

Hyperspectral Anomaly Detection via Background and Potential Anomaly Dictionaries Construction

Ning Huyan, Xiangrong Zhang[✉], Senior Member, IEEE, Huiyu Zhou, and Licheng Jiao, Fellow, IEEE

Abstract—In this paper, we propose a new anomaly detection method for hyperspectral images based on two well-designed dictionaries: background dictionary and potential anomaly dictionary. In order to effectively detect an anomaly and eliminate the influence of noise, the original image is decomposed into three components: background, anomalies, and noise. In this way, the anomaly detection task is regarded as a problem of matrix decomposition. Considering the homogeneity of background and the sparsity of anomalies, the low-rank and sparse constraints are imposed in our model. Then, the background and potential anomaly dictionaries are constructed using the background and anomaly priors. For the background dictionary, a joint sparse representation (JSR)-based dictionary selection strategy is proposed, assuming that the frequently used atoms in the overcomplete dictionary tend to be the background. In order to make full use of the prior information of anomalies hidden in the scene, the potential anomaly dictionary is constructed. We define a criterion, i.e., the anomalous level of a pixel, by using the residual calculated in the JSR model within its local region. Then, it is combined with a weighted term to alleviate the influence of noise and background. Experiments show that our proposed anomaly detection method based on potential anomaly and background dictionaries construction can achieve superior results compared with other state-of-the-art methods.

Index Terms—Anomaly detection, background dictionary, hyperspectral images (HSIs), joint sparse representation (JSR), low rank, potential anomaly dictionary.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) are of wide spectral range and high spectral resolution [1], [2], so it contains rich spectral information to discriminate physical properties of different materials [3]. Therefore, HSI finds many applications in different areas, such as land cover classification [4]–[7], urban change detection [8], and crop monitoring [9]. Especially, it is more appealing to detect interesting materials, e.g., targets [10] or anomalies which are very different from

background pixels in an image scene [11]. Anomaly detection [12] based on the HSI has been widely studied in the fields of agriculture [13], mineral exploration [14], maritime rescue [15], and military defense [16], [17]. Different from the supervised target detection, anomaly detection is achieved without knowing the prior information of targets [18]. The difference of statistical distributions between background [19] and anomalies can be utilized for detection [20]. However, due to the absence of spectral information of anomalies, anomaly detection brings more challenges to traditional detection methods.

In recent years, many anomaly detection methods based on the HSI have emerged. As the pioneering work in this area, Chen and Reed [21] justified the assumption that most optical clutters can be modeled as a whitened Gaussian random process with a rapid space-varying mean and a slow-varying covariance. Based on this assumption, they proposed a new constant false alarm rate detector based on generalized likelihood ratio test for multidimensional image data. After that, Reed and Yu [22] developed a method to deal with the signal patterns with nonnegligible and unknown intensities in several optical bands, that is, the well-known Reed–Xiaoli (RX) detector. However, the assumption that the background of an HSI follows a zero mean and unknown covariance multivariate normal distribution is too strong to satisfy in reality. Thus, some modified RX methods were proposed, such as linear-RX and iterative linear-RX [23], weighted-RX [24], segmented-RX [25], subspace-RX [26], kernel-RX [27], and regularized-RX [28]. Recently, a new cluster kernel RX [29] has also been proposed, which groups the background of HSI into clusters and applies a fast eigenvalue decomposition algorithm to achieve anomaly detection. Unfortunately, it is difficult to perfectly model the background utilizing a manually designed distribution form.

To avoid using an inappropriate data distribution, many other methods were also attempted, in which sparse representation [30] has shown great advantages in computer vision, such as face recognition [31], image super-resolution [32], and image denoising [33]. It has also been used in hyperspectral target detection. Chen *et al.* [11] proposed a target detection method based on sparse representation by utilizing a local dual window to construct an adaptive background dictionary and a global target dictionary. In [34], a target detection method based on joint sparse representation (JSR) and multi-task learning has also been proposed. For anomaly detection, due to the fact that there is no prior information about

Manuscript received June 2, 2018; revised August 15, 2018; accepted September 17, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61772400, Grant 61501353, Grant 61772399, Grant 91438201, and Grant 61573267. The work of H. Zhou was supported in part by U.K. EPSRC under Grant EP/N508664/1, Grant EP/R007187/1, and Grant EP/N011074/1 and in part by the Royal Society-Newton Advanced Fellowship under Grant NA160342. (Corresponding author: Xiangrong Zhang.)

N. Huyan, X. Zhang, and L. Jiao are with Xidian University, Xi'an 710071, China (e-mail: xrzhang@ieee.org).

H. Zhou is with the Institute of Electronics, Communications and Information Technology, Queen's University Belfast, Belfast BT3 9DP, U.K.

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

Digital Object Identifier 10.1109/TGRS.2018.2872590

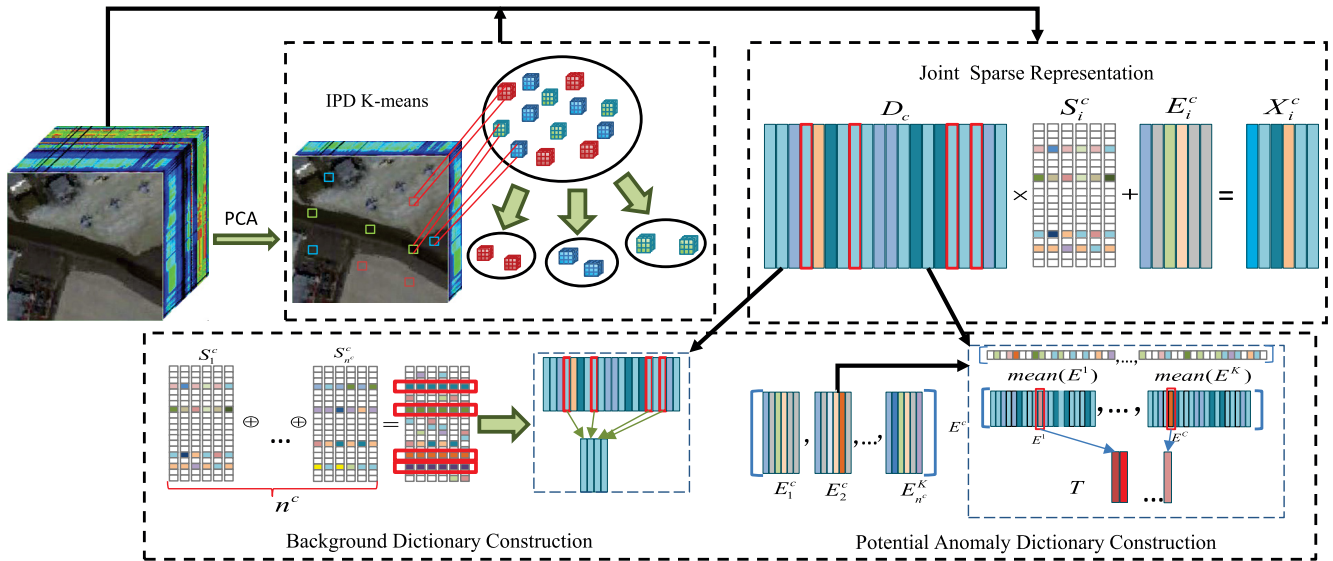


Fig. 1. Schematic of our dictionary construction method for hyperspectral anomaly detection. First, the PCA is used to get a low-dimensional 3-D cube. Then, the original image is separated into many 3-D regions by a fixed-size window. After that, the IPD-based K -means is used to group the regions into several classes. The original spectral pixels corresponding to each region are reconstructed using the JSR model with class-based overcomplete dictionary. The coefficient of JSR is used to construct the background, and the residual in JSR is used to construct the potential anomaly dictionary.

a target, existing methods try to make full use of background information. Most of them assumed that background pixels in the center of a local region can be represented by the combination of other pixels in the region, while anomalous pixels cannot. With this assumption, in [35], an anomaly detection method was presented based on JSR of background, which utilized the characteristics of JSR, that all the similar pixels within a local region can be jointly represented in the same low-dimensional subspaces. In [36], a novel sparsity score estimation framework based on sparse representation was proposed for anomaly detection.

Since HSI usually has large homogeneous regions [10] whose majority of pixels have similar spectral characteristics, the structure can be represented by an underlying subspace using a subspace learning method. The commonly used methods include principal component analysis (PCA) [37] and robust PCA (RPCA) [38]. For better multisubspace learning, low-rank representation (LRR) [39] was proposed. It attempts to minimize the rank of a dictionary with the corresponding coefficients while decomposing the original data into low rank and noise components. For HSI data, an LRR technique has been used in classification [40], [41] and denoising [42]. Recently, this method was also utilized to model the problem of anomaly detection [43], [44] based on the assumption that the background has low-rank properties, while the anomalies demonstrate sparse properties. In [43], the observed data were decomposed into background and anomaly parts. The coefficients of the model were constrained to be both low rank and sparse in order to obtain the global and local structures of the background. Because of the effectiveness of low-rank and sparsity constraints for background modeling, this method is achieved a promising performance. However, anomaly and noise distributions are aliasing owing to similar sparse characteristics. Therefore, it is a challenge to distinguish

noise from anomaly component. In [44], both background and anomaly priors were considered and each part of the original data was modeled separately. It was assumed that the background has the low-rank property, while an anomaly owns the sparsity property, so the observed data were decomposed into background, anomaly, and noise parts by extending the RPCA model. An anomaly was then detected based on the Mahalanobis distance.

To accurately model background and anomaly information, we propose a new HSI anomaly detection method based on the well-designed background and potential anomaly dictionaries utilizing low-rank and sparse representation (LRSR) strategy. We decompose the original data into three components, i.e., background, anomaly, and noise. The background component is constrained to have a low-rank property due to the homogeneity of HSIs. The anomaly component is constrained to have the sparsity property. More importantly, two well-designed dictionaries are constructed to constitute our model. The JSR model is used to depict the pure background without the influence of clutters and anomalies. In order to make full use of the anomaly information hidden in the HSI, a potential anomaly dictionary is constructed. The atoms of the dictionary are selected according to the anomalous level (AL) utilizing the residual calculated in the JSR model combined with a weighted term. With the carefully constructed background and potential anomaly dictionaries, the HSI can be separated into background, anomaly, and noise components using a sparse and low-rank decomposition model. The schematic of the proposed algorithm is shown in Fig. 1. The main contribution in our proposed anomaly detection method based on potential anomaly and background dictionary construction (PAB-DC) can be summarized as follows.

- 1) A new low-rank and sparsity-based anomaly detection model is proposed with two well-designed dictionaries,

i.e., background and potential anomaly dictionaries, so that the original data can be properly decomposed into background, anomaly, and noise components.

- 2) For the background dictionary construction, local and nonlocal similarities of each region in the scene are considered, and we propose to use the coefficients of the JSR to select the atoms.
- 3) Making full use of the anomaly prior information hidden in the scene, we propose to construct a potential anomaly dictionary utilizing the residual of each local region by the JSR model.

The remainder of this paper is organized as follows. The detailed introduction of our method is given in Section II. The experimental results and discussions are reported in Section III. Finally, we conclude this paper in Section IV.

II. PROPOSED METHOD

A. Background, Anomaly, and Noise Decomposition Model

Let an HSI data be denoted as $\mathbb{X} \in \mathbb{R}^{h \times w \times d}$, where d is the number of the spectral bands, and h and w are the spatial size of the data. For convenience, we transform the 3-D cube \mathbb{X} into a 2-D matrix $\mathbf{X} = \{x_i\}_{i=1}^n \in \mathbb{R}^{d \times n}$, where each column of \mathbf{X} is a spectral pixel vector in the HSI and $n = h \times w$ is the number of the pixels. In this paper, we formulate the anomaly detection task as a matrix decomposition problem. The HSI data matrix is decomposed into three components: background, anomaly, and noise. Considering that there usually exists a strong correlation between background pixels, which can be represented by the combination of other background pixels, and we want to distinguish the anomalies and noise simultaneously, our decomposition model is formulated as

$$\mathbf{X} = \mathbf{BZ} + \mathbf{A} + \mathbf{E} \quad (1)$$

where \mathbf{BZ} is the background component, $\mathbf{B} = [b_1, b_2, \dots, b_{n_B}]$ is the background dictionary, n_B is the number of the atoms in the dictionary, $\mathbf{Z} = [z_1, z_2, \dots, z_n]$ are the corresponding representation coefficients, and $\mathbf{A} = [a_1, a_2, \dots, a_n]$ and $\mathbf{E} = [e_1, e_2, \dots, e_n]$ are the anomaly and noise components, respectively. Intuitively, the whole spectral space can be divided into several underlying subspaces. For an HSI, pixels in a local region are most likely homogeneous, so we assume that the background holds a low-rank property. For the noise component, it has been investigated that there are mainly two kinds of noise existing in HSIs, including sparse noise (strip and deadline) and Gaussian random noise [45]. Compared with l_2 and l_1 norms, $l_{2,1}$ norm is more robust to describe both sparse noise and Gaussian random noise. Therefore, in this paper, $l_{2,1}$ norm is utilized to model the noise. Thus, the objective function can be further formulated as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{A}, \mathbf{E}} \quad & \text{rank}(\mathbf{Z}) + \beta \|\mathbf{A}\|_l + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{BZ} + \mathbf{A} + \mathbf{E} \end{aligned} \quad (2)$$

where $\beta > 0$ and $\lambda > 0$ are the coefficients used to balance all the components. $\text{rank}(\mathbf{Z})$ represents the rank of matrix \mathbf{Z} which is the coefficient matrix of the low-rank representation.

The $l_{2,1}$ norm is defined as the sum of the l_2 norm of the columns in a matrix, i.e.,

$$\|\mathbf{E}\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^d (e_{i,j})^2} \quad (3)$$

which attempts to enforce each element of the matrix to approach zero except for some outliers.

It is difficult to estimate anomalies using a particular distribution, because anomalies may be different in the same HSIs. Thus, pixels with significant differences from the background are extracted and used as potential prior of the other anomalous pixels. Intuitively, anomalies chosen as the atoms of the potential anomaly dictionary are related to the other ones hidden in the data set. Therefore, a hidden anomaly can be represented by the linear combination of the predetected strong anomaly atoms in the dictionary, namely, $\mathbf{A} = \mathbf{TS}$, where $\mathbf{T} = [t_1, t_2, \dots, t_{n_T}]$ is the potential anomaly dictionary, n_T is the number of atoms in potential anomaly dictionary, and $\mathbf{S} = [s_1, s_2, \dots, s_n]$ is the corresponding coefficient matrix. However, when generating the potential anomaly dictionary, we may mistakenly include some pixels that are not anomalies. To avoid this situation, we assume that only the potential anomaly atoms are active when reconstructing an anomaly pixel. The atoms in the potential anomaly dictionary are expected to be the supportive bases for the reconstruction of an anomaly. Since these pixels are randomly distributed in the scene, anomalous pixels retain sparse characteristics. In this way, we constrain the coefficients matrix to be sparse. As a result, anomalous pixels are reconstructed using the atoms as few as possible from the potential anomaly dictionary. The above-mentioned formulation is nonconvex and NP-hard. Fortunately, under certain conditions [38], the problem of finding a low-rank approximation for a given matrix can be solved by minimizing its nuclear norm. Now, the model for our proposed anomaly detection becomes as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{S}, \mathbf{E}} \quad & \|\mathbf{Z}\|_* + \beta \|\mathbf{S}\|_1 + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{BZ} + \mathbf{TS} + \mathbf{E} \end{aligned} \quad (4)$$

where $\|\cdot\|_*$ denotes the matrix's nuclear norm. The response value of each pixel belonging to anomalies can be calculated by the l_2 norm of each column of the anomaly component $\mathbf{A} = \mathbf{TS}$, i.e., $r_i = \|a_i\|_2$ $i = 1, 2, \dots, n$. Finally, anomalies can be determined by a predefined threshold.

B. Background Dictionary Construction

For the pixels within a local region of an HSI, they may share common structures. Therefore, these regions can be jointly approximated by a sparse linear combination of a few common atoms. The objective of the JSR [46] is

$$\begin{aligned} \min \quad & \|\psi\|_{\text{row},0} \\ \text{s.t.} \quad & \mathbf{U} = \mathbf{V}\psi + \mathbf{R} \end{aligned} \quad (5)$$

where $\mathbf{U} = [u_1, u_1, \dots, u_L]$ is a 2-D matrix with L spectral pixels flattened from the 3-D local region cube and ψ is the corresponding representation coefficients where only a few

rows are nonzero, $\mathbf{R} = [r_1, r_2, \dots, r_L]$ are the residuals after construction based on \mathbf{V} and ψ , and $\|\cdot\|_{\text{rows},0}$ denotes the nonzero rows of ψ .

The JSR model represents each spectral pixel within a local region using common atoms chosen from an overcomplete dictionary. Thus, it captures the common part from the original region by the linear combination of a few atoms, which reflects the consistent spectral information in this region. Obviously, for the anomaly detection task, the common part within a region tends to be regarded as the background, so we choose the atoms that are frequently used as the bases to reconstruct background. All the chosen atoms from different classes of background materials form the background dictionary. Now, the challenge is how to construct an overcomplete dictionary for the JSR model and how to design the metric to measure the frequently used atoms chosen from the dictionary.

For the first problem, a simple method to construct the dictionary in the JSR model is to utilize all the original spectral pixels in HSI. However, it is time-consuming to calculate each small region over such a large dictionary. Although a wide range of spectral information from different materials may boost the representation ability of the JSR model, it actually degrades the ability of describing a particular material. This is because too many pixels from different classes are involved when reconstructing a local region. Consequently, the global dictionary may be confused when intending to extract a common structure.

Considering the nonlocal similarity between the local regions, we use an extended k -means clustering algorithm to group all the pixels into several clusters so that each group contains a similar underlying structure for reducing complexity. Let the small region under investigation be denoted as a 3-D cube $\mathbf{U} \in \mathbb{R}^{\text{win} \times \text{win} \times d}$ of size $\text{win} \times \text{win} \times d$, where win is the window size. Its 2-D form is $\mathbf{U} \in \mathbb{R}^{d \times L}$, where $L = \text{win} \times \text{win}$. The image patch distance (IPD) [47] is used to measure the distance between two regions, which is defined as

$$o_{\text{IPD}}(\mathbf{U}_P, \mathbf{U}_Q) = o(P(u_i), Q(u_j)) \\ = \sum_{h=1}^L \max \begin{pmatrix} \min_{b \in Q(u_j)} o(a_h, b) \\ \min_{a \in P(u_i)} o(b_h, a) \end{pmatrix} \quad (6)$$

where $o(a, b)$ is a nonlocal spectral similarity function, and (here it is the Euclidean distance), and \mathbf{U}_P and \mathbf{U}_Q are two 2-D matrices representing two regions selected from the regions. The calculation procedure is shown in Fig. 2.

With the increase in the number of classes, the number of the regions in each class usually decreases. As a result, the number of the regions in the smaller clusters is insufficient to construct an overcomplete dictionary by the JSR model. To solve this problem, those nonovercomplete classes will be merged into the nearest class based on the IPD distance.

In this way, we group all the regions into several clusters. For each cluster, the overcomplete dictionary is made up of the overall spectral items in the class which are used to reconstruct each local region. The frequently used atoms in the dictionary are treated as the final background dictionary atoms.

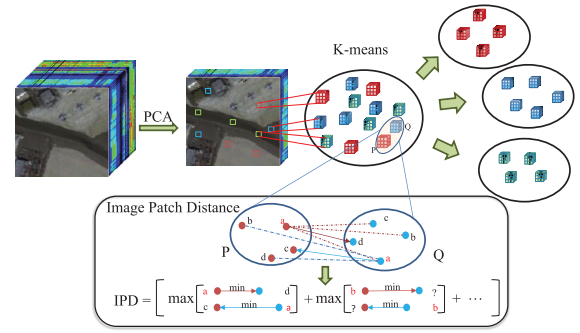


Fig. 2. Procedure of IPD-based K -means.

The ultimate formulation of the JSR model for each class is as

$$\min \|\psi_i^c\|_{\text{row},0} \\ \text{s.t. } \mathbf{U}_i^c = \mathbf{G}^c \psi_i^c + \mathbf{R}_i^c \quad (7)$$

where \mathbf{U}_i^c is the i th region of the c th class, $i = [1, 2, \dots, n^c]$, $n^c = [n^1, n^2, \dots, n^K]$ and $c = [1, 2, \dots, K]$, n^c is the number of the regions in the c th class and K is the number of classes, $\mathbf{G}^c = [g_1^c, g_2^c, \dots, g_{n^c}^c]$ is the overcomplete dictionary of the c th class with n^c atoms, and ψ_i^c and \mathbf{R}_i^c are the corresponding coefficients and residual, respectively.

The second problem is to measure the representation frequency of each atom in each class. In this paper, the frequency is defined as the sum of the JSR coefficients after normalization. It can be expressed as

$$P^c = \frac{1}{\gamma} \sum_{i=1}^{n^c} \sum_{j=1}^L |\psi_{i,j}^c| \\ \gamma = \text{sum} \left(\sum_{i=1}^{n^c} \sum_{j=1}^L |\psi_{i,j}^c| \right) \quad (8)$$

where $\psi_{i,j}^c$ is the j th column of the i th region in the c th class, γ is a normalization term, P^c is a vector, in which the value of each element reflects the weighted frequency chosen as the background dictionary atom, and $\text{sum}(\cdot)$ denotes the elementwise sum of a vector. We sort it in descending order and the top atoms are chosen as background dictionary atoms for the c th class. Finally, all the chosen atoms from each class construct the background dictionary \mathbf{B} .

C. Potential Anomaly Dictionary Construction

We believe that some anomalous pixels with strong responses to the background can be detected by the JSR model and they can be considered as the prior in order to detect the other anomalies. If a region is not completely homogenous, that is, there are outliers or different pixels, these pixels are more likely anomalies. Such heterogeneity in a local region is reflected by the residual computed by the JSR model. Therefore, the pixels in a region leading to a large reconstruction residual are claimed to be anomalous pixels. Meanwhile, in order to reduce the influence of isolated noise in the scene, we use the region-based residual to select the potential anomaly atoms. For the central pixel in a region,

a larger average residual of its neighboring pixels means that the central pixel is most likely an anomaly. Here, the mean residual of the regions is regarded as the anomaly response of the central pixel.

For the i th region of the c th class, the response of the anomaly is calculated by

$$\bar{\mathbf{R}}_i^c = \frac{1}{L} \sum_{j=1}^L \|r_{i,j}^c\|_2 \quad (9)$$

where $r_{i,j}^c$ is the j th column of the i th region in c th class. Therefore, the error in the c th class is

$$\mathbf{R}_{\text{mean}}^c = [\bar{\mathbf{R}}_1^c, \bar{\mathbf{R}}_2^c, \dots, \bar{\mathbf{R}}_{n^c}^c]. \quad (10)$$

The average residual of each region in K classes is concatenated, and then, we define the response value of a pixel being anomaly, namely, AL as

$$\begin{aligned} \text{AL} &= \frac{1}{\chi} [\mathbf{R}_{\text{mean}}^1, \mathbf{R}_{\text{mean}}^2, \dots, \mathbf{R}_{\text{mean}}^K] \\ \chi &= \sum_{c=1}^K \sum_{i=1}^{n^c} \bar{\mathbf{R}}_i^c = \frac{1}{L} \sum_{c=1}^K \sum_{i=1}^{n^c} \sum_{j=1}^L \|r_{i,j}^c\|_2 \end{aligned} \quad (11)$$

where χ is the normalization part and AL is a vector in which each element is the anomaly response value of a region.

However, it is not enough to extract obvious anomalous pixels. On the one hand, noise is always involved. On the other hand, for a complicated scene, a region may contain different types of background materials that may be mistakenly regarded as potential anomaly pixels. In other words, not all the pixels with large reconstruction residuals are anomalies. For this concern, we take into account the importance of atoms participating in reconstructing other pixels. Compared with other atoms, the selected background and anomaly atoms are more significant. Therefore, we can eliminate the atoms that do not take part in reconstruction in order to alleviate the interference.

In addition, anomalies are selected as atoms only when the current regions contain an obvious anomaly structure. In order to identify these truly anomalous pixels, we utilize the difference between the JSR coefficients when reconstructing the background and anomaly. The way to estimating the atoms of the background dictionary actually includes two levels of information. The first level is the times of an atom being selected. It describes the participation quantity of each atom in a certain class. The second level is the absolute value of the corresponding coefficients as shown in (8), which describes the importance degree of an atom in region reconstruction. For the first level, the selection times of the background atoms are always high, while the selection times of anomaly atoms are low when reconstructing each of the local regions. For the second level, the chosen anomaly atoms have strong coefficients only when reconstructing a local region including many anomalies. So the frequently used atoms for each class tends to be background, while those atoms that only participate in reconstructing a region containing anomalies tends to be anomalies. The anomaly atoms in each class show the property of low selection frequency and large coefficient. We define

the anomalous weight (AW) as follows to depict the above-mentioned issue:

$$\text{AW}^c = \frac{P^c}{F^c} \quad (12)$$

where F^c is a vector in which each element reflects the select times of each atom, which can be written as

$$F^c = \sum_{i=1}^{n^c} \sum_{j=1}^L \text{sgn}(|\varphi_{i,j}^c|)$$

where sgn is the sign function. The weighted AL can be expressed as

$$\begin{aligned} \text{AL} &= \frac{1}{\chi} [\mathbf{R}_{\text{mean}}^1, \mathbf{R}_{\text{mean}}^2, \dots, \mathbf{R}_{\text{mean}}^K] \odot \mathbf{AW} \\ \chi &= \sum_{c=1}^K \sum_{i=1}^{n^c} \bar{\mathbf{R}}_i^c = \frac{1}{L} \sum_{c=1}^K \sum_{i=1}^{n^c} \sum_{j=1}^L \|r_{i,j}^c\|_2 \end{aligned} \quad (13)$$

where \odot denotes the elementwise multiplication and $\mathbf{AW} = \{\text{AW}^1, \text{AW}^2, \dots, \text{AW}^K\}$. We sort them in a descending order and choose the top pixels as the atoms of the potential anomaly dictionary \mathbf{T} .

D. Optimization and Computational Complexity

To solve the problem shown in (4), for convenience, two auxiliary variables \mathbf{J} and \mathbf{L} are introduced to make the objective function separable. Thus, the problem can be converted to the following form:

$$\begin{aligned} \min_{\mathbf{J}, \mathbf{E}, \mathbf{Z}, \mathbf{S}, \mathbf{L}} \quad & \|\mathbf{J}\|_* + \beta \|\mathbf{L}\|_1 + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{BZ} + \mathbf{TS} + \mathbf{E}, \quad \mathbf{Z} = \mathbf{J}, \quad \mathbf{S} = \mathbf{L}. \end{aligned} \quad (14)$$

We solve (14) by utilizing the augmented Lagrange multiplier method reported in [48], which is implemented by updating one variable with others being fixed

$$\begin{aligned} \ell &= \|\mathbf{J}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \beta \|\mathbf{L}\|_1 \\ &+ \langle \mathbf{Y}_1, \mathbf{X} - \mathbf{BZ} - \mathbf{E} - \mathbf{TS} \rangle + \langle \mathbf{Y}_2, \mathbf{Z} - \mathbf{J} \rangle \\ &+ \langle \mathbf{Y}_3, \mathbf{S} - \mathbf{L} \rangle + \frac{\mu}{2} (\|\mathbf{X} - \mathbf{BZ} - \mathbf{E} - \mathbf{TS}\|_F^2 \\ &\quad + \|\mathbf{Z} - \mathbf{J}\|_F^2 + \|\mathbf{S} - \mathbf{L}\|_F^2) \end{aligned} \quad (15)$$

where \mathbf{Y}_1 , \mathbf{Y}_2 , and \mathbf{Y}_3 are the Lagrange multipliers and $\mu > 0$ is a penalty parameter. The problem can be resolved using the following steps.

- 1) Fix $\mathbf{E}, \mathbf{L}, \mathbf{S}, \mathbf{Z}$ and update \mathbf{J} . The objective can be derived as

$$\min_{\mathbf{J}} \|\mathbf{J}\|_* + \frac{\mu}{2} \left\| \mathbf{J} - \left(\mathbf{Z} + \frac{\mathbf{Y}_2}{\mu} \right) \right\|_F^2. \quad (16)$$

- 2) Fix $\mathbf{J}, \mathbf{L}, \mathbf{S}, \mathbf{Z}$ and update \mathbf{E} . The objective can be derived as

$$\min_{\mathbf{E}} \lambda \|\mathbf{E}\|_{2,1} + \frac{\mu}{2} \left\| \mathbf{E} - \left(\mathbf{X} - \mathbf{BZ} - \mathbf{TS} + \frac{\mathbf{Y}_1}{\mu} \right) \right\|_F^2. \quad (17)$$

- 3) Fix $\mathbf{J}, \mathbf{E}, \mathbf{S}, \mathbf{Z}$ and update \mathbf{L} . The objective can be derived as

$$\min_{\mathbf{L}} \beta \|\mathbf{L}\|_1 + \frac{\mu}{2} \left\| \mathbf{L} - \left(\mathbf{S} + \frac{\mathbf{Y}_3}{\mu} \right) \right\|_F^2. \quad (18)$$

The nuclear norm, l_1 and $l_{2,1}$ norms, can be solved by singular value thresholding [49], soft thresholding [50], and $l_{2,1}$ norm minimization operator [39]. The complete procedure is summarized in Algorithm 1.

Algorithm 1 Algorithm for Solving Background, Anomaly, and Noise Decomposition

Input: Data matrix \mathbf{X} , parameters $\lambda > 0$ and $\beta > 0$

Initialize: $\mathbf{Z} = \mathbf{J} = \mathbf{S} = \mathbf{L} = 0$, $\mathbf{E} = 0$, $\mathbf{Y}_1 = \mathbf{Y}_2 = \mathbf{Y}_3 = 0$, $\mu = 10^{-6}$, $\mu_{\max} = 10^{10}$, $\rho = 1.2$, $\varepsilon = 10^{-6}$

Output: $\mathbf{Z}, \mathbf{E}, \mathbf{S}$

1. While not converged do
2. Fix others and update \mathbf{J} by Eq. (16)
3. Fix others and update \mathbf{L} by Eq. (17)
4. Fix others and update \mathbf{E} by Eq. (18)
5. Fix others and update \mathbf{Z} by

$$\mathbf{Z} := (\mathbf{B}^T \mathbf{B} + \mathbf{I})^{-1} [\mathbf{B}^T \mathbf{X} - \mathbf{B}^T \mathbf{TS} - \mathbf{B}^T \mathbf{E} + \mathbf{J} + (\mathbf{B}^T \mathbf{Y}_1 - \mathbf{Y}_2)/\mu]$$

6. Fix others and update \mathbf{S} by

$$\mathbf{S} := (\mathbf{T}^T \mathbf{T} + \mathbf{I})^{-1} (\mathbf{T}^T \mathbf{X} - \mathbf{T}^T \mathbf{BZ} - \mathbf{T}^T \mathbf{E} + \mathbf{L} + (\mathbf{T}^T \mathbf{Y}_1 - \mathbf{Y}_3)/\mu)$$

7. Update the three Lagrange multipliers

$$\mathbf{Y}_1 := \mathbf{Y}_1 + \mu(\mathbf{X} - \mathbf{BZ} - \mathbf{E} - \mathbf{DS})$$

$$\mathbf{Y}_2 := \mathbf{Y}_2 + \mu(\mathbf{Z} - \mathbf{J})$$

$$\mathbf{Y}_3 := \mathbf{Y}_3 + \mu(\mathbf{S} - \mathbf{L})$$

8. Update the parameter μ ,

$$\mu = \min(\rho\mu, \mu_{\max})$$

9. Check the convergence conditions

$$\|\mathbf{X} - \mathbf{BZ} - \mathbf{DS}\|_F < \varepsilon$$

$$\|\mathbf{Z} - \mathbf{J}\|_F < \varepsilon$$

$$\|\mathbf{S} - \mathbf{L}\|_F < \varepsilon$$

10. End While
-

The computation of our method includes those of Algorithm 1 and dictionary construction. For the first aspects, the major computation is Step 2, which requires computing the singular value decomposition of an $n \times n$ matrix. Therefore, it is time-consuming if n is large. However, the computational cost of this step can be easily reduced using the method reported in [39]. The optimal solution \mathbf{Z}^* (with respect to the variable \mathbf{Z}) to (4) always lies within the subspace spanned by the rows of \mathbf{A} . This means that \mathbf{Z}^* can be factorized into $\mathbf{Z}^* = \mathbf{P}^* \tilde{\mathbf{Z}}^*$, where \mathbf{P}^* can be computed in advance by orthogonalizing the columns of \mathbf{B}^T . Therefore, our model

shown in (4) can be rewritten as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{S}, \mathbf{E}} \quad & \|\tilde{\mathbf{Z}}\|_* + \beta \|\mathbf{S}\|_1 + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t} \quad & \mathbf{X} = \mathbf{B}^* \tilde{\mathbf{Z}} + \mathbf{TS} + \mathbf{E} \end{aligned} \quad (19)$$

where $\mathbf{B}^* = \mathbf{BP}^*$. Since the number of rows of $\tilde{\mathbf{Z}}$ is at most r_b (the rank of matrix \mathbf{B} $r_b \ll n$), therefore the computation complexity of Step 2 is $O(r_b^3)$. Noted that, $\mathbf{X} = \{x_i\}_{i=1}^n \in \mathbb{R}^{d \times n}$, so the computation complexity of Steps 3 and 4 is $O(dnr_B)$. For the dictionary \mathbf{B} and potential anomaly dictionary \mathbf{T} , their column numbers are n_B and n_T , respectively. In Steps 5 and 6, the term $(\mathbf{B}^T \mathbf{B} + \mathbf{I})^{-1}$ and $(\mathbf{T}^T \mathbf{T} + \mathbf{I})^{-1}$ can be calculated in advance, so the computation complexity of these two steps are $O(nr_B^2)$ and $O(nn_T^2)$, respectively. Therefore, the complexity of Algorithm 1 is $O(n_s(r_b^3 + dnr_B + nr_B^2 + nn_T^2))$, where n_s is the number of iterations. The major computation of the dictionary construction addresses on solving the JSR model. In this paper, the Orthogonal Matching Pursuit-Cholesky-based method [51] is used to solve (14). However, the dictionaries in the JSR model vary across different classes. For convenience, we consider that there is only one dominant class so that we can reconstruct the regions using one dictionary. In this way, the upper bound of the computational complexity can be calculated as $O(nk(2ndL + nL^2 + 2n + k^2))$, where k is the sparsity level. The computational complexity of the IPD k means is $O(n_k(d_r L K n))$, where d_r is the number of the bands after dimensionality reduction, K is the number of clusters, and n_k is the number of the iterations. Therefore, the computation complexity of the entire algorithm is $(n_k(d_r L K n) + nk(2ndL + nL^2 + 2n + k^2) + n_s(r_b^3 + dnr_B + nr_B^2 + nn_T^2))$.

Finally, our proposed PAB-DC is summarized in Algorithm 2. The construction of the background and potential anomaly dictionaries is shown in Fig. 1.

III. EXPERIMENTS RESULTS

In this section, we first compare our proposed PAB-DC method with the widely used anomaly detection algorithms, such as Global-RX, Local-RX [22], the detector based on collaborative representation (CRD) [52], and the LRSR-based detector [43] on five real HSI data sets. The CRD method was proposed based on the assumption that the background can be approximately represented by its spatial neighborhoods, while anomalies cannot. The LRSR is a detector which tries to model the global and local structure of background using low rank and sparse constraints. In addition, the effectiveness of the potential anomaly dictionary is investigated and the parameters analysis is given. All the experiments are conducted on a workstation with Intel Xeon Processor E5-2630 v3 2.40 GHz \times 32 and 64 GB RAM.

A color detection map is provided to illustrate the results. The colors range from dark blue to bright yellow which reflect the responses of pixels in the current component, that is, the brighter the pixels, the stronger responses to the corresponding component. In addition, for qualitative comparisons, the receiver operating characteristic (ROC) curves with pointwise confidence intervals are used in our experiments. The bias-corrected and accelerated percentile method that is a

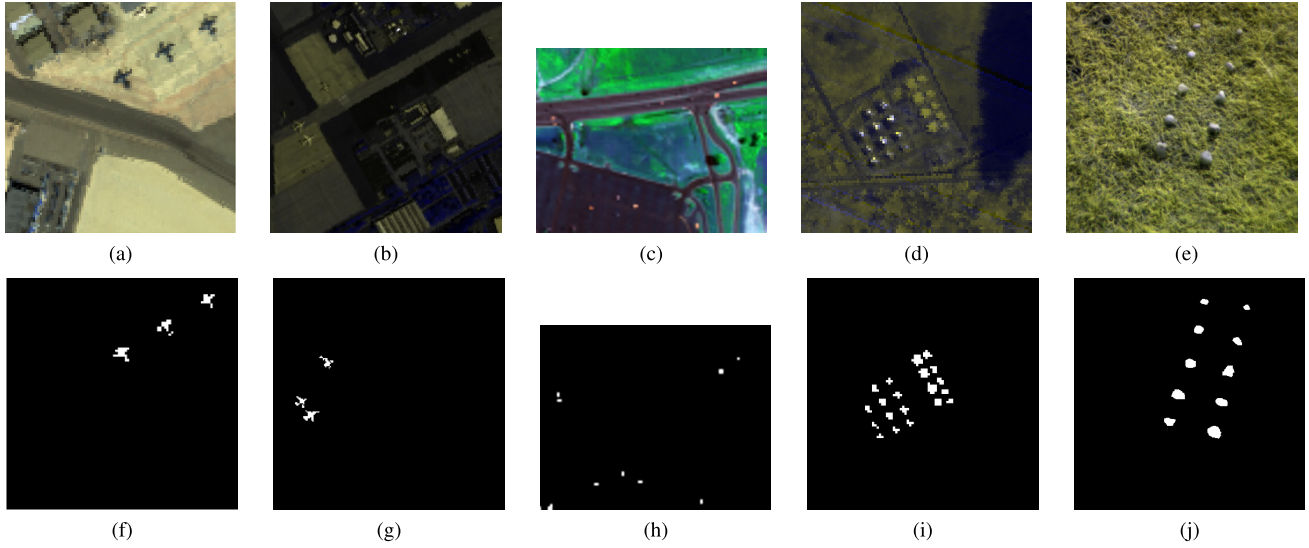


Fig. 3. Image descriptions. (a) False color image of AVIRIS-I. (b) False color image of AVIRIS-II. (c) False color image of the HYDICE data set. (d) False color image of the Urban data set. (e) False color image of the Cri data set. (f) Ground truth of AVIRIS-I. (g) Ground truth of AVIRIS-II. (h) Ground truth of the HYDICE data set. (i) Ground truth of the Urban data set. (j) Ground truth of the Cri data set.

Algorithm 2 Algorithm for PAB-DC

Input: 2D data matrix \mathbf{X} , 3D data cube \mathbf{R} , parameters $\lambda > 0$ and $\beta > 0$, K , M , η , W , ρ

Output: $\mathbf{Z}, \mathbf{E}, \mathbf{S}$

1. Generate a local region \mathbf{U}_i ($i = 1, 2, 3, \dots, n$) from the dataset with a window of size $win \times win$;

$$\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n]$$

2. Obtain the class label of each \mathbf{U}_i using a K-means method based on the IPD on the regions after dimensionality reduction by PCA;
- 3) Group the regions into K classes;

$$\mathbf{U} = [\mathbf{U}^1, \mathbf{U}^2, \dots, \mathbf{U}^K] = [\mathbf{U}_1^1, \mathbf{U}_2^1, \dots, \mathbf{U}_{n_1}^1] \cup [\mathbf{U}_1^2, \mathbf{U}_2^2, \dots, \mathbf{U}_{n_2}^2] \cup \dots \cup [\mathbf{U}_1^K, \mathbf{U}_2^K, \dots, \mathbf{U}_{n_K}^K]$$

4. Calculate the representation coefficient ψ_i^c and residual \mathbf{R}_i^c for each region in each class based on the JSR model by Eq. (7);
5. Generate the background dictionary by Eq. (8);
6. Generate the potential anomaly dictionary by Eq. (13);
7. Apply the low rank and spare decomposition model using the background and potential anomaly dictionaries from

Algorithm 1

kind of bootstrap-based confidence bound estimation method is used to estimate the confidence intervals on the level of confidence being 95% [53], [54]. The area under curve (AUC) is summarized with upper bound and lower bound based on the confidence intervals.

A. Data Set Description

The first data set was provided in [43]. It was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)

over San Diego, CA, USA. The spatial resolution is 3.5 m per pixel. It has 224 spectral bands in the wavelengths ranging from 370 to 2510 nm. After removing the bands that correspond to the water absorption regions, low SNR, and poor quality (1–6, 33–35, 97, 107–113, 153–166, and 221–224), 189 bands are utilized in our experiment. The whole data set has an image size of 400×400 . There are two kinds of airplanes in the scene that are treated as an anomaly. From top-left of this hyperspectral data set, a region of 100×100 pixels is selected as AVIRIS-I for testing. The three airplanes are regarded as an anomaly in the scene. The anomalous pixels refer to the main body and edges of the airplanes with a total of 57 pixels. Fig. 3(a) and (f) shows the false color image and the ground-truth map of AVIRIS-I data set, respectively.

The second data set AVIRIS-II is a 200×200 area selected from the AVIRIS image, which is located at the center of the San Diego region. Compared with AVIRIS-I, this data set has more different types of background materials, including roofs, grasses, shadow, roads, and so on. The three airplanes with 134 pixels are regarded as anomalous pixels. The false color image and the ground-truth map are shown in Fig. 3(b) and (g), respectively.

The third data set used in the experiment is obtained from an aircraft platform with a Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor. The image has a spectral resolution 10 nm and a spatial resolution of 1 m. It covers an urban area that comprises a vegetation area, a construction area, and several roads, including some vehicles. The whole data set has a size of 307×307 pixels. In this experiment, a subscene 80×100 on the upper right of the whole scene is used. The 21 anomalous pixels are about vehicles with different sizes [52]. The false color image and the ground-truth map are shown in Fig. 3(c) and (h), respectively.

The fourth data set is from an open Airport-Beach-Urban data set [54]. The sample images in this data set were manually extracted from large images downloaded from the

TABLE I
AUC COMPARISON OF THE METHODS WITH CONFIDENCE INTERVAL

	Global-RX	Local-RX	CRD	LRSR	PAB-DC
AVIRIS-I	0.9091 ^{+0.0268} -0.0369	0.6914 ^{+0.0668} -0.0908	0.9530 ^{+0.0118} -0.0206	0.9779 ^{+0.0072} -0.0109	0.9950 ^{+0.0032} -0.0105
AVIRIS-II	0.8870 ^{+0.0191} -0.0268	0.8159 ^{+0.0735} -0.0419	0.8765 ^{+0.0165} -0.0285	0.7710 ^{+0.0239} -0.0452	0.9186 ^{+0.0071} -0.0080
HYDICE	0.9867 ^{+0.0074} -0.0209	0.8983 ^{+0.0838} -0.1560	0.9885 ^{+0.0087} -0.0146	0.9303 ^{+0.0312} -0.0671	0.9907 ^{+0.0043} -0.0077
Urban	0.9946 ^{+0.0012} -0.0017	0.9157 ^{+0.0148} -0.0256	0.9309 ^{+0.0095} -0.0135	0.9872 ^{+0.0047} -0.0053	0.9937 ^{+0.0020} -0.0026
Cri	0.9134 ^{+0.0069} -0.0078	0.7543 ^{+0.0084} -0.0141	0.6737 ^{+0.0248} -0.0117	0.5545 ^{+0.0135} -0.0111	0.9661 ^{+0.0032} -0.0041

AVIRIS website. We use the data set with the size of 100×100 , as shown in Fig. 3(d) and (i). It was collected over Texas Coast on August 29, 2010. The spatial resolution is 17.2 m per pixel. The noisy bands in the original images had been removed, and the ground truth is manually labeled with the help of the Environment for Visualizing Images software [55].

The fifth data set was acquired by the Nuance Cri hyperspectral sensor. The spectral resolution of this data set is 10 nm. The image scene covers an area of 400×400 pixels, with 46 spectral bands in the wavelengths ranging from 650 to 1100 nm. The 10 rocks in this scene can be regarded as anomaly to be detected which is different from the grassy background, as shown in Fig. 3(e) and (j) [44].

B. Detection Performance

The detection performance of our proposed PAB-DC is evaluated and compared with four other state-of-the-art detectors: Global-RX, Local-RX, CRD, and LRSR. The dual windows (win_{in} , win_{out}) in the Local-RX and CRD are set as (3, 5) for the HYDICE data set and (7, 13) for the others. The number of clusters of LRSR and the number of the pixels chosen for constructing the background dictionary are set to be 15 and 20, respectively, as in [43] for all data sets. The parameters β and λ in our model are 0.01 and 0.1, which is quite stable across different data sets except for Cri data set, where λ is 10 to get the best performance. The window size in PAB-DC is 1×1 for the HYDICE data set, because the anomalies presented in it are isolated pixels, while it is set to be 3×3 for the other data sets empirically. The number of classes K is determined by the number of background materials in different data sets. Therefore, for the Urban data set, K is 5; for the AVIRIS-I and HYDICE data sets, K are 10; and for the AVIRIS-II and Cri data sets, K are 15. The number of atoms chosen to construct the background dictionary varies for different classes. Thus, we define the parameter η to present the percentage of the chosen atoms who have the top frequencies in each class. For AVIRIS-I, AVIRIS-II, HYDICE, Urban and Cri data sets, it is set to be 5%, 5%, 3%, 1%, and 1%, respectively, to get the best performance. In order to construct the overcomplete potential anomaly dictionary, the number of atoms should be larger than the number of bands. Therefore, the number of potential anomaly dictionary atoms ρ for the experiment data sets is 200, 200, 100, 200, and 100.

The color detection maps are shown in Fig. 4. It can be seen that the Global-RX, CRD, and LRSR can correctly detect the

anomalous pixels in different data sets, but their responses are not strong. Due to the sensitivity to the window size, the Local-RX cannot detect the anomaly in the AVIRIS-I and AVIRIS-II data sets. In addition, the response of anomalies in the other three data sets is not very strong compared with other methods. The PAB-DC method can get the strongest responses of anomalies for all data sets, especially for the AVIRIS-II and Cri data sets. The LRSR method also gets a good performance in these five data sets, but on the AVIRIS-II and Cri data sets, it cannot get strong responses because no anomaly information is used. In contrast, PAB-DC obviously obtains strong responses in these two data sets, and it is benefited from the construction of a potential anomaly dictionary.

In addition, to quantitatively compare the performance of the proposed method with the other four methods, the ROC curves with pointwise confidence intervals are shown in Fig. 5 and the AUC values with upper and lower bounds are given in Table I. The best results for each data set in Table I are highlighted in bold. In Fig. 5, we can see that the Global-RX can obtain stable results for different data sets because of its global characteristic. The CRD method shows a good performance except for Cri data set. It shows advantage compared with the Local-RX method, this is because the CRD method can model the background more exactly, while the multivariate normal distribution-based Local-RX cannot always hold different kinds of background. However, the CRD method also suffers from the sensitivity to the window size, and the local window cannot obtain the global background information, so it cannot get good performance on the Cri data set which has a cluttered background. Since the low-rank model tries to catch the global information of the whole scene, it is very benefit for the simple background. Thus, the LRSR method shows a good performance on the AVIRIS-I, HYDICE, and Urban data sets shown in Fig. 5(a), (c), and (d), respectively. However, it cannot get good performance in the AVIRIS-II and Cri data sets due to their complex background. Our proposed PAB-DC method gets the best performance in the AVIRIS-I, AVIRIS-II, HYDICE, and Cri data sets, as shown in Table I and Fig. 5. It also gets a comparable performance on the Urban data set following the Global-RX method as shown in the fourth line of Table I.

To further reveal the procedure of our PAB-DC method, we illustrate the segmentation map, the background dictionary atoms, the potential anomaly dictionary atoms, the background part, the anomaly part, and the noise part in Fig. 6(a)–(f), respectively. The JSR model can obtain the informative

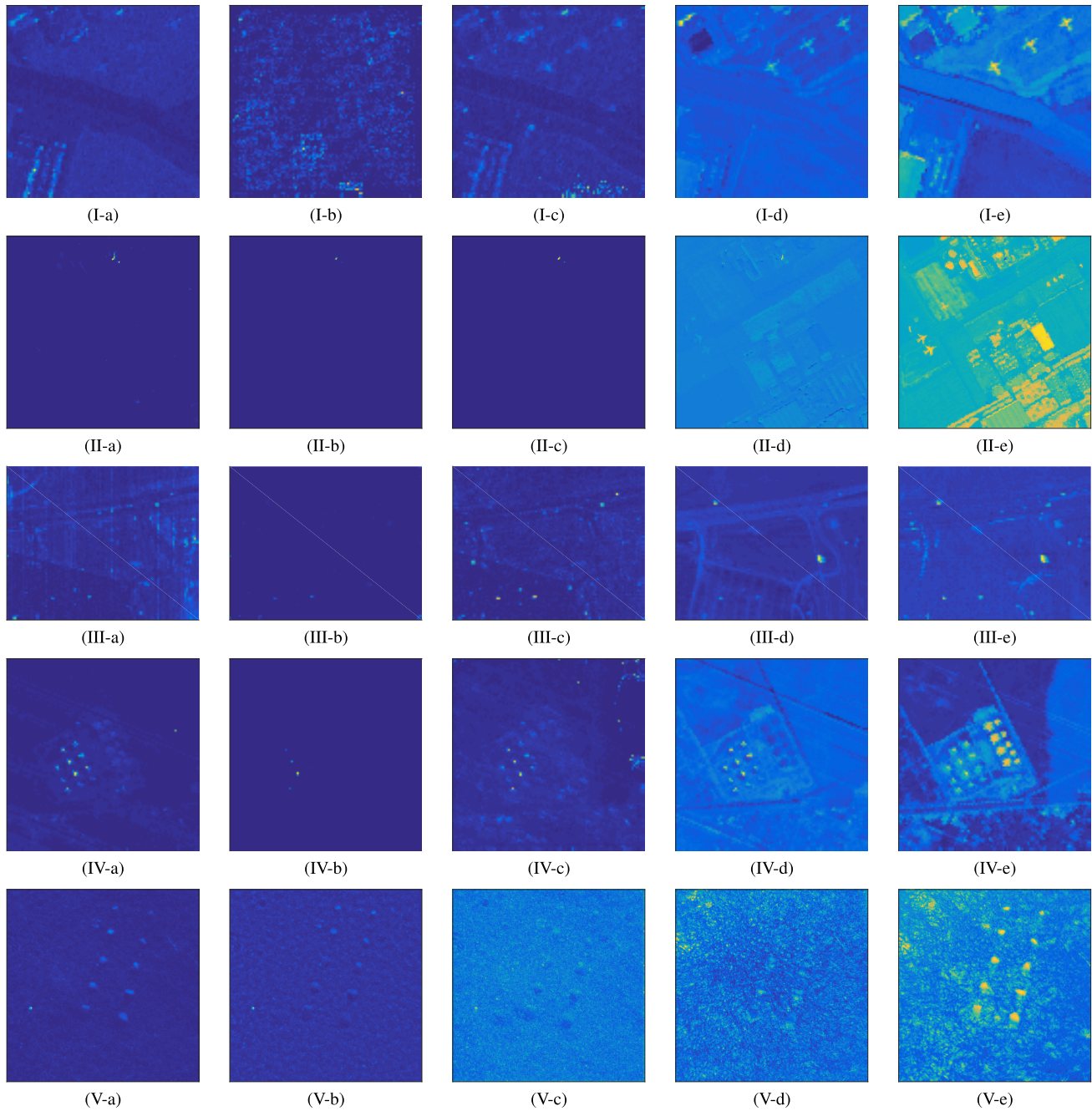


Fig. 4. I is the results using the AVIRIS-I data set. II is the results using the AVIRIS-II data set. III is the results using the HYDICE data set. IV is the results using the Urban data set. V is the results using the Cri data set. (a) Globa-RX. (b) Local CRX. (c) CRD. (d) LRSR. (e) PAB-DC.

background support to construct purer and more reliable background dictionary. Combining with the strong robustness of low-rank constraint, our method can capture the background across different data sets shown in Fig. 6(b) and (d). As shown in Fig. 6(c), in most cases, the potential anomaly dictionary can extract some obviously anomalies in advance, which enhances the performance of subsequent detection tasks. In addition, the procedure of clustering in our PAB-DC method is quite different, which can be seen as a blockwise clustering strategy, resulting in smoother segmentation as shown in Fig. 6(a). It is beneficial to areawise anomaly detection just as in the AVIRIS-I, AVIRIS-II and Cri data sets. Meanwhile, for pixelwise anomaly detection, we can simply set the window

size being 1 as in the HYDICE data sets, and the PAB-DC achieves better results as well. In Fig. 6(d)–(f), the background, anomaly, and noise parts are visualized. The PAB-DC can effectively decompose the original data into these three parts, and we can obviously see that the noise part can capture the strong interference around the edges and cluttered background, which makes the anomaly detection map more smooth.

C. Parameter Analysis

1) *Effectiveness Evaluation of Potential Anomaly Dictionary*: To evaluate the effectiveness of the potential anomaly dictionary, we compare the PAB-DC with the one without it, which is implemented by setting the potential anomaly

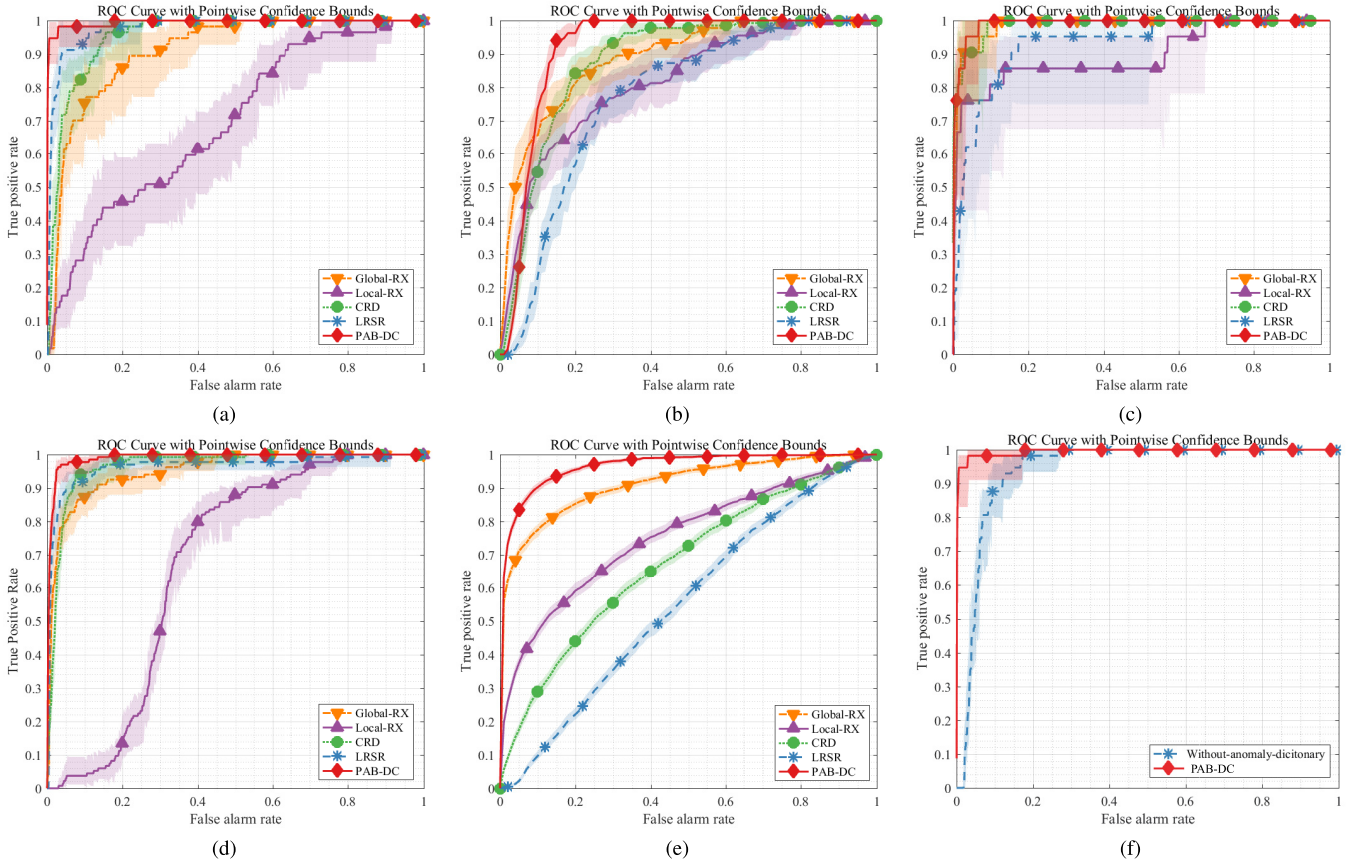


Fig. 5. Experimental results. (a) ROC curves comparison on the AVIRIS-I data set. (b) ROC curves comparison on the AVIRIS-II data set. (c) Curves ROC comparison on the HYDICE data set. (d) ROC curves comparison on the Urban data set. (e) ROC curves comparison on the Cri data set. (f) ROC curves of the PAB-DC method with and without potential anomaly dictionary.

dictionary as a zero vector. For convenience, we conduct the experiments on the AVIRIS-I data set. We set the window size as 3×3 in this experiment. The number of clusters is set as 10 empirically. The number of atoms to construct the background dictionary varies for different classes. The parameter η is 5%. The number of the potential dictionary atoms ρ is slightly greater than that of bands or 0 in the comparison experiment. In addition, we reveal the coefficients obtained in our model to show if the anomaly atoms in the potential anomaly dictionary represent anomalies in the scene.

Fig. 7 shows the coefficients of the chosen anomalies pixels in the potential dictionary, where the red bar is the position of the anomalies detected by the potential anomaly dictionary. The blue bar reflects the coefficients of each atom. We can see that there are five anomalous pixels chosen in the potential anomaly dictionary, two of which have very large coefficients when they are used to reconstruct other anomalies. It is noticed that the coefficients of other pixels in the potential anomaly dictionary are relatively small. It is illustrated that the anomaly part can be well reconstructed using the atoms in the potential anomaly dictionary, and thus, the anomaly part has the ability to identify the anomalous pixels in the scene. Fig. 5(f) shows the ROC curve of these two methods. The AUC value produced by the PAB-DC is 0.9950, while the AUC value of the PAB-DC without the potential anomaly dictionary is 0.9409. From the results, we can see that the

potential anomaly dictionary indeed enhances the detection performance.

2) *Analysis of the Window Size and Cluster Number:* The first parameter that we discuss is the window size used in our experiments. It influences the number of pixels within each region and also affects the construction of two dictionaries by the JSR model. The number of classes is another important parameter, which determines the size of the dictionary. As the increase of the class number, the number of pixels in each class may decrease, which results in the reduction of regions in each class. We evaluate the performance of the PAB-DC considering these two parameters together on the AVIRIS-I data set. The numbers of classes that we choose are 5, 10, 15, 20, and 25 and the window size is 3, 5, 7, and 9, respectively. The parameters β and λ of the objective function is fixed to $\beta = 0.001$ and $\lambda = 0.1$. The results are shown in Fig. 8.

It can be seen that the PAB-DC is robust with the change of the window size, which is different from the other local detectors. Because we group the data on the region level into several classes, it is beneficial to utilizing the relatively global information without affecting the spectral information of the other materials. In addition, for AVIRIS-I, the anomaly is not too large nor too small. However, with the increase of the window size, the computational cost in the procedure of JSR increases. Therefore, in the experiment, a 3×3 window is used.

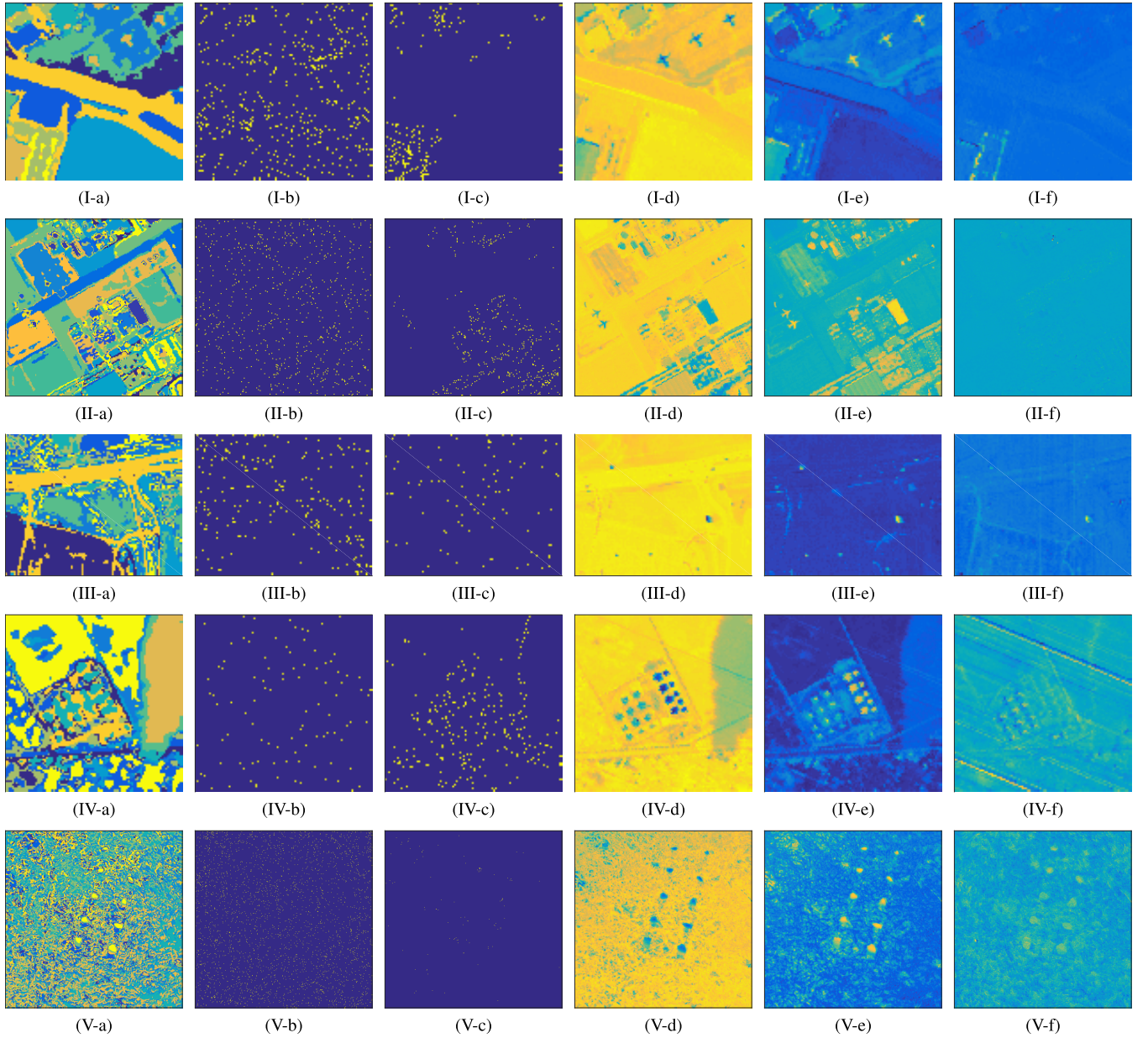


Fig. 6. I is the results using the AVIRIS-I data set. II is the results using the AVIRIS-II data set. III is the results using the HYDICE data set. IV is the results using the Urban data set. V is the results using the Cri data set. (a) Visualization of segmentation map. (b) Visualization of the chosen atoms in the background dictionary. (c) Visualization of the chosen atoms in the potential anomaly dictionary. (d) Visualization of background component. (e) Visualization of anomaly component. (f) Visualization of noise component.

3) *Analysis of Parameters β and λ* : β and λ in the objective function are the parameters to balance the background, anomaly, and noise parts. We analyze their effects on the performance together. Both β and λ are chosen from 0.0001, 0.001, 0.01, 0.1, 1, 2, and 3. In this experiment, the number of the classes is fixed to 10 and the window size is set to 3×3 . The experiment is conducted on the AVIRIS-I data set. The result is shown in Fig. 9, where the x-axis and the y-axis represent λ and β , respectively. We can see that the parameter has slight effects on the AUC ranging from 0.0001 to 3. When $\lambda < 0.01$, the result of the PAB-DC is poor. When it is larger than 0.01, the AUC tends to be stable. Therefore, in our experiment, we choose $\beta = 0.01$ and $\lambda = 0.1$.

4) *Analysis of Parameters of ρ and η* : The parameter η is the percentage of the selected atoms for each class to

construct the background dictionary, and ρ is the number of atoms selected to construct the background dictionary. In our experiment, η is chosen from 0.001, 0.01, 0.05, 0.1, 0.3, and 0.5 and ρ is chosen from 10, 50, 100, 200, 250, 300, and 500. The number of the classes is fixed to 10 and the window size is set to 3×3 . The parameters β and λ are as those suggested earlier. The result is shown in Fig. 10. It can be seen that the PAB-DC is not sensitive to the number of background atoms. Even when 0.1% pixels are chosen, they are sufficient to represent the background due to the low-rank property of the background. When the number of potential anomaly dictionary atoms is too small, the dictionary is not overcomplete, which leads to the fluctuations of the AUC value. With the increase of the number of atoms, the results tend to be stable. However, too many nonanomaly pixels involved in will degrade the

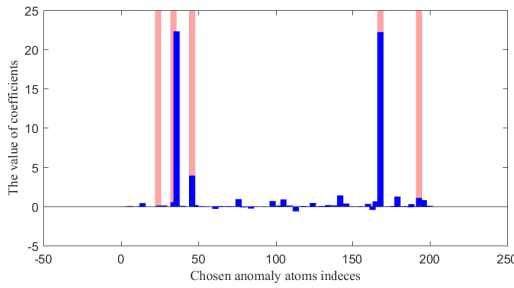


Fig. 7. Coefficients of the chosen anomaly pixels in the potential anomaly dictionary. The red bar reflects the true anomaly pixels chosen in the potential anomaly dictionary. The blue bar presents the value of the coefficient corresponding to each atom in the potential dictionary.

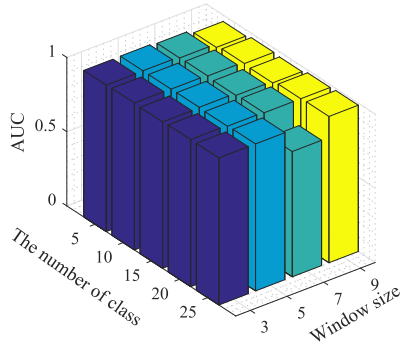


Fig. 8. Analysis of the window size and class number on the AVIRIS-I data set.

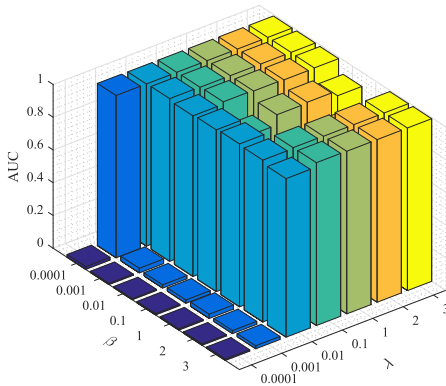


Fig. 9. Analysis of parameters β and λ on the AVIRIS-I data set.

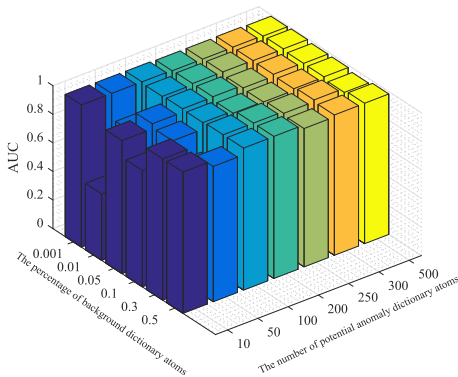


Fig. 10. Analysis of parameters of ρ and η on the AVIRIS-I data set.

effectiveness of the anomaly part in the process. Therefore, we set the number of potential dictionary atoms to be 200, which is slightly greater than the number of bands.

IV. CONCLUSION

In this paper, we have presented a new HSI anomaly detection method based on background and potential anomaly dictionaries utilizing an LRSR strategy, denoted as PAB-DC, where the original data are decomposed into background, anomaly, and noise parts. For the background part, the low-rank representation is used to capture the underlying subspaces of different background materials. For the anomaly part, the sparse representation is utilized to simulate the property of anomalies. In addition, different from other commonly used methods that only consider the background information, we proposed to use both the background and the estimated anomaly spectral information to enhance the performance. The background dictionary is constructed by introducing the JSR model with the grouped overcomplete dictionary for each class. The potential anomaly dictionary is built to catch strong anomalous pixels using the residual computed in the JSR model. The superiority of our proposed PAB-DC method was demonstrated in five real data experiments through the comparison with four standard methods.

ACKNOWLEDGMENT

The authors would like to thank Prof. Q. Du for providing suggestions and discussions.

REFERENCES

- [1] M. Borengasser, W. S. Hungate, and R. Watkins, "Hyperspectral remote sensing: Principles and applications," *Iasri*, vol. 31, no. 12, pp. 1249–1259, Mar. 2007.
- [2] 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.
- [3] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
- [4] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [5] S. Yang, M. Wang, P. Li, L. Jin, B. Wu, and L. Jiao, "Compressive hyperspectral imaging via sparse tensor and nonlinear compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5943–5957, Nov. 2015.
- [6] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.
- [7] Q. Wang, Z. Meng, and X. Li, "Locality adaptive discriminant analysis for spectral-spatial classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2077–2081, Nov. 2017.
- [8] A. Ertürk, M.-D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 708–719, Feb. 2016.
- [9] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. B. Jupp, and J. S. Pearlman, "Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1246–1259, Jun. 2003.
- [10] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.
- [11] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Feb. 2011.
- [12] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 3123–3134, Dec. 2015.
- [13] N. K. Patel, C. Patnaik, S. Dutta, A. M. Shekh, and A. J. Dave, "Study of crop growth parameters using airborne imaging spectrometer data," *Int. J. Remote Sens.*, vol. 22, no. 12, pp. 2401–2411, Aug. 2001.

- [14] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.
- [15] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.
- [16] R. Zhao, B. Du, and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1227–1234, Apr. 2014.
- [17] S. Khazai, A. Safari, B. Mojaradi, and S. Homayouni, "An approach for subpixel anomaly detection in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 769–778, Apr. 2013.
- [18] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.
- [19] M. Rossacci, D. Manolakis, J. Cipar, R. Lockwood, T. Cooley, and J. Jacobson, "Effects of dimensionality reduction on the statistical distribution of hyperspectral backgrounds," *Proc. SPIE*, vol. 6302, p. 63020H, Sep. 2006.
- [20] F. M. Mindrup, T. J. Bihl, and K. W. Bauer, "Modeling noise in a framework to optimize the detection of anomalies in hyperspectral imaging," in *Intelligent Engineering Systems Through Artificial Neural Networks: Smart Systems Engineering: Computational Intelligence in Architecting Complex Engineering Systems*, vol. 20. New York, NY, USA: American Society of Mechanical Engineers, 2010, pp. 517–524.
- [21] J. Y. Chen and I. S. Reed, "A detection algorithm for optical targets in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 1, pp. 46–59, Jan. 2007.
- [22] 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.
- [23] J. P. Williams, T. J. Bihl, and K. W. Bauer, "Towards the mitigation of correlation effects in anomaly detection for hyperspectral imagery," *J. Defence Model. Simul.*, vol. 10, no. 3, pp. 263–273, Feb. 2013.
- [24] 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.
- [25] 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, p. 258, Apr. 2010.
- [26] A. P. Schaum, "Hyperspectral anomaly detection beyond RX," *Proc. SPIE*, vol. 6565, p. 656502, Apr. 2007.
- [27] 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.
- [28] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," *Proc. SPIE*, vol. 6966, pp. 696604-1–696604-12, Apr. 2008.
- [29] 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.
- [30] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.
- [31] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [32] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
- [33] W. Dong, X. Li, D. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 457–464.
- [34] Y. Zhang, B. Du, L. Zhang, and T. Liu, "Joint sparse representation and multitask learning for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 894–906, Feb. 2017.
- [35] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.
- [36] R. Zhao, B. Du, and L. Zhang, "Hyperspectral anomaly detection via a sparsity score estimation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3208–3222, Jun. 2017.
- [37] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.
- [38] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.
- [39] 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.
- [40] S. Jia, X. Zhang, and Q. Li, "Spectral-spatial hyperspectral image classification using $\ell_{1/2}$ regularized low-rank representation and sparse representation-based graph cuts," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2473–2484, Jun. 2015.
- [41] A. Sumarsono and Q. Du, "Low-rank subspace representation for supervised and unsupervised classification of hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4188–4195, Sep. 2016.
- [42] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.
- [43] 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.
- [44] 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.
- [45] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, Mar. 2002.
- [46] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
- [47] H. Pu and B. Wang, "Novel similarity measure-based nonlinear dimensionality reduction methods for hyperspectral imagery," in *Proc. IEEE IGARSS*, Jul. 2013, pp. 414–417.
- [48] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York, NY, USA: Academic, 2014.
- [49] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [50] Z. Lin, M. Chen, and Y. Ma. (Sep. 2010). "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices." [Online]. Available: <https://arxiv.org/abs/1009.5055>
- [51] R. Rubinfeld, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," Dept. Comput. Sci., Israel Inst. Technol., Haifa, Israel, Tech. Rep. CS-2008-08, Jan. 2008.
- [52] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [53] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Apr. 2008.
- [54] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [55] Y. Xing and R. B. Gomez, "Hyperspectral image analysis using ENVI (environment for visualizing images)," *Proc. SPIE*, vol. 4383, pp. 79–87, Jun. 2001.



Ning Huiyan received the B.S. degree from Xidian University, Xi'an, China, in 2015, where he is currently pursuing the Ph.D. degree with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education of China.

His research interests include compressed sensing, low-rank and sparse representation, pattern recognition and machine learning, hyperspectral image classification, and target detection.



Xiangrong Zhang (SM'14) received the B.S. and M.S. degrees from the School of Computer Science, Xidian University, Xi'an, China, in 1999 and 2003, respectively, and the Ph.D. degree from the School of Electronic Engineering, Xidian University, in 2006.

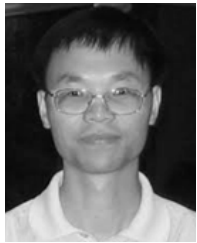
She was a Visiting Scientist with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA, from 2015 to 2016. She is currently a Professor with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of

Education, Xidian University. Her research interests include pattern recognition, machine learning, and remote sensing image analysis and understanding.



Licheng Jiao (SM'89–F'17) received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

He has authored or co-authored over 150 scientific papers. His research interests include signal and image processing, nonlinear circuit and systems theory, learning theory and algorithms, optimization problems, wavelet theory, and data mining.



Huiyu Zhou received the B.Eng. degree in radio technology from the Huazhong University of Science and Technology, Wuhan, China, the M.S. degree in biomedical engineering from the University of Dundee, Dundee, U.K., and the Doctor of Philosophy degree in computer vision from Heriot-Watt University, Edinburgh, U.K.

He is currently a Reader with the Department of Informatics, University of Leicester, Leicester, U.K. He has taken part in the consortiums of a number of research projects in medical image processing,

computer vision, intelligent systems, and data mining.