher local densities.

For each data point  $i$ , two quantities, i.e., the local density  $\rho_i$  and its distance  $\delta_i$  between point  $i$  and those with higher densities are computed. Both of these two quantities depend only on the distances  $d_{i,j}$ . Here, the Euclidean distance is considered to measure  $d_{i,j}$  between data points

$$d_{i,j} = \|\mathbf{y}_i - \mathbf{y}_j\|_2. \quad (1)$$

Then, the local density  $\rho_i$  could be calculated as follows:

$$\rho_i = \sum_j \chi(d_{i,j} - d_c) \quad (2)$$

where  $d_c$  is the cutoff distance which behaves as the radius of search region and  $\chi(x)$  is an indicator function with the following definition:

$$\chi(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{if } x \geq 0. \end{cases} \quad (3)$$

It can be easily found that  $\rho_i$  is actually the number of neighbors that are closer than  $d_c$  to the point  $i$ . After obtaining  $\rho_i$ ,  $\delta_i$  is measured by computing the minimum distance between point  $i$  and other points with higher densities

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{i,j}). \quad (4)$$

If the point  $i$  has the highest local density,  $\delta_i$  is conventionally taken  $\delta_i = \max_j (d_{i,j})$ . Then, the cluster centers can be recognized as points whose values of  $\delta_i$  are anomalously large.

The change of  $d_c$  only makes a difference on the relative magnitude of  $\rho_i$  in different data points, implying that the clustering results are stable under different choice of  $d_c$ . The cutoff distance  $d_c$  is often chosen so that the average number of neighbors is around 1%–2% of the total number of points [30].

With the local density  $\rho_i$  and distance  $\delta_i$  in hand, the cluster centers can then be found using a ranking procedure. As aforementioned, the cluster centers are those points with high  $\rho_i$  and anomalous large  $\delta_i$ . This observation is illustrated by a “decision graph” shown in Fig. 1(a) with simulated hyperspectral data, where  $\delta_i$  is plotted as a function of  $\rho_i$ . For details of this data, see the first simulated data set in Section IV. Four representative points marked with different colors in Fig. 1(a) are regarded as the cluster centers. To be specific, the cluster centers can be determined by the plot of a ranking score defined as

$$\gamma_i = \rho_i \times \delta_i. \quad (5)$$

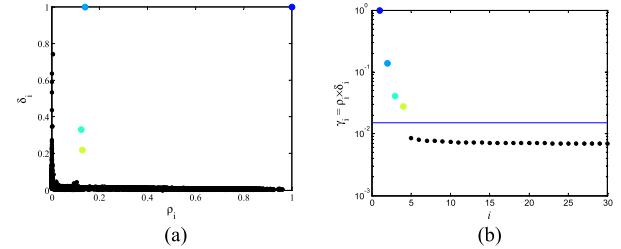


Fig. 1. Cluster analysis on the simulated hyperspectral data. (a) Decision graph, where the cluster centers are highlighted with different colors. (b) Ranking score  $\gamma_i = \rho_i \times \delta_i$  sorted in descending order.

Fig. 1(b) shows that this score is by definition large for cluster centers.

### B. Decomposition of Background, Anomaly, and Noise for Hyperspectral Imagery

Consider a hyperspectral cube with resized samples  $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^n \in \mathbb{R}^{d \times n}$ , where each column denotes a spectral vector with  $d$  bands and  $n$  stands for the total number of pixels. The data matrix  $\mathbf{Y}$  can be represented as a superposition of three parts: background, anomaly, and noise

$$\mathbf{Y} = \mathbf{L} + \mathbf{S} + \mathbf{N} \quad (6)$$

where  $\mathbf{L}$  represents the background,  $\mathbf{S}$  denotes the sparse component, namely anomalies, and  $\mathbf{N}$  stands for the noise. Some criteria can be imposed on matrices  $\mathbf{L}$  and  $\mathbf{S}$  to achieve a unique decomposition. For instance, the LRaSMD technique uses the low rankness and sparsity to characterize  $\mathbf{L}$  and  $\mathbf{S}$ , respectively, [24]. However, LRaSMD cannot separate these components effectively.

Generally, the background pixels in hyperspectral imagery can be represented by some other background pixels. On the other hand, some anomalous pixels with significant difference from the background can be detected and used as the potential prior of other anomalous pixels [26]. These strong anomalies can be chosen to represent the other anomalies hidden in the scene. Given the background dictionary  $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m] \in \mathbb{R}^{d \times m}$  with  $m$  atoms and the potential anomaly dictionary  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r] \in \mathbb{R}^{d \times r}$  with  $r$  atoms, all the pixels can be approximately represented by the atoms in the union dictionary  $[\mathbf{B} \cup \mathbf{A}]$

$$\mathbf{Y} = [\mathbf{B} \cup \mathbf{A}] \begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix} + \mathbf{N} = \mathbf{B}\mathbf{X} + \mathbf{A}\mathbf{Z} + \mathbf{N} \quad (7)$$

where  $\mathbf{X}$  and  $\mathbf{Z}$  are the representation coefficients associated with the background dictionary  $\mathbf{B}$  and potential anomaly dictionary  $\mathbf{A}$ , respectively. With reasonable constraints imposed on matrices  $\mathbf{X}$  and  $\mathbf{Z}$ , the background and the anomalous components should be well separated.

## III. PROPOSED METHOD

In this section, we firstly present an approach for the construction of union dictionary based on the density peak-based clustering. Then, the decomposition model utilizing the union dictionary is proposed to separate the hyperspectral imagery

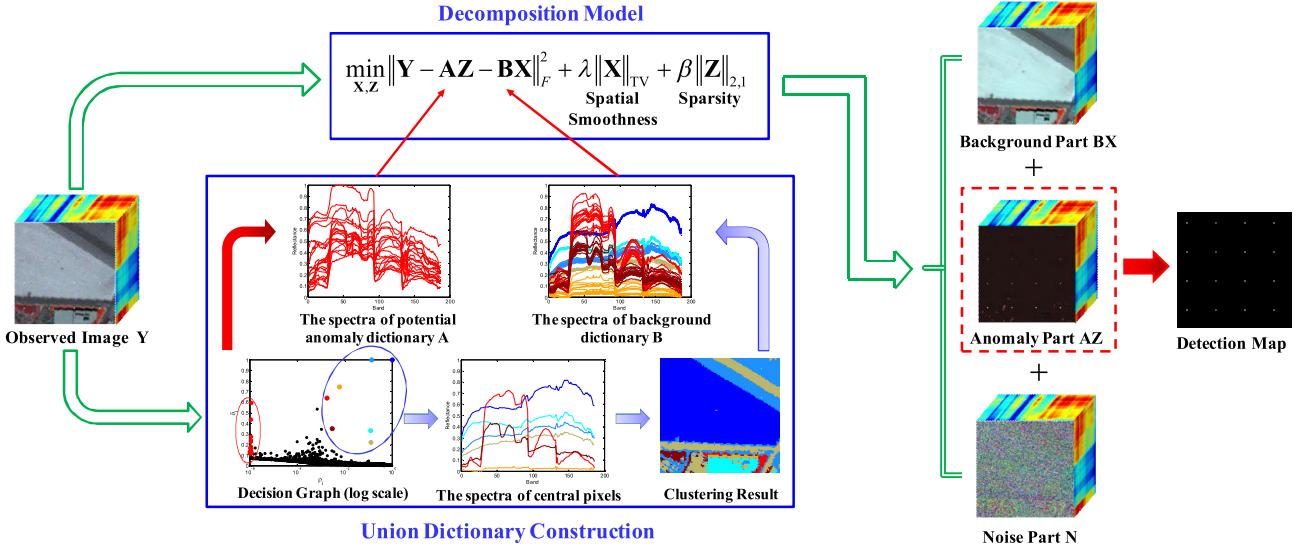


Fig. 2. Schematic flowchart of the proposed method for hyperspectral anomaly detection.

into three components, i.e., background, anomalies, and noise, in which the TV and sparsity regularizers are incorporated to facilitate the separation, followed by the optimization procedure via alternating direction method of multipliers (ADMM) [31]. The schematic flowchart of the proposed method for hyperspectral anomaly detection is shown in Fig. 2. In the following, each step of the proposed method will be described in detail.

#### A. Construction of Union Dictionary

In sparse representation for supervised target detection, the union dictionary consisting of the background dictionary and target dictionary is usually given *a priori* [32], [33]. However, in the case of anomaly detection, neither the background dictionary nor the anomaly dictionary is known in advance. The clustering methods that group all the pixels into several classes are preferred to construct the background dictionary, such as  $k$ -means [20], [26] or  $k$ -means++ [34], mean shift [35], and density-based spatial clustering of applications with noise (DBSCAN) [36]. However, these clustering approaches have a common drawback that choosing appropriate parameters can be nontrivial. To avoid this, we adopt the density peak-based clustering and further modify it properly to fit the cluster analysis of hyperspectral data. It is worth mentioning that this clustering algorithm has been successfully utilized in several hyperspectral applications, such as band selection [37], [38] and classification with noisy label [39].

In this article, we use a soft Gaussian kernel function to compute the local density  $\rho_i$  instead of the hard threshold function in (2)

$$\rho_i = \sum_{j=1, j \neq i}^n \exp\left(-\left(\frac{d_{i,j}}{d_c}\right)^2\right). \quad (8)$$

The adoption of Gaussian kernel function can reduce possible statistical errors [37], [39].  $d_c$  is empirically set to the

value as follows:

$$d_c = \mathbf{d}_s(p), p = \langle 0.02 \times n \times (n-1) \rangle \quad (9)$$

where  $\mathbf{d}_s$  is the sorted result of all  $d_{i,j}$  ( $i \neq j$ ) in an ascending order and  $\langle \cdot \rangle$  stands for the round operation. Then,  $\rho_i$  and  $\delta_i$  are normalized to  $[0, 1]$  according to the following rule:

$$\rho_i = \frac{\rho_i - \min(\rho)}{\max(\rho) - \min(\rho)}, \quad \delta_i = \frac{\delta_i - \min(\delta)}{\max(\delta) - \min(\delta)}. \quad (10)$$

In the task of dictionary-based anomaly detection, the background dictionary should cover all the background classes. However, the ranking score in (5) can recognize the dominated background classes effectively while omit some sparse background classes with less pixels since  $\rho_i$  of centers in these classes are relatively low. Hence, we give more importance to the role of  $\delta_i$ , and the ranking score is modified as

$$\gamma_i = \rho_i \times \delta_i^2. \quad (11)$$

Some anomalous pixels may also have a large  $\delta_i$ , and their local densities  $\rho_i$  are far less than those of centers in background classes so that these anomalous pixels will not be selected. To make the process of identifying the centers with anomalously large ranking scores automatic, we empirically define the number of centers as the first  $k$  satisfies

$$|\lg(\gamma_{k+1}) - \lg(\gamma_{k+2})| < \eta \quad \text{and} \quad |\lg(\gamma_{k+2}) - \lg(\gamma_{k+3})| < \eta \quad (12)$$

where  $\eta$  is a predefined threshold. After the central pixel for each background cluster has been identified, each remaining pixel is assigned to the same cluster as its nearest neighbor of higher density. Finally, the clusters with few pixels (less than 1% of the total number of pixels) are merged into the nearest class based on the distance between the cluster centers. In this way, all the pixels including both background and the anomalies are grouped into several clusters.

For the background dictionary  $\mathbf{B}$ , the central pixel in each cluster can be selected as the atom to form the background dictionary. However, the center only is often insufficient to

represent a background material due to the fact that the spectrum is often affected by environmental conditions. In order to boost the representation ability of the background dictionary, the other  $P - 1$  pixels which give the highest ranking scores in each cluster measured by (11s) as well as the central pixel are chosen to constitute the background dictionary. It is worth noting that the ranking score in (11) takes both  $\rho_i$  and  $\delta_i$  into account; thus, in each cluster, the other selected atoms are at a certain distance to the center pixel, leading to more diversity of the spectrum. On the other hand, some of the background clusters may contain the anomalies, and the value of  $P$  cannot be too large to avoid the anomalies being mistakenly selected as the atoms in the background dictionary.

For the potential anomaly dictionary  $\mathbf{A}$ , the anomalous pixels with significant difference from the background can be considered isolated points. These pixels have a very small  $\rho_i$  and a large  $\delta_i$  and thus can be easily picked from the decision graph. Specifically, another score for each point is computed by

$$\phi_i = \delta_i / \rho_i. \quad (13)$$

We sort the scores in a descending order and choose the first  $r$  pixels as the atoms in the potential anomaly dictionary.

### B. TV and Sparsity Regularized Decomposition Model

Considering that the natural background usually has spatial smoothness, it is very likely that two spatial neighboring pixels  $\mathbf{y}_i$  and  $\mathbf{y}_j$  have similar representation vectors  $\mathbf{x}_i$  and  $\mathbf{x}_j$  associated with the background dictionary  $\mathbf{B}$ . The TV regularizer [40], [41], which offers the potential of constraining the smoothness of neighboring pixels [27], [42], [43], is incorporated to improve the estimation of background part

$$TV(\mathbf{X}) = \sum_{\{i,j\} \in \varepsilon} \|\mathbf{x}_i - \mathbf{x}_j\|_1 \quad (14)$$

where  $\varepsilon$  denotes the set of the pairs of each pixel and its spatial neighbors (in both horizontal and vertical dimensions). Let  $\mathbf{H}_h$  denote the linear operator calculating the differences between the vectors in  $\mathbf{X}$  corresponding to adjacent pixels in horizontal direction, i.e.,  $\mathbf{H}_h \mathbf{X} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N]$ , in which  $\mathbf{d}_i = \mathbf{x}_i - \mathbf{x}_{i_h}$  reflects the difference between the  $i$ th pixel and its horizontal neighbor. Likewise, the vertical difference can be computed through  $\mathbf{H}_v \mathbf{X} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N]$ . Thus, the TV term defined in (14) can be rewritten in an equivalent way as follows:

$$TV(\mathbf{X}) = \left\| \begin{bmatrix} \mathbf{H}_h \mathbf{X} \\ \mathbf{H}_v \mathbf{X} \end{bmatrix} \right\|_{1,1} = \|\mathbf{H} \mathbf{X}\|_{1,1} \quad (15)$$

where the  $\ell_{1,1}$ -norm is defined as  $\|\mathbf{X}\|_{1,1} = \sum_{i=1}^n \|\mathbf{x}_i\|_1$ .

On the other hand, we expect that the atoms in potential anomaly dictionary  $\mathbf{A}$  are active only when representing an anomalous pixel and the representation vectors of the background pixels associated with  $\mathbf{A}$  tend to be zero. Since the anomalous pixels are rare in the scene, the representation coefficients  $\mathbf{Z}$  should be column sparse. To achieve this, the  $\ell_{2,1}$ -norm [18], which is defined as the sum of  $\ell_2$ -norm of

the column in a matrix, is utilized to constrain  $\mathbf{Z}$ . Compared with  $\ell_1$ -norm, the  $\ell_{2,1}$ -norm is superior in encouraging the column sparsity, i.e., most of the columns of  $\mathbf{Z}$  being zero except for those corresponding to the anomalies.

Finally, the data fidelity term  $\|\mathbf{Y} - \mathbf{BX} - \mathbf{AZ}\|_F^2$  is used to model the noise, which is usually assumed identically and independently distributed Gaussian random variables [25]. With these terms in place, the optimization problem of the proposed method can be formulated as

$$\min_{\mathbf{X}, \mathbf{Z}} \|\mathbf{Y} - \mathbf{BX} - \mathbf{AZ}\|_F^2 + \lambda \|\mathbf{H} \mathbf{X}\|_{1,1} + \beta \|\mathbf{Z}\|_{2,1} \quad (16)$$

where  $\lambda$  and  $\beta$  are the tradeoff parameters used to balance these terms. Once the optimization for problem (16) is finished, the anomalous degree of each test pixel can be determined by the response of the anomalous part  $\mathbf{S}$  as follows:

$$\left\| [\mathbf{S}^*]_{:,i} \right\|_2 = \left\| [\mathbf{AZ}^*]_{:,i} \right\|_2 = \sqrt{\sum_{j=1}^d \left( [\mathbf{AZ}^*]_{j,i} \right)^2}. \quad (17)$$

### C. Optimization Procedure

To solve the proposed optimization problem in (16), three auxiliary variables  $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$  are introduced to make the objective function separable. Thus, the original problem can be converted to the following equivalent form:

$$\begin{aligned} & \min_{\{\mathbf{V}\}, \mathbf{X}, \mathbf{Z}} \|\mathbf{Y} - \mathbf{BX} - \mathbf{AZ}\|_F^2 + \lambda \|\mathbf{V}_2\|_{1,1} + \beta \|\mathbf{V}_3\|_{2,1} \\ & \text{s.t. } \mathbf{V}_1 = \mathbf{X}, \mathbf{V}_2 = \mathbf{HV}_1, \mathbf{V}_3 = \mathbf{Z}. \end{aligned} \quad (18)$$

The problem (18) can be solved by utilizing ADMM. The augmented Lagrangian function is formulated as

$$\begin{aligned} \mathcal{L}(\{\mathbf{V}\}, \mathbf{X}, \mathbf{Z}, \{\mathbf{D}\}, \mu) &= \|\mathbf{Y} - \mathbf{BX} - \mathbf{AZ}\|_F^2 + \lambda \|\mathbf{V}_2\|_{1,1} + \beta \|\mathbf{V}_3\|_{2,1} \\ &+ \frac{\mu}{2} (\|\mathbf{V}_1 - \mathbf{X} - \mathbf{D}_1\|_F^2 + \|\mathbf{V}_2 - \mathbf{HV}_1 - \mathbf{D}_2\|_F^2 \\ &+ \|\mathbf{V}_3 - \mathbf{Z} - \mathbf{D}_3\|_F^2) \end{aligned} \quad (19)$$

where  $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$  are the Lagrange multipliers and  $\mu$  is a positive penalty parameter. Note that the objective function  $\mathcal{L}$  is convex regarding each variable with other variables fixed despite the fact that  $\mathcal{L}$  is not convex with respect to all the variables simultaneously. Therefore, the optimization can be achieved by successively minimizing  $\mathcal{L}$  with respect to one variable while fixing the others. The detailed procedure is presented as follows:

- 1) The optimization problem for  $\mathbf{X}$  is

$$\begin{aligned} \mathbf{X}^{(t+1)} &= \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{BX} - \mathbf{AZ}^{(t)}\|_F^2 \\ &+ \frac{\mu^{(t)}}{2} \left\| \mathbf{V}_1^{(t)} - \mathbf{X} - \mathbf{D}_1^{(t)} \right\|_F^2 \end{aligned} \quad (20)$$

where  $t$  is the current index of the iteration. This problem has a closed-form solution as follows:

$$\begin{aligned} \mathbf{X}^{(t+1)} &= (2\mathbf{B}^T \mathbf{B} + \mu \mathbf{I})^{-1} \\ &\times \left[ 2\mathbf{B}^T (\mathbf{Y} - \mathbf{AZ}^{(t)}) + \mu^{(t)} (\mathbf{V}_1^{(t)} - \mathbf{D}_1^{(t)}) \right]. \end{aligned} \quad (21)$$

- 2) The optimization problem for  $\mathbf{Z}$  can be written as follows:

$$\begin{aligned}\mathbf{Z}^{(t+1)} = \arg \min_{\mathbf{Z}} & \| \mathbf{Y} - \mathbf{B}\mathbf{X}^{(t+1)} - \mathbf{A}\mathbf{Z} \|_F^2 \\ & + \frac{\mu^{(t)}}{2} \left\| \mathbf{V}_3^{(t)} - \mathbf{Z} - \mathbf{D}_3^{(t)} \right\|_F^2.\end{aligned}\quad (22)$$

Likewise, the problem has a closed-form solution as

$$\begin{aligned}\mathbf{Z}^{(t+1)} = & (2\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})^{-1} \\ & \times \left[ 2\mathbf{A}^T (\mathbf{Y} - \mathbf{B}\mathbf{X}^{(t+1)}) + \mu^{(t)} (\mathbf{V}_3^{(t)} - \mathbf{D}_3^{(t)}) \right].\end{aligned}\quad (23)$$

- 3) The optimization problem for  $\mathbf{V}_1$  can be written as

$$\begin{aligned}\mathbf{V}_1^{(t+1)} = \arg \min_{\mathbf{V}_1} & \left\| \mathbf{V}_1 - \mathbf{X}^{(t+1)} - \mathbf{D}_1^{(t)} \right\|_F^2 \\ & + \left\| \mathbf{V}_2^{(t)} - \mathbf{H}\mathbf{V}_1 - \mathbf{D}_2^{(t)} \right\|_F^2\end{aligned}\quad (24)$$

which has the following analytical form:

$$\mathbf{V}_1^{(t+1)} = (\mathbf{H}^T \mathbf{H} + \mathbf{I})^{-1} \left( \mathbf{H}^T (\mathbf{V}_2^{(t)} - \mathbf{D}_2^{(t)}) + \mathbf{X}^{(t+1)} + \mathbf{D}_1^{(t)} \right).\quad (25)$$

Here, we resort to the discrete Fourier transform diagonalization to efficiently implement the update rule in (25).

- 4) The variable  $\mathbf{V}_2$  can be updated by solving the following problem:

$$\mathbf{V}_2^{(t+1)} = \arg \min_{\mathbf{V}_2} \lambda \|\mathbf{V}_2\|_1 + \frac{\mu^{(t)}}{2} \left\| \mathbf{V}_2 - \mathbf{H}\mathbf{V}_1^{(t+1)} - \mathbf{D}_2^{(t)} \right\|_F^2\quad (26)$$

whose solution is given by

$$\mathbf{V}_2^{(t+1)} = \text{soft} \left( \mathbf{H}\mathbf{V}_1^{(t+1)} + \mathbf{D}_2^{(t)}, \frac{\lambda}{\mu^{(t)}} \right)\quad (27)$$

in which  $\text{soft}(\cdot, \cdot)$  is the shrinkage operator [44].

- 5) The optimization problem for  $\mathbf{V}_3$  to be solved is

$$\mathbf{V}_3^{(t+1)} = \arg \min_{\mathbf{V}_3} \beta \|\mathbf{V}_3\|_{2,1} + \frac{\mu^{(t)}}{2} \left\| \mathbf{V}_3 - \mathbf{Z}^{(t+1)} - \mathbf{D}_3^{(t)} \right\|_F^2\quad (28)$$

whose solution can be obtained as follows:

$$\mathbf{V}_3^{(t+1)} = \Omega_{\beta/\mu^{(t)}} (\mathbf{Z}^{(t+1)} + \mathbf{D}_3^{(t)})\quad (29)$$

where  $\Omega$  is the solving operator for  $\ell_{2,1}$ -minimization problem [18]. Note that the order of updating rules for the above five variables can be exchanged. Finally, before advancing into the next iteration, the Lagrange multipliers  $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$  and the penalty parameter  $\mu$  are updated by the following equations:

$$\begin{aligned}\mathbf{D}_1^{(t+1)} &= \mathbf{D}_1^{(t)} - (\mathbf{V}_1^{(t+1)} - \mathbf{X}^{(t+1)}) \\ \mathbf{D}_2^{(t+1)} &= \mathbf{D}_2^{(t)} - (\mathbf{V}_2^{(t+1)} - \mathbf{H}\mathbf{V}_1^{(t+1)}) \\ \mathbf{D}_3^{(t+1)} &= \mathbf{D}_3^{(t)} - (\mathbf{V}_3^{(t+1)} - \mathbf{Z}^{(t+1)})\end{aligned}\quad (30)$$

$$\mu^{(t+1)} = \min(\rho \mu^{(t)}, \mu_{\max}).\quad (31)$$

According to the above derivation and description, the optimization procedure for the proposed TVSDM can be summarized in **Algorithm 1**.

---

**Algorithm 1** Optimization Procedure for TVSDM

---

**Input:** data matrix  $\mathbf{Y} \in \mathbb{R}^{d \times n}$ , background dictionary  $\mathbf{B}$ , potential anomaly dictionary  $\mathbf{A}$ , tradeoff parameters  $\lambda$  and  $\beta$ .

**Initialize:** all variables  $\mathbf{X}, \mathbf{Z}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$  are set to zero matrices,  $\mu_0 = 1e-3$ ,  $\mu_{\max} = 1e10$ ,  $\rho = 1.2$ ,  $\varepsilon_1 = 1e-4$ ,  $t = 0$ .

**While** not converged, **do**

1) Update  $\mathbf{X}$  and  $\mathbf{Z}$  with (21) and (23), respectively.

2) Update  $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$  with (25), (27) and (29), respectively.

3) Update  $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$  and  $\mu$  with (30) and (31), respectively.

4) Check the convergence condition

$$\|\mathbf{V}_1 - \mathbf{X}\|_F + \|\mathbf{V}_2 - \mathbf{H}\mathbf{V}_1\|_F + \|\mathbf{V}_3 - \mathbf{Z}\|_F < \varepsilon_1$$

5)  $t \leftarrow t + 1$ .

**end while**

**Output:** obtain detection map with (17).

---

#### IV. EXPERIMENTAL RESULTS

In this section, both simulated and real hyperspectral data sets are used to evaluate the proposed anomaly detection method. Especially, two simulated data sets are generated to give a comprehensive analysis on the performance of TVSDM and three real data sets are employed to validate the effectiveness of TVSDM in practical situations. The receiver operating characteristic (ROC) curve [45] which describes the tradeoff between the probability of detection and false alarm rate along with the area under ROC curve (AUC) [46] is employed to assess the detection performance quantitatively. The experimental platform is a computer with an octa-core Intel CPU Xeon CPU 3.07 and 3.06 GHz processors, 48 GB memory, and 64-bit Windows 7. All the experiments are carried out in MATLAB R2013b.

##### A. Experiments on Simulated Hyperspectral Data Sets

The two simulated hyperspectral data sets are generated based on a real scene which was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) from San Diego airport region, CA, USA. The raw hyperspectral imagery consists of 224 spectral bands in wavelengths ranging from 370 to 2510 nm and the spatial resolution is 3.5 m/pixel. In our experiments, 186 bands are retained after removal of the low-signal-to-noise ratio (SNR) bands and those affected by water vapor absorption (1–6, 33–35, 94–97, 107–113, 153–166, and 221–224). The original scene shown in Fig. 3(a) contains  $400 \times 400$  pixels and two portions of  $100 \times 100$  pixels (in red squares) with different background compositions and distributions are selected to form the simulated data sets. Here, we adopt the target implantation method in [47] to generate the anomalous pixels. Specifically, a synthetic anomalous pixel with spectrum  $\mathbf{z}$  and specific spatial location is generated by fractionally embedding a desired anomaly with spectrum  $\mathbf{t}$  into

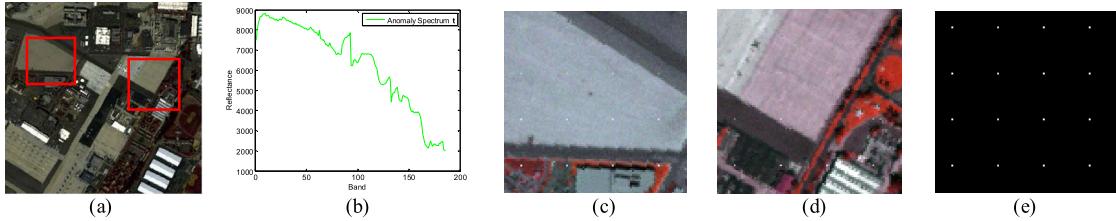


Fig. 3. Simulated hyperspectral data sets. (a) False color image of the whole scene. (b) Spectrum of the desired anomaly  $\mathbf{t}$ . (c) False color image of the first simulated data set. (d) False color image of the second simulated imagery. (e) Ground-truth map of the embedded anomalies.

the background pixel at the same place

$$\mathbf{z} = f \cdot \mathbf{t} + (1 - f) \cdot \mathbf{b}. \quad (32)$$

In each simulated data set, 16 anomalous targets have been implanted, which are distributed in four rows and four columns. The mixing fractions  $f$  corresponding to the anomalies in the same row remain unchanged and take value of 0.05, 0.1, 0.2, and 0.4 for different rows from top to bottom, respectively. The desired anomaly is chosen from one plane in the original scene, the spectrum  $\mathbf{t}$  of which is shown in Fig. 3(b). Furthermore, to evaluate the detection performance in the presence of noise, zero-mean white Gaussian noise is added to the simulated data with SNR = 30 dB. Here, the definition of SNR is given as follows:

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

where  $\mathbb{E}[\cdot]$  stands for the expectation operator;  $\mathbf{y}$  and  $\mathbf{e}$  represent the spectrum and the additive noise of a pixel, respectively. The false color images of the two simulated data sets and the ground-truth map of the embedded anomalies are provided in Fig. 3(c)–(e), respectively.

It should be pointed out that these two simulated data sets are challenging for anomaly detection in two aspects: 1) all the embedded anomalous targets are at a relatively low mixing fraction  $f$ , resulting in a weak anomaly detection problem, which is difficult for most of detectors and 2) there are some sparse background materials in the scene, which may cause some false alarms easily in the detection map.

In the following, we utilize the first simulated data set to conduct a comprehensive analysis on the proposed TVSDM, including the dictionary construction, the selection of the involved parameters, and the performance gain of TVSDM by step-wise adding the regularization terms. Then, the proposed TVSDM is compared with other anomaly detectors using both the two simulated data sets in order to achieve a more reliable comparison.

1) *Dictionary Construction*: Here, we first analyze the cluster centers found by the density peak-based clustering with the improved ranking score in (11). The threshold  $\eta$  in (12) is empirically set to 0.1 and fixed in the following. The decision graph and the ranking process are shown in Fig. 4(a) and (b), respectively. Clearly, in addition to the centers found in Fig. 1(a) that correspond to the dominated background classes, more representative points can be picked out with the improved ranking score. Note that Fig. 4(a) is actually the result of zooming in the leftmost region in Fig. 1(a).

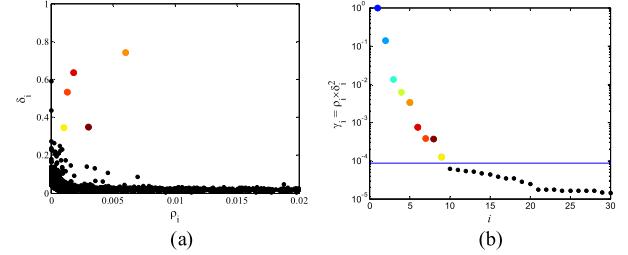


Fig. 4. Cluster analysis on the first simulated data set. (a) Decision graph, where the cluster centers are highlighted with different colors. (b) Ranking score  $y_i = \rho_i \times \delta_i^2$  sorted in descending order.

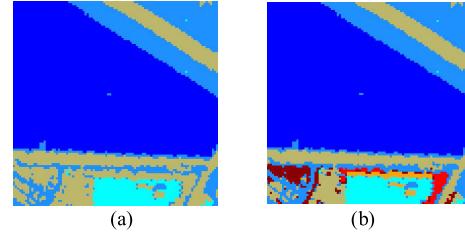


Fig. 5. Cluster results of the first simulated data set with different ranking scores, in which pixels are colored according to the cluster to which they are assigned. (a) Segmentation map obtained with  $y_i = \rho_i \times \delta_i$ . (b) Segmentation map obtained with  $y_i = \rho_i \times \delta_i^2$ .

After the cluster centers have been found, the assignment of other pixels can then be performed. The segmentation maps obtained by different ranking scores are shown in Fig. 5(a) and (b), respectively. The clusters are distinguished with different colors, which are in accordance with those of the representative points in Figs. 2 or 4. It is noteworthy that there are nine representative points or cluster centers found in Fig. 4(b), while seven clusters are finally obtained in Fig. 5(b) because two other clusters with few pixels are merged into the nearest cluster. An important observation is that the clustering result in Fig. 5(b) covers some sparse background classes, i.e., those colored with orange, red, and dark red, respectively, compared with that in Fig. 5(a).

Fig. 6(a) shows the spectra of the selected background pixels ( $P = 20$ ) based on the clustering result in Fig. 5(b). Note here that for a better distinction, we display the intrinsic spectra before adding the noise. It can be observed that the sparse background classes have distinct spectra from those of dominated background. The coverage of these spectra in background dictionary can greatly enhance its representation

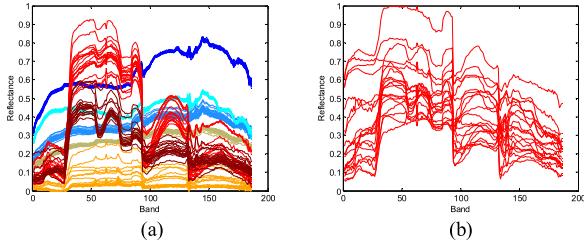


Fig. 6. Spectra of the selected pixels in (a) background dictionary and (b) potential anomaly dictionary.

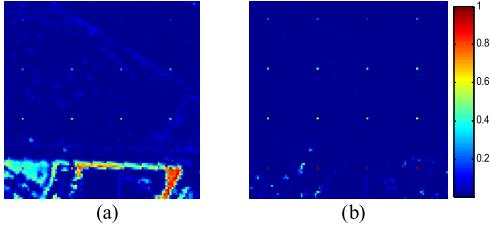


Fig. 7. 2-D plots of detection results obtained by the proposed detector using background dictionary consists of (a) pixels of only dominated background classes and (b) pixels of dominated as well as sparse background classes.

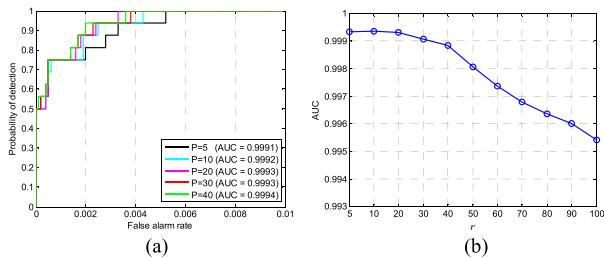


Fig. 8. Analyses about the effects of parameters  $P$  and  $r$  on the detection performance. (a) Variation of ROC curves under different values of  $P$ . (b) Variation of AUC scores under different values of  $r$ .

ability. Meanwhile, the spectra of the potential anomalous pixels ( $r = 20$ ) are also plotted in Fig. 6(b), where some spectra are with significant difference from those in Fig. 6(a).

Furthermore, the detection results obtained by TVSDM using dictionaries based on different clustering results are displayed in Fig. 7(a) and (b), respectively. The tradeoff parameters in the proposed TVSDM are set to  $\lambda = 0.1$  and  $\beta = 1$ . The selection of parameters will be analyzed in Section IV-A2. It can be observed from Fig. 7(a) that there are many false alarms in the bottom of the scene. However, these false alarms can be well suppressed in Fig. 7(b) owing to the fact that the background dictionary in this case covers all classes of background atoms.

2) *Selection of Parameters*: First, we investigate the effects of parameters  $P$  and  $r$  on the detection performance of the proposed TVSDM. From Fig. 8(a), it can be concluded that the detection results in terms of ROC curves obtained with different  $P$  are almost the same. This indicates that our proposed TVSDM is robust to  $P$ . Thus, we fix  $P = 20$  throughout the experiments. Fig. 8(b) shows the variation of detection performance with regard to AUC score under different values of  $r$ . The proposed TVSDM obtains satisfactory result with

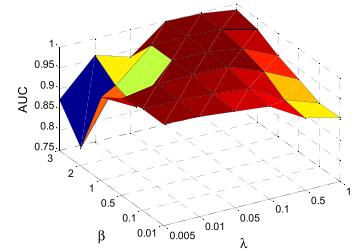


Fig. 9. Performance variation of the proposed detector under different combinations of tradeoff parameters  $\beta$  and  $\lambda$ .

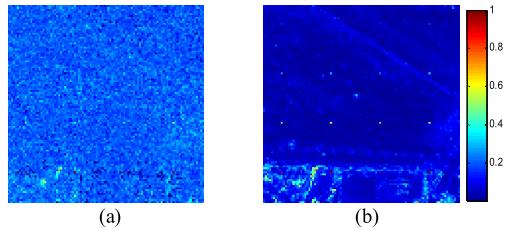


Fig. 10. 2-D plots of detection results obtained by (a) proposed TVSDM without any regularization terms ( $\lambda = 0, \beta = 0$ ) and (b) proposed TVSDM without the sparsity regularization term ( $\lambda = 0.1, \beta = 0$ ).

few potential anomalous atoms. This can be explained by the fact that the strong anomalous pixels are far away from the background clusters in the feature space and, therefore, these pixels generally exist in the first several pixels with highest ranking scores  $\phi_i$ . As  $r$  increases, the performance drops slightly due to the potential anomaly dictionary contains more background pixels. As a consequence, the value of  $r$  is set to 20 in our experiments.

Next, we illustrate the performance of the proposed detector when jointly considering the two tradeoff parameters  $\beta$  and  $\lambda$  in Fig. 9. It reveals that, when both  $\beta$  and  $\lambda$  take moderate values, the performance is satisfactory and the change of AUC score is not obvious. Thus, for simplicity, we empirically set  $\beta = 1$  and  $\lambda = 0.1$  in all the following experiments since the parameter tuning is generally a nontrivial and computationally expensive task.

3) *Effects of Regularization Terms*: In this section, we investigate the individual contributions of the incorporated regularizations, i.e., TV term and  $\ell_{2,1}$ -norm. The proposed TVSDM without any regularization terms ( $\lambda = 0$  and  $\beta = 0$ ) and that without the sparsity regularization term ( $\lambda = 0.1$  and  $\beta = 0$ ) are performed on the first simulated data set. The 2-D plots of detection results are shown in Fig. 10(a) and (b), respectively. It can be observed that the TV term is good at removing the noise and keeping the edges between the background and the anomalies. However, there are still some false alarms caused by the pixels in the bottom of the scene. The reason is that the potential anomaly dictionary also contains some background atoms. Therefore, the  $\ell_{2,1}$ -norm is incorporated with the aim of forcing the pixels that are not anomalous to be represented only using the atoms in the background dictionary. Thus, the false alarms can be suppressed to a large extent, which can be seen from Fig. 7(b).

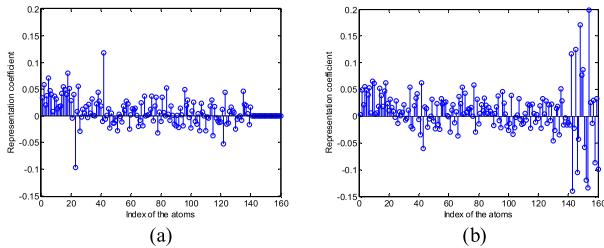


Fig. 11. Representation coefficients of a background pixel obtained by (a) proposed TVSDM ( $\lambda = 0.1$ ,  $\beta = 1$ ) and (b) proposed TVSDM without the sparsity regularization term ( $\lambda = 0.1$ ,  $\beta = 0$ ).

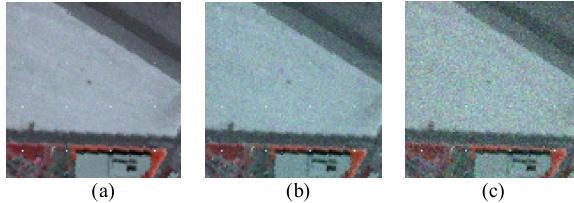


Fig. 12. False color images of the first simulated data set with different levels of zero-mean white Gaussian noise. (a) 30, (b) 25, and (c) 20 dB.

Another observation from Fig. 10(b) is that the weak anomalies in the first row are not so obvious since there is still some noise interference. To illustrate this, we plot the representation coefficients of a background pixel obtained by the proposed TVSDM with and without the sparsity regularization term, as shown in Fig. 11(a) and (b), respectively. The first 140 atoms in the union dictionary are background ones and the last 20 are potential anomalous ones. From Fig. 11(b), it is evident that the proposed TVSDM without the sparsity constraint tends to use significant weights for the potential anomalous atoms to represent the noise part so as to make the reconstruction error [the first term in (16)] as small as possible. However, this problem has been well solved in Fig. 11(a), where the  $\ell_{2,1}$ -norm constrains those weights to be zero since the noise cannot be represented by the atoms in the potential anomaly dictionary. As a result, the weak anomalies stand out with the noise being further suppressed.

**4) Comparison With Different Anomaly Detectors Using the First Simulated Data Set:** We now compare the detection performance of the proposed TVSDM with two traditional anomaly detectors, i.e., LRX [7], local KRX (LKRX) [8], and three state-of-the-art anomaly detectors including CRD [15], LRASR [20], and PAB-DC [26] using the first simulated data set. The CRD is a representation-based method. The LRASR is a background dictionary-based anomaly detector, in which both the low rankness and sparsity are imposed to characterize the background part. The PAB-DC is also a union dictionary-based detector which utilizes the low rankness and  $\ell_1$ -norm to characterize the background and anomalies, respectively. Besides, in order to investigate the robustness of different anomaly detectors to noise, we corrupt the first simulated data set with different levels of zero-mean Gaussian noise (30, 25, and 20 dB, respectively). Fig. 12(a)–(c) shows the corresponding false color images.

TABLE I  
AUC SCORES OF THE DETECTION RESULTS REPORTED IN FIG. 14

Algorithm	LRX	LKRX	CRD	LRASR	PAB-DC	TVSDM
30 dB	0.9895	0.9900	0.9976	0.9180	0.9905	<b>0.9993</b>
25 dB	0.9862	0.9890	0.9839	0.8960	0.9847	<b>0.9987</b>
20 dB	0.9389	0.9580	0.9614	0.8182	0.9540	<b>0.9969</b>

The parameters involved in these approaches are optimized or set to the recommended values in the literatures. Specifically, the window sizes ( $w_{\text{out}}$ ,  $w_{\text{in}}$ ) in LRX and CRD are both set to (7, 3) after extensive searching. For LKRX, the window sizes are consistent with that in LRX and the pixels falling into the outer window region are employed to calculate the Gram matrix  $\mathbf{K}_b$  [8]. In our experiments, the Gaussian RBF kernel  $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/c)$  is utilized and the width  $c$  is set to  $2^3$  after tuning. The tradeoff parameters  $\beta$  and  $\lambda$  in LRASR are both set to 0.1, which is a default setting suggested in [20]. For PAB-DC, the parameters  $\beta$  and  $\lambda$  are set to 1 and 2, respectively, to achieve optimal performance. The dictionaries in LRASR and PAB-DC are chosen the same as that constructed in our proposed method for a fair comparison.

The 2-D plots of the detection results are presented in Fig. 13(a)–(c). From these figures, it is obvious that all detectors can discriminate between the background and the anomalous pixels with relatively high mixing fraction ( $f = 0.4$ ). However, these compared detectors are unable to distinguish well the weak anomalies in the first row. Compared with LRX, LKRX can detect more anomalies, which may be owing to that the anomalies and background are better separated in the feature space. The anomalies in the detection maps of PAB-DC and TVSDM are better separated from noise than that of LRASR. Overall, the proposed TVSDM provides a clear discrimination between background and all the embedded anomalous pixels. Compared with PAB-DC, the detection results of TVSDM are superior to those of PAB-DC, especially when the noise level is high, i.e., SNR = 20 dB. This should indicate that the TV term is more powerful than low rankness in suppressing the noise. The ROC curves in Fig. 14(a)–(c) show that the proposed detector can achieve 100% detection rate with a very low false alarm rate. It is noteworthy that LRX and CRD are effective in suppressing the background. Hence, they can achieve comparable detection rates with TVSDM when the false alarm rate is low. Nevertheless, LRX and CRD are inferior to the proposed TVSDM since the weak anomalies are missed in the detection maps. This fact is also reflected in the ROC curves, where the detection rates of LRX and CRD achieve 100% with high false alarm rates. The AUC scores which provide overall assessment of the detection performance are listed in Table I. The proposed detector achieves the highest scores as expected. Particularly when SNR = 20 dB, the AUC score of TVSDM is 0.035 higher than the second best, i.e., CRD, which is a significant gap. These results demonstrate that the proposed TVSDM outperforms the other compared detectors.

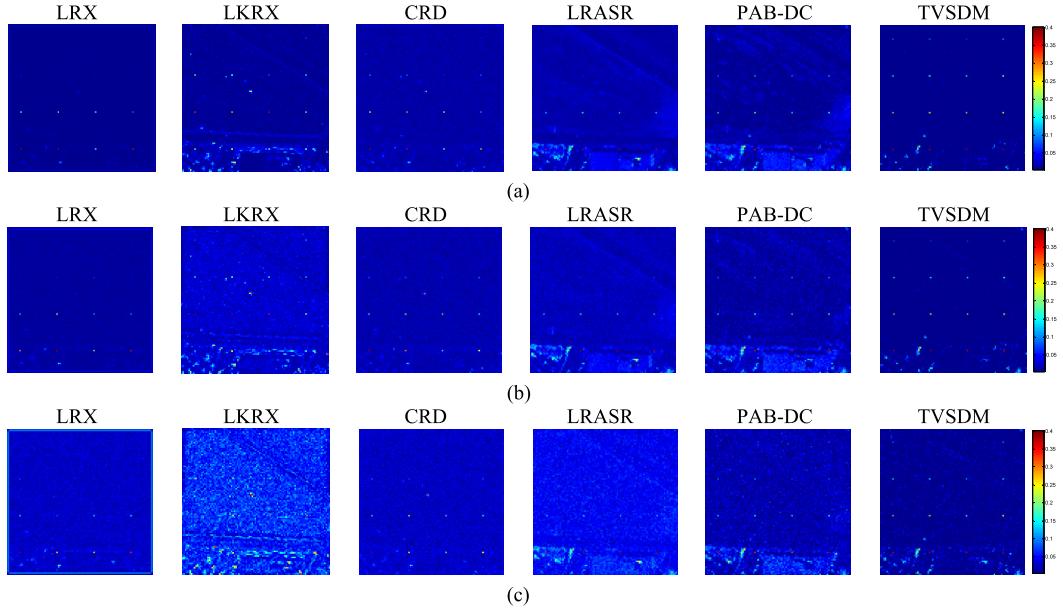


Fig. 13. 2-D plots of detection results of different anomaly detectors on the first simulated data set with different levels of noise. (a) 30, (b) 25, and (c) 20 dB.

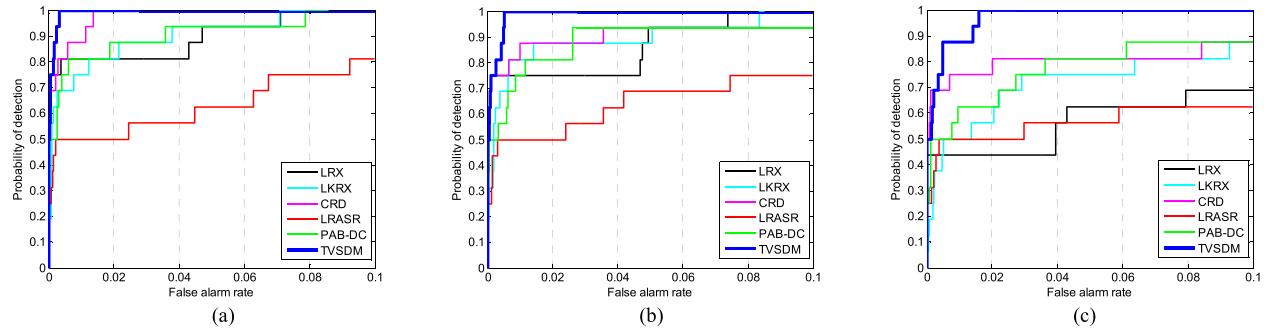


Fig. 14. ROC curves of different anomaly detectors on the first simulated data set with different levels of noise. (a) 30, (b) 25, and (c) 20 dB.

TABLE II  
AUC SCORES AND COMPUTATIONAL TIME (MEAN  $\pm$  STD) FOR DIFFERENT ANOMALY DETECTORS BASED ON 20 RUNS

Algorithm	LRX	LKRX	CRD	LRASR	PAB-DC	TVSDM
AUC	$0.9805 \pm 0.0129$	$0.9853 \pm 0.0053$	$0.9836 \pm 0.0196$	$0.8777 \pm 0.0269$	$0.9876 \pm 0.0025$	$0.9988 \pm 0.0004$
Time (s)	$66.62 \pm 9.25$	$41.75 \pm 1.16$	$76.28 \pm 12.10$	$45.92 \pm 4.93$	$60.73 \pm 6.47$	$81.44 \pm 6.67$

Finally, we randomly change the positions of the implanted anomalies and perform the preceding experiment for 20 times to investigate the stability of different detectors. Here, the noise level is fixed as SNR = 30 dB. Table II provides the mean value and the standard deviation (std) of AUC scores. The proposed TVSDM is the best and most stable detector. We also report the computational costs in Table II and all the methods take a relatively long time.

5) *Comparison With Different Anomaly Detectors Using the Second Simulated Data Set:* Here, we use the second simulated data set to further validate the effectiveness of TVSDM in comparison with other anomaly detectors.

Likewise, the density peak-based clustering is first performed. The decision graph and the clustering result are shown in Fig. 15(a) and (b), respectively. In this simulated data set,

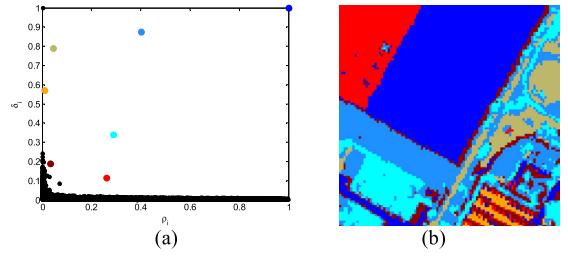


Fig. 15. Cluster analysis on the second simulated data set. (a) Decision graph. (b) Segmentation map.

seven clusters have been identified, also including the sparse background class in the bottom right corner of the scene, which is colored in orange in the clustering result.

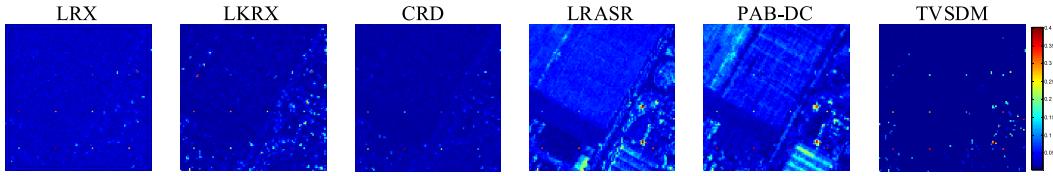


Fig. 16. 2-D plots of detection results of different anomaly detectors on the second simulated data set.

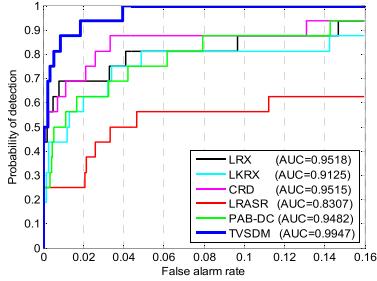


Fig. 17. ROC curves and AUC scores of different anomaly detectors on the second simulated data set.



Fig. 18. HYDICE data set. (a) False color image of the whole scene. (b) False color image of the detection area. (c) Ground-truth map of the anomalies.

Next, we evaluate the detection performance of TVSDM as well as the compared detectors. The window sizes ( $w_{\text{out}}, w_{\text{in}}$ ) in LRX and CRD are set to (9, 3) and (5, 3), respectively. The width  $c$  in LKRX is set to 2. For PAB-DC, the parameters  $\beta$  and  $\lambda$  are set to 2 and 1, respectively. The 2-D plots of the detection results are presented in Fig. 16. In the detection maps of LRX and CRD, the background is effectively suppressed. However, the noise still remains and some anomalies are missed. For LRASR and PAB-DC, the removal of background and noise is unsatisfactory, especially for the sparse background residing in the bottom right corner of the scene. By contrast, the background and noise are better removed in the detection map of TVSDM and the anomalies are more obvious. Quantitative results in terms of ROC curves and AUC scores shown in Fig. 17 also indicate that the proposed TVSDM achieves the best detection result.

#### B. Experiments on Real Hyperspectral Data Sets

In this section, we utilize three widely used real-world hyperspectral data sets, acquired from different data collection equipments, with various background materials and anomaly distributions to evaluate the practical performance of the proposed TVSDM.

The first one was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) airborne sensor over an urban residential area with a spatial resolution of 1 m. The raw imagery contains  $307 \times 307$  pixels and

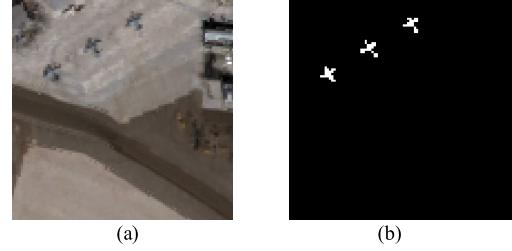


Fig. 19. San Diego data set. (a) False color image of the detection area. (b) Ground-truth map of the anomalies.

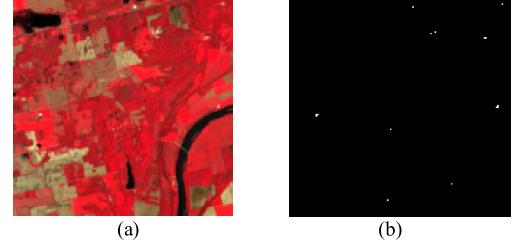


Fig. 20. Hyperion data set. (a) False color image. (b) Ground-truth map of the anomalies.

210 spectral bands, as shown in Fig. 18(a). In our experiments, 160 bands are preserved after removing the low-SNR and water vapor absorption bands (1–4, 76, 87, 101–112, 136–153, and 197–210). A subscene in the upper right area of the original image is cropped. The anomalous targets in this data set are the vehicles and roofs with different sizes. The image scene and the ground-truth map are illustrated in Fig. 18(b) and (c), respectively.

The second data set used for real-world detection is also a portion of the San Diego airport imagery. The upper left region with  $100 \times 100$  pixels is selected as the test imagery. This scene mainly contains roofs, shadows, and grass, and the anomalies to be detected are three airplanes. The false color image and the ground-truth map are depicted in Fig. 19(a) and (b), respectively.

The third one was downloaded from the EO-1 satellite image website, called the Hyperion data set [6]. This data set covers an agricultural region of the State of Indiana, USA with 242 bands and a spectral resolution of 10 m ranging from 357 to 2576 nm. After eliminating some low-SNR and uncalibrated bands, 149 bands are retained. A subimagery containing  $150 \times 150$  pixels with available ground truth of the anomalies is used for evaluation. In this data set, the anomalies are the storage silos and the roofs. The false color image and the ground-truth map are shown in Fig. 20(a) and (b), respectively.

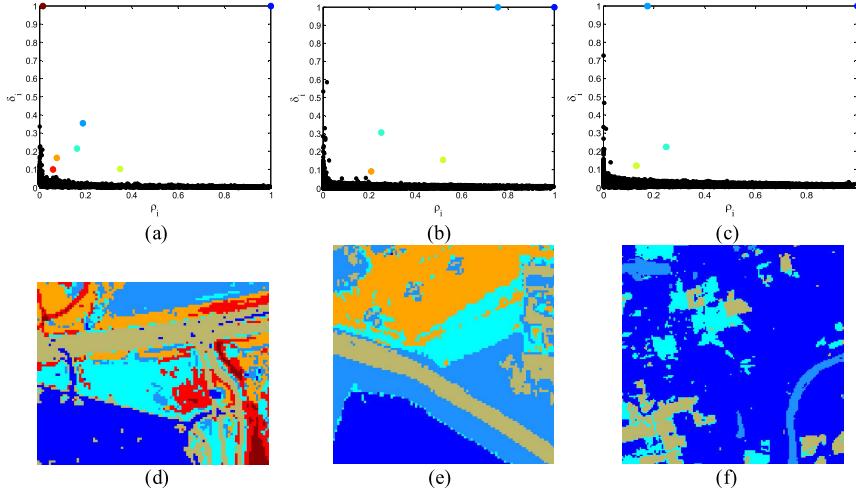


Fig. 21. Cluster analyses on three real data sets. The decision graphs of (a) HYDICE, (b) San Diego, and (c) Hyperion. The segmentation maps of (d) HYDICE, (e) San Diego, and (f) Hyperion.

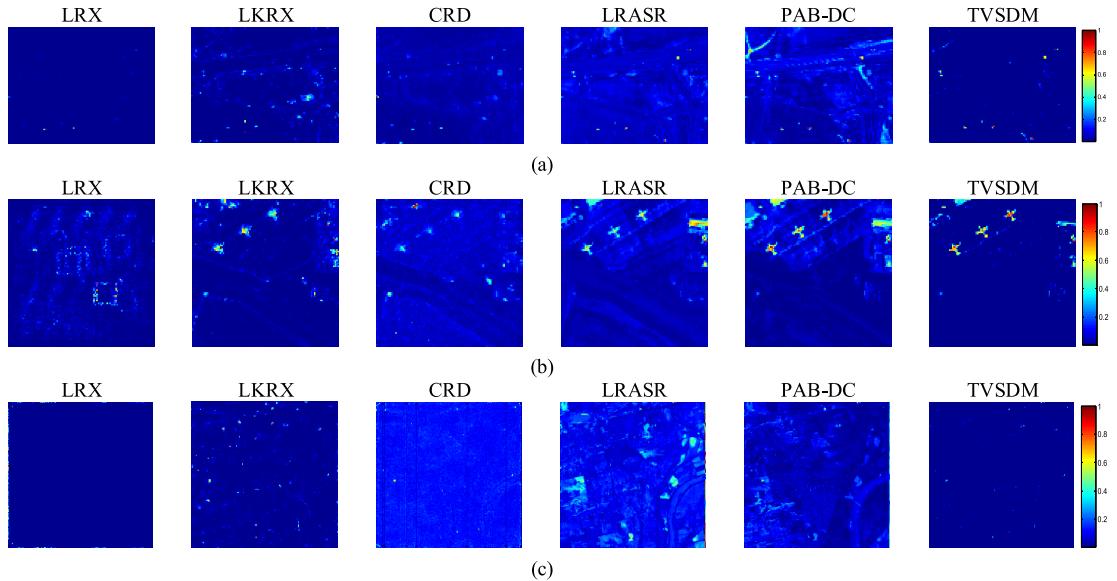


Fig. 22. 2-D plots of detection results of different anomaly detectors on three real data sets. (a) HYDICE. (b) San Diego. (c) Hyperion.

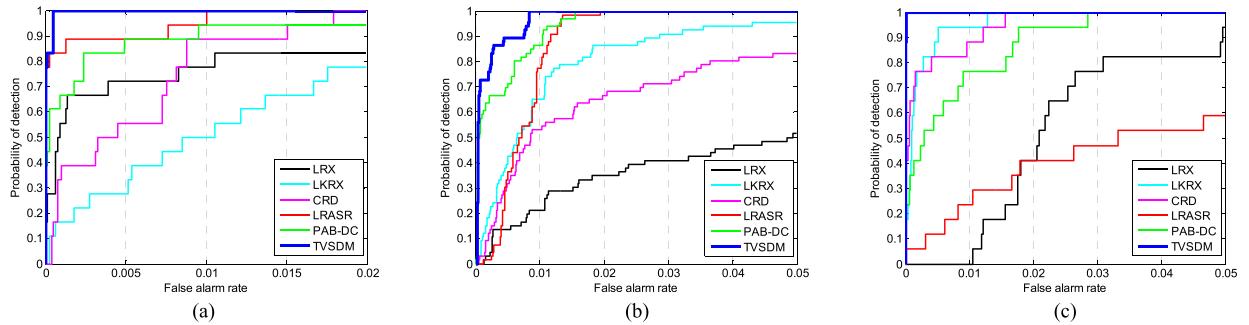


Fig. 23. ROC curves of different anomaly detectors on three real data sets. (a) HYDICE. (b) San Diego. (c) Hyperion.

For the HYDICE urban data, we first show the decision graph and the obtained cluster result in Fig. 21(a) and (d), respectively. Seven clusters are identified including six dominated background classes and one sparse background class (colored in dark red). The window sizes ( $w_{out}, w_{in}$ ) in LRX and

CRD are both set to (5, 3) and the width  $c$  in LKRX is set to 2 after extensive searching. The optimal setting for the tradeoff parameters in PAB-DC is  $\beta = 0.01$ ,  $\lambda = 2$ . Fig. 22(a) shows the 2-D plots of the detection results obtained by different detectors. The anomalies in the result of TVSDM are more

TABLE III  
AUC SCORES AND COMPUTATIONAL TIMES OF DIFFERENT ANOMALY DETECTORS ON THE THREE REAL DATA SETS

Algorithm		LRX	LKRX	CRD	LRASR	PAB-DC	TVSDM
HYDICE	AUC	0.9404	0.9828	0.9947	0.9989	0.9946	<b>0.9999</b>
	Time (s)	33.89	11.28	36.63	38.33	40.90	74.17
San Diego	AUC	0.8383	0.9886	0.9725	0.9925	0.9967	<b>0.9984</b>
	Time (s)	107.07	860.22	173.48	37.89	48.71	60.15
Hyperion	AUC	0.9737	0.9979	0.9972	0.9973	0.9978	<b>0.9999</b>
	Time (s)	220.72	3353.45	251.24	96.53	110.26	70.26

distinct since the background has been well suppressed. The quantitative results in terms of ROC curves and AUC scores are presented in Fig. 23(a) and Table III, respectively. These results also show that the proposed TVSDM outperforms the other detectors.

For the real San Diego data, the cluster analysis is performed first and the obtained decision graph as well as the assignment map are shown, respectively, in Fig. 21(b) and (e). For this data set, five dominated background classes have been found. Through extensive search, the window sizes ( $w_{\text{out}}, w_{\text{in}}$ ) in LRX and CRD are set to (17, 11) and (15, 7), respectively, and the width  $c$  in LKRX is set to 2. The tradeoff parameters in PAB-DC are set as  $\beta = 1$  and  $\lambda = 2$ . The 2-D plots of the detection results are displayed in Fig. 22(b). Clearly, the anomalies in the results of dictionary-based approaches are visually more evident than those of sliding windows-based approaches. This is because the anomalies in this image are relatively large and reside closely. Therefore, it is hard to choose appropriate sizes of the sliding windows. On the other hand, compared with LRASR and PAB-DC, the background is suppressed better in the proposed detector and thus the anomalies stand out. The ROC curves and AUC scores are also shown in Fig. 23(b) and Table III, respectively. The proposed TVSDM achieves 100% detection rate with the lowest false alarm rate (about 0.008) and the highest AUC score. It can be concluded that TVSDM has the best detection performance among all the compared detectors.

For the Hyperion data, the decision graph and the segmentation map are provided in Fig. 21(c) and (f), respectively, with four dominated background classes being identified. In this case, the window sizes ( $w_{\text{out}}, w_{\text{in}}$ ) in LRX and CRD are set to (17, 5) and (11, 3), respectively, and the width  $c$  in LKRX is set to 1 after repeated trials. The tradeoff parameters in PAB-DC  $\beta$  and  $\lambda$  are set to 1 and 2, respectively. Fig. 22(c) illustrates the 2-D plots of the detection results obtained by different methods. We can find that the proposed TVSDM performs much better than other compared detectors in suppressing the background and the detection result is very close to the ground-truth map. The ROC curves and AUC scores, which are provided in Fig. 23(c) and Table III, respectively, confirm the proposed TVSDM is a promising method for detecting the anomalies in hyperspectral images.

The runtime on the three real data sets is also reported in Table III. It can be observed that the LKRX takes much longer times than LRX for the San Diego and Hyperion data sets, which is because the outer window regions in these cases

are relatively large. On the whole, the proposed TVSDM has an acceptable computational cost.

### C. Summary

The experimental results on both simulated and real hyperspectral data sets demonstrate the advantages of our proposed anomaly detection method. To summarize, the proposed TVSDM has the following three main features.

- 1) *Effectiveness:* The proposed TVSDM utilizes a union dictionary to represent the background and the anomalies, respectively. The TV and sparsity regularizers are incorporated to achieve better separation. The experimental results show that the proposed TVSDM is highly effective in detecting the anomalies especially the weak ones and maintains a very low false alarm rate when the detection rate achieves 100%.
- 2) *Robustness to the Noise:* The incorporated TV term is powerful in removing the Gaussian noise. Besides, the noise can be further separated from the anomalous part with the potential anomaly dictionary. The experimental results on the first simulated data set show that the detection accuracy of TVSDM remains satisfactory even when the noise level is high.
- 3) *Convenience in Choosing Parameters:* The proposed strategy for the construction of union dictionary spares us from setting proper parameters beforehand in clustering. On the other hand, the detection performance of the proposed TVSDM is insensitive to the choices of the numbers of atoms, i.e.,  $P$  and  $r$ . Meanwhile, the tradeoff parameters  $\beta$  and  $\lambda$  are fixed throughout the experiments and the satisfactory detection results are obtained for all the test hyperspectral data sets.

## V. CONCLUSION

In this article, we have presented a novel hyperspectral anomaly detection method based on the TVSDM. The observed hyperspectral data are decomposed into three components: background, anomaly, and noise. A union dictionary consisting of both background and potential anomalous atoms is utilized to represent the background and anomalies, respectively. For the background part, the TV term is imposed to preserve the spatial relationships of background. For the anomaly part, the sparsity-inducing regularizer is used to characterize the anomalies. Moreover, for the construction of union dictionary, we present a new strategy based on a modified version of density peak-based clustering algorithm. This strategy avoids the troublesome issue of assigning parameters in clustering. Based on the clustering result, some representative pixels including both background and anomalous ones can be picked out simultaneously to form the union dictionary. The experimental results on both simulated and real hyperspectral data sets demonstrate that the proposed TVSDM provides much better detection results than several state-of-the-art anomaly detectors.

One relevant topic deserving future research is the setting of tradeoff parameters in the proposed TVSDM. Although satisfactory results have been obtained for all the test data sets

with a fixed setting, the performance of the proposed method on each real data set might not be optimal. In the future work, we will consider to address this issue with a recently proposed approach named multiobjective optimization [48].

## REFERENCES

- [1] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [2] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
- [3] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.
- [4] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.
- [5] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: Clutter adaptation in anomaly detection," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1855–1871, Aug. 2000.
- [6] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.
- [7] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [8] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [9] H. Goldberg, H. Kwon, and N. M. Nasrabadi, "Kernel eigenspace separation transform for subspace anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 581–585, Oct. 2007.
- [10] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.
- [11] S. Matteoli, M. Diani, and G. Corsini, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Opt. Eng.*, vol. 49, no. 4, pp. 1–16, Apr. 2010.
- [12] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.
- [13] N. Billor, A. S. Hadi, and P. F. Velleman, "BACON: Blocked adaptive computationally efficient outlier nominators," *Comput. Statist. Data Anal.*, vol. 34, no. 3, pp. 279–298, Sep. 2000.
- [14] Q. Ling, Y. Guo, Z. Lin, and W. An, "A constrained sparse representation model for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2358–2371, Apr. 2019.
- [15] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [16] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.
- [17] W. Wang, S. Li, H. Qi, B. Ayhan, C. Kwan, and S. Vance, "Identify anomaly component by sparsity and low rank," in *Proc. IEEE WHISPERS*, Tokyo, Japan, Jun. 2015, pp. 1–4.
- [18] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [19] Y. Niu and B. Wang, "Hyperspectral anomaly detection based on low-rank representation and learned dictionary," *Remote Sens.*, vol. 8, no. 4, p. 289, Mar. 2016.
- [20] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [21] Z. Wu, W. Zhu, J. Chanussot, Y. Xu, and S. Osher, "Hyperspectral anomaly detection via global and local joint modeling of background," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3858–3869, Jul. 2019.
- [22] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerosp. Conf.*, Mar. 2004, pp. 1818–1824.
- [23] S. Matteoli, M. Diani, and G. Corsini, "Impact of signal contamination on the adaptive detection performance of local hyperspectral anomalies," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 1948–1968, Apr. 2014.
- [24] S.-Y. Chen, S. Yang, K. Kalpakis, and C.-I. Chang, "Low-rank decomposition-based anomaly detection," *Proc. SPIE*, vol. 8743, May 2013, Art. no. 87430N.
- [25] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.
- [26] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.
- [27] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.
- [28] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using Mahalanobis distance-regularized tensor RPCA," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2919–2930, May 2018.
- [29] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3034–3047, Jun. 2019.
- [30] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
- [31] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
- [32] M. Dao, C. Kwan, B. Ayhan, and T. D. Tran, "Burn scar detection using cloudy MODIS images via low-rank and sparsity-based models," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Washington, DC, USA, Dec. 2016, pp. 177–181.
- [33] M. Dao, C. Kwan, S. Bernabe, A. J. Plaza, and K. Koperski, "A joint sparsity approach to soil detection using expanded bands of WV-2 images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1869–1873, Dec. 2019.
- [34] D. Ma, Y. Yuan, and Q. Wang, "Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation," *Remote Sens.*, vol. 10, no. 5, p. 745, May 2018.
- [35] Y. Qu *et al.*, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.
- [36] S. Song, H. Zhou, Y. Yang, and J. Song, "Hyperspectral anomaly detection via convolutional neural network and low rank with density-based clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3637–3649, Sep. 2019.
- [37] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2016.
- [38] X. Luo, R. Xue, and J. Yin, "Information-assisted density peak index for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1870–1874, Oct. 2017.
- [39] B. Tu, X. Zhang, X. Kang, G. Zhang, and S. Li, "Density peak-based noisy label detection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1573–1584, Mar. 2019.
- [40] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.
- [41] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 460–489, Jan. 2005.
- [42] 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.
- [43] S. Yang and Z. Shi, "Hyperspectral image target detection improvement based on total variation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2249–2258, May 2016.
- [44] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Univ. Illinois Urbana-Champaign, Champaign, IL, USA, Tech. Rep. UILU-ENG-09-2215, 2009.
- [45] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Apr. 2008.

- [46] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 646–650, Jul. 2011.
- [47] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.
- [48] Z. Cai and Y. Wang, "A multiobjective optimization-based evolutionary algorithm for constrained optimization," *IEEE Trans. Evol. Comput.*, vol. 10, no. 6, pp. 658–675, Dec. 2006.



**Tongkai Cheng** received the B.S. degree in information display and optoelectronic technology from South China University of Technology, Guangzhou, China, in 2016. He is pursuing the Ph.D. degree in circuits and systems with the Department of Electronic Engineering, Fudan University, Shanghai, China.

His main research interests include hyperspectral target detection, machine learning, and pattern recognition.



**Bin Wang** (Senior Member, IEEE) received the B.S. degree in electronic engineering and the M.S. degree in communication and electronic systems from Xidian University, Xi'an, China, in 1985 and 1988, respectively, and the Ph.D. degree in system science from Kobe University, Kobe, Japan, in 1999.

After his graduation in 1988, he was with Xidian University as a Teacher. From 1999 to 2000, he was with the Communications Research Laboratory, Ministry of Posts and Telecommunications, Kobe, as a Research Fellow, working on magnetoencephalography signal processing and its application for brain science. Then, as a Senior Supervisor, he was with the Department of Etching, Tokyo Electron AT Ltd., Yamanashi, Japan, from 2000 to 2002, dealing with the development of advanced process control systems for etching semiconductor equipment. Since September 2002, he has been with the Department of Electronic Engineering, Fudan University, Shanghai, China, where he is a Full Professor and the Leader of the Image and Intelligence Laboratory. He has published more than 120 scientific articles in important domestic and international periodicals. He is the holder of several patents. His main research interests include multispectral/hyperspectral image analysis, automatic target detection and recognition, pattern recognition, signal detection and estimation, and machine learning.

Dr. Wang is an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (IEEE JSTARS).