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 (HSI) 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, respectively. 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 (PAB-DC) can achieve superior results compared with other state-of-the-art methods.**

***Index Terms*—Hyperspectral images, anomaly detection, low rank, joint sparse representation, background dictionary, 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][5][6][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 form background pixels in an image scene [11]. Anomaly detection [12] based on 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 HSI have emerged. As the pioneering work in this area, Reed and Chen [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 (CFAR) detector, based on generalized maximum likelihood ration (GLRT) for multi-dimensional image data. After that, Reed and Yu 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 [22]. 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 (CKRX) [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 and multi-task learning has also been proposed. For anomaly detection, due to the fact that there is no prior information about 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 joint sparse representation (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 homogenous 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

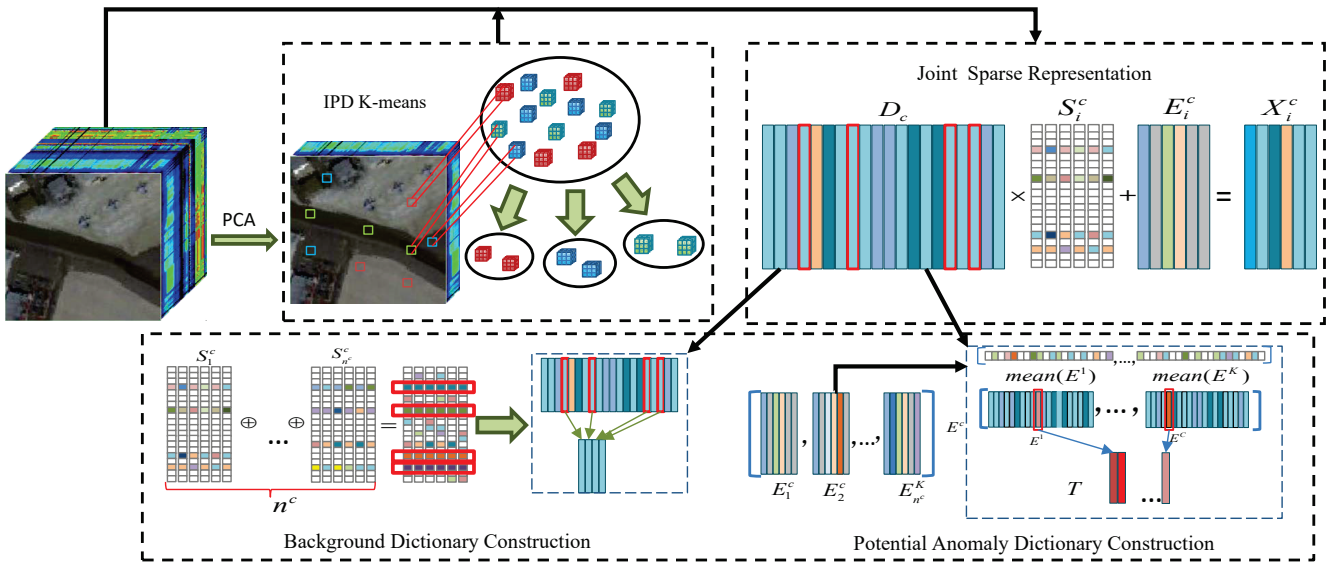

Fig. 1. Schematic illustration of our dictionary construction method for hyperspectral anomaly detection. Firstly the PCA is used to get a low dimensional 3D cube. Then the original image is separate into many 3D regions by a fixed-size window. After that, the IPD based K-means is used to group the regions into several classes. For the original spectral pixels corresponding to each region are reconstructed using 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.

principal component analysis (RPCA) [38]. For better multi-subspace learning, low rank representation (LLR) [39] was proposed. It attempts to minimize the rank of a dictionary with corresponding coefficients whilst 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 whilst the anomalies demonstrate sparse properties. In [43] the observed data was decomposed into background and anomaly parts. The coefficients of the model was 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 achieved promising performance. However, anomaly and noise distributions are aliasing owing to similar sparse characteristics. So, 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 were 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 well-designed background and potential anomaly dictionaries utilizing low rank and sparse representation 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 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 flowchart 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 dictionaries 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 our work 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 3D cube $\mathbb{X}$ into a 2D 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 our work, 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:

$$\mathbf{X} = \mathbf{BZ} + \mathbf{A} + \mathbf{E} \tag{1}$$

where $\mathbf{BZ}$ is the background component, $\mathbf{B} = [b_1, b_2, ..., b_{n_B}]$ is the background dictionary, and n_B is the number of the atoms in the dictionary, $\mathbf{Z} = [z_1, z_2, ..., z_n]$ are the corresponding representation coefficients, and $\mathbf{A} = [a_1, a_2, ..., a_n]$ and $\mathbf{E} = [e_1, e_2, ..., e_n]$ 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. So in our work, $l_{2,1}$ norm is utilized to model the noise. Thus, the objective function can be further formulated as:

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

where $\beta > 0$ and $\lambda > 0$ are the coefficients used to balance all the components. $rank(\mathbf{Z})$ 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.,

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

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 dataset. So a hidden anomaly can be represented by the linear combination of the pre-detected strong anomaly atoms in the dictionary, namely, $\mathbf{A} = \mathbf{TS}$, where $\mathbf{T} = [t_1, t_2, ..., t_{n_T}]$ is the potential anomaly dictionary, n_T is the number of atoms in potential anomaly dictionary and $\mathbf{S} = [s_1, s_2, ..., s_n]$ 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 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 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:

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

where $\|*\|_*$ 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., $r_i = \|a_i\|_2 i = 1, 2, ..., n$. Finally, anomalies can be determined by a pre-defined threshold.

B. Background dictionary construction

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

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

where $\mathbf{U} = [u_1, u_1, ..., u_L]$ is a 2D matrix with L spectral pixels flattened from the 3D local region cube and ψ is the corresponding representation coefficients where only a few rows are non-zero, $\mathbf{R} = [r_1, r_2, ..., r_L]$ are the residuals after construction based on $\mathbf{V}$ and ψ, and $\|*\|_{rows,0}$ denotes the non-zero 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 reflect 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 non-local 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 3D cube $\mathbb{U} \in \mathbb{R}^{win \times win \times d}$ of size $win \times win \times d$ where win is the window size. Its 2D form is $\mathbf{U} \in \mathbb{R}^{\mathbf{d} \times \mathbf{L}}$, where $L = win \times win$. The image patch distance (IPD) [47] is used to measure the distance between two regions, which is defined as

$$\begin{aligned} o_{IPD}(\mathbf{U_P}, \mathbf{U_Q}) &= o(P(u_i), Q(u_j)) \\ &= \sum_{h=1}^{L} \max \begin{pmatrix} \min\limits_{b \in Q(u_j)} o(a_h, b) \\ \min\limits_{a \in P(u_i)} o(b_h, a) \end{pmatrix} \end{aligned} \tag{6}$$

where $o(a, b)$ is a non-local spectral similarity function, and (here it is the Euclidean distance), and $\mathbf{U_P}$ and $\mathbf{U_Q}$ are two 2D matrices representing two regions selected from the regions. The calculation procedure is illustrated in Fig. 2.

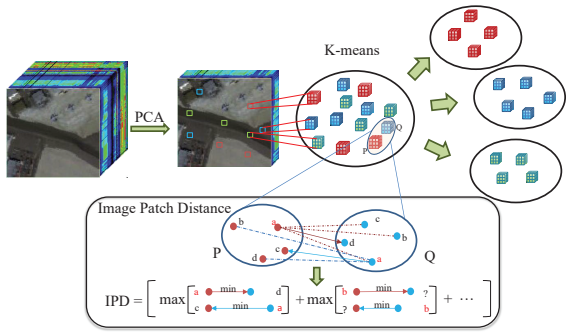

Fig. 2. The procedure of IPD based K-means.

With the increase of 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 non-overcomplete 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. The ultimate formulation of the JSR model for each class is as

$$\begin{aligned} &\min \|\psi_i^c\|_{row,0} \\ &s.t. \mathbf{U}_i^c = \mathbf{G}^c \psi_i^c + \mathbf{R}_i^c \end{aligned} \tag{7}$$

where $\mathbf{U}_i^c$ is the i^{th} region of the c^{th} class, $i = [1, 2, ..., n^c]$, $n^c = [n^1, n^2, ..., n^K]$ and $c = [1, 2, ..., K]$, n^c is the number of the regions in the c^{th} class and K is the number of classes, $\mathbf{G}^c = [g_1^c, g_2^c, ..., g_{n^c}^c]$ 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 our work, the frequency is defined as the sum of the joint sparse representation coefficients after normalization. It can be expressed as:

$$\begin{aligned} P^c &= \frac{1}{\gamma} \sum_{i=1}^{n^c} \sum_{j=1}^{L} \left| \psi_{i,j}^c \right| \\ \gamma &= sum(\sum_{i=1}^{n^c} \sum_{j=1}^{L} \left| \psi_{i,j}^c \right|) \end{aligned} \tag{8}$$

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, and P^c is a vector, in which the value of each element reflects the weighted frequency chosen as the background dictionary atom, $sum(\cdot)$ denotes the element-wise 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, to reduce the influence of noise and which may also have large residuals. So, 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 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:

$$\bar{\mathbf{R}}_i^c = \frac{1}{L} \sum_{j=1}^{L} \left\| r_{i,j}^c \right\|_2 \tag{9}$$

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

$$\mathbf{R}_{mean}^c = [\bar{\mathbf{R}}_1^c, \bar{\mathbf{R}}_2^c, \ldots, \bar{\mathbf{R}}_{n^c}^c] \tag{10}$$

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

$$\begin{aligned} AL &= \frac{1}{\chi}[\mathbf{R}^1_{mean}, \mathbf{R}^2_{mean}, ..., \mathbf{R}^K_{mean}] \\ \chi &= \sum_{c=1}^{K}\sum_{i=1}^{n^c} \bar{\mathbf{R}}^c_i = \frac{1}{L}\sum_{c=1}^{K}\sum_{i=1}^{n^c}\sum_{j=1}^{L} \left\| r^c_{i,j} \right\|_2 \end{aligned} \tag{11}$$

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. So we can eliminate the atoms which 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 show in Eq. (15), 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 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 not always used but with significant coefficients when reconstruct a region containing anomaly pixels 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 issue

$$AW^c = \frac{P^c}{F^c} \tag{12}$$

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

$$F^c = \sum_{i=1}^{n^c}\sum_{j=1}^{L} \mathrm{sgn}(\left|\varphi^c_{i,j}\right|)$$

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

$$\begin{aligned} AL &= \frac{1}{\chi}[\mathbf{R}^1_{mean}, \mathbf{R}^2_{mean}, ..., \mathbf{R}^K_{mean}] \odot \mathbf{AW} \\ \chi &= \sum_{c=1}^{K}\sum_{i=1}^{n^c} \bar{\mathbf{R}}^c_i = \frac{1}{L}\sum_{c=1}^{K}\sum_{i=1}^{n^c}\sum_{j=1}^{L} \left\| r^c_{i,j} \right\|_2 \end{aligned} \tag{13}$$

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

Algorithm 1

Input: Data matrix 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

$$\begin{aligned} \mathbf{Z} := (\mathbf{B}^T\mathbf{B} + \mathbf{I})^{-1}[&\mathbf{B}^T\mathbf{X} - \mathbf{B}^T\mathbf{T}\mathbf{S} - \mathbf{B}^T\mathbf{E} \\ &+ \mathbf{J} + (\mathbf{B^T}\mathbf{Y_1} - \mathbf{Y_2})/\mu] \end{aligned}$$

6. Fix others and update S by

$$\begin{aligned} \mathbf{S} := (\mathbf{T}^T\mathbf{T} + \mathbf{I})^{-1}(&\mathbf{T}^T\mathbf{X} - \mathbf{T}^T\mathbf{B}\mathbf{Z} - \mathbf{T}^T\mathbf{E} \\ &+ \mathbf{L} + (\mathbf{T}^T\mathbf{Y}_1 - \mathbf{Y}_3)/\mu) \end{aligned}$$

7. Update the three Lagrange multipliers

$$\begin{aligned} \mathbf{Y_1} &:= \mathbf{Y_1} + \mu(\mathbf{X} - \mathbf{B}\mathbf{Z} - \mathbf{E} - \mathbf{D}\mathbf{S}) \\ \mathbf{Y_2} &:= \mathbf{Y_2} + \mu(\mathbf{Z} - \mathbf{J}) \\ \mathbf{Y_3} &:= \mathbf{Y_3} + \mu(\mathbf{S} - \mathbf{L}) \end{aligned}$$

8. Update the parameter μ,

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

9. Check the convergence conditions

$$\begin{aligned} &\|\mathbf{X} - \mathbf{B}\mathbf{Z} - \mathbf{D}\mathbf{S}\|_F < \varepsilon \\ &\|\mathbf{Z} - \mathbf{J}\|_F < \varepsilon \\ &\|\mathbf{S} - \mathbf{L}\|_F < \varepsilon \end{aligned}$$

10. End While

D. Optimization and computational complexity

To solve the problem shown in Eq. (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:

$$\begin{aligned} &\min_{J,E,Z,S,L} \|\mathbf{J}\|_* + \beta\|\mathbf{L}\|_1 + \lambda\|\mathbf{E}\|_{2,1} \\ &s.t.\ \ \mathbf{X} = \mathbf{B}\mathbf{Z} + \mathbf{T}\mathbf{S} + \mathbf{E}, \mathbf{Z} = \mathbf{J}, \mathbf{S} = \mathbf{L} \end{aligned} \tag{14}$$

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

$$\begin{aligned}\ell &= \|\mathbf{J}\|_* + \lambda\|\mathbf{E}\|_{2,1} + \beta\|\mathbf{L}\|_1 + \\ &< \mathbf{Y_1}, \mathbf{X} - \mathbf{BZ} - \mathbf{E} - \mathbf{TS} > + < \mathbf{Y_2}, \mathbf{Z} - \mathbf{J} > \\ &+ < \mathbf{Y_3}, \mathbf{S} - \mathbf{L} > + \frac{\mu}{2}(\|\mathbf{X} - \mathbf{BZ} - \mathbf{E} - \mathbf{TS}\|_F^2 \\ &+ \|\mathbf{Z} - \mathbf{J}\|_F^2 + \|\mathbf{S} - \mathbf{L}\|_F^2)\end{aligned} \tag{15}$$

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:

$$\min_J \|\mathbf{J}\|_* + \frac{\mu}{2}\left\|\mathbf{J} - (\mathbf{Z} + \frac{\mathbf{Y_2}}{\mu})\right\|_F^2 \tag{16}$$

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

$$\min_E \lambda\|\mathbf{E}\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{E} - (\mathbf{X} - \mathbf{BZ} - \mathbf{TS} + \frac{\mathbf{Y_1}}{\mu})\right\|_F^2 \tag{17}$$

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

$$\min_L \beta\|\mathbf{L}\|_1 + \frac{\mu}{2}\left\|\mathbf{L} - (\mathbf{S} + \frac{\mathbf{Y_3}}{\mu})\right\|_F^2 \tag{18}$$

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

The computation of our method include those of Algorithm 1 and dictionary construction. For the first aspects, the major computation is Step 2, which requires computing the SVD of an $n \times n$ matrix. So 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 Eq. (4) always lies within the subspace spanned by the rows of 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 $\mathbf{B^T}$. So our model shown in Eq. (4) can be rewrite as:

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

where $\mathbf{B}^* = \mathbf{BP}^*$. Since the number of rows of $\tilde{Z}$ is at most r_b (the rank of matrix $\mathbf{B}$ $r_b << n$), so 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 Step 3 and 4 are $O(dnr_B)$. For the dictionary $\mathbf{B}$ and potential anomaly dictionary $\mathbf{T}$, their column numbers are n_B and n_T respectively. In Step 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 our work, the OMP-Cholesky based method [51] is used to solve Eq. (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-kmeans is $O(n_k(d_rLKn))$, 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 $(n_k(d_rLKn) + nk(2ndL + nL^2 + 2n + k^2) + n_s(r_B^3 + dnr_B + nr_B^2 + nn_T^2))$.

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

Algorithm 2

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

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

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

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 ;

$$\begin{aligned}\mathbf{U} = [\mathbf{U}^1, \mathbf{U}^2, ..., \mathbf{U}^K] = [\mathbf{U}_1^1, \mathbf{U}_2^1, ...\mathbf{U}_{n^1}^1] \cup \\ [\mathbf{U}_1^2, \mathbf{U}_2^2, ...\mathbf{U}_{n^2}^2] \cup ... \cup [\mathbf{U}_1^K, \mathbf{U}_2^K, ...\mathbf{U}_{n^K}^K]\end{aligned}$$

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 I**

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 low rank and sparse representation based detector (LRSR) [43] on five real HSI datasets. The CRD method was proposed based on the assumption that the background can be approximately represented by its spatial neighborhoods, while anomalies

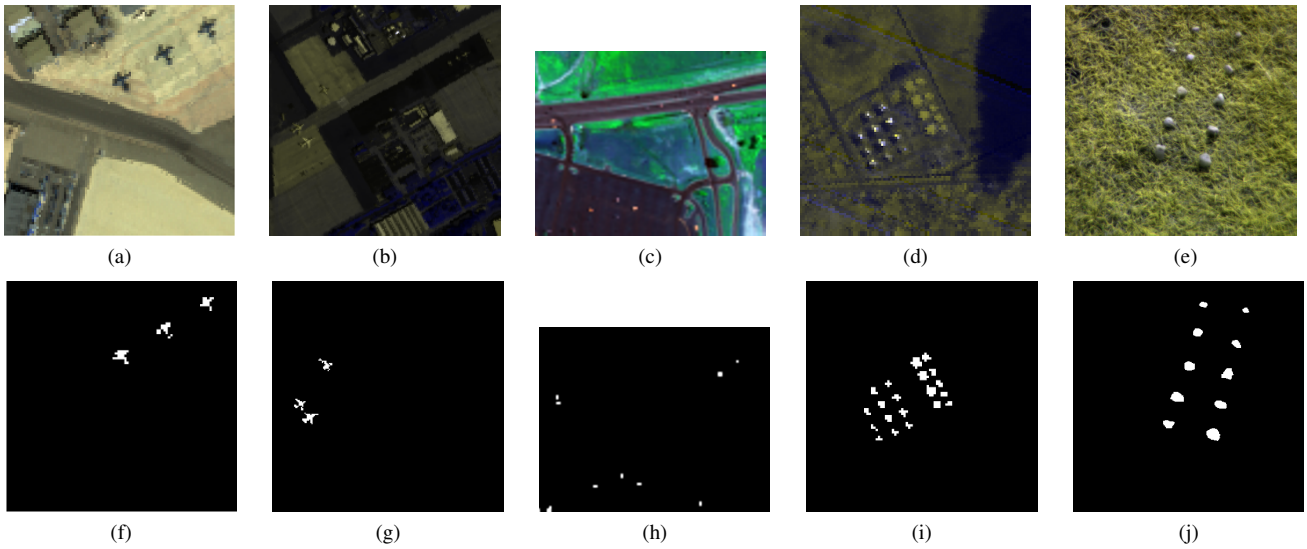

Fig. 3. Image descriptions: (a) The false color image of AVIRIS-I; (b) The false color image of AVIRIS-II; (c) The false color image of the HYDICE dataset; (d) The false color image of the Urban dataset; (e) The false color image of Cri dataset; (f) The groundtruth of AVIRIS-I; (g) The groundtruth of AVIRIS-II; The groundtruth of HYDICE dataset; The groundtruth of Urban dataset; The groundtruth of Cri dataset.

cannot. The LRSR method using both low rank and sparse to constrain the background. 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 2.40GHz $\times$ 32 and 64 GB RAM.

A color detection map is provided to illustrate the results. The colors ranges from dark blue to bright yellow which reflects 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 which is a 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. Dataset description

The first dataset was provided by [43]. It was collected by the Airborne Visible/Infrared Imaging Spectrometer over San Diego, CA, USA (AVIRIS). 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 dataset has an image size of 400×400 . There are two kinds of airplanes in the scene that are treated as anomaly. From up-left of this hyperspectral dataset, a region of 100×100 pixels selected as AVIRIS-I for testing. The three airplanes are regarded as 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) show the false color image and the groundtruth map of AVIRIS-I dataset, respectively.

The second dataset AVIRIS-II is a 200×200 area selected from the AVIRIS image, which is located at the center of San Diego region. Compared to AVIRIS-I this dataset 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 groundtruth map are shown in Fig.3 (b) and (g), respectively.

The third dataset 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 10nm and a spatial resolution of 1m. It covers an urban area that comprises a vegetation area, a construction area, and several roads including some vehicles. The whole dataset 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 groundtruth map are shown in Fig.3 (c) and (h).

The fourth dataset is from an open Airport-Beach-Urban dataset [54]. The sample images in this dataset were manually extracted from large images downloaded from the AVIRIS Web site. We use the dataset with the size of 100×100 as shown in Fig.3 (d) and (i). It was collected over Texas Coast in 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 (ENVI) software [55].

The fifth dataset was acquired by the Nuance Cri hyperspectral sensor. The spectral resolution of this dataset is 10nm. The image scene covers an area of 400×400 pixels, with 46 spectral bands in the wavelengths ranging from 650 to 1100nm. The ten 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 HYDICE dataset 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 datasets. The parameters β and λ in our model are 0.01 and 0.1, which is quite stable across different datasets except for Cri dataset where λ is 10 to get the best performance. The window size in PAB-DC is 1×1 for HYDICE dataset because the anomalies presented in it are isolated pixels while it set to be 3×3 for the other datasets empirically. The number of classes K is determined by the number of background materials in different datasets. So, for Urban dataset K is 5, for AVIRIS-I and HYDICE datasets K are 10 and for AVIRIS-II and Cri dataset K are 15. The number of atoms chosen to construct the background dictionary varies for different classes. So, 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 datasets it is set to be 5%, 5%, 3%, 1%, 1% respectively to get the best performance. In order to construct the over-complete potential anomaly dictionary, the number of atoms should be larger than the number of bands. So, the number of potential anomaly dictionary atoms ρ for the experiment datasets are 200, 200, 100, 200, 100 respectively.

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

Additionally, to quantitatively compare the performance of the proposed method with the other four methods, the ROC curves with point-wise 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 dataset in Table I are highlighted in bold. In Fig. 5 we can see that the Global-RX can obtain stable results for different datasets because of its global characteristic. The CRD method shows good performance except for Cri dataset. It shows advantage compared with Local-RX method, this is because that, the CRD method can model the background more exactly, while the multivariate normal distribution based Local-RX cannot always hold different kinds 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 dataset which has a cluttered background. Since the low rank model tries to catch the global information of whole scene, it is very benefit for the simple background. So, the LRSR method shows good performance on the AVIRIS-I, HYDICE and Urban datasets illustrated in Fig. 5 (a) (c) and (d). However, it cannot get good performance in AVIRIS-II and Cri datasets due to their complex background. Our proposed PAB-DC method gets the best performance in AVIRIS-I, AVIRIS-II, HYDICE and Cri datasets as shown in Table I and Fig. 5. It also gets a comparable performance on Urban dataset following Global-RX method as show 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) (b) (c) (d) and (e). The JSR model can obtain the informative 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 dataset illustrated 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 block-wise clustering strategy, resulting in smoother segmentation as show in Fig. 6 (a). It is beneficial to area-wise anomaly detection just as in AVIRS-I, AVIRS-II and Cri datasets. Meanwhile, for pixel-wise anomaly detection we can simply set the window size being 1 as in HYDICE datasets, and the PAB-DC achieves better results as well. In 6 (d) (e) and (f), the background, anomaly and noise parts are visualized. The PAB-DC can effectively decomposes 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 dictionary as a zero vector. For convenience, we conduct the experiments on the AVIRIS-I dataset. 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

TABLE I
THE AUC COMPARISON OF THE METHODS WITH CONFIDENCE INTERVAL

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

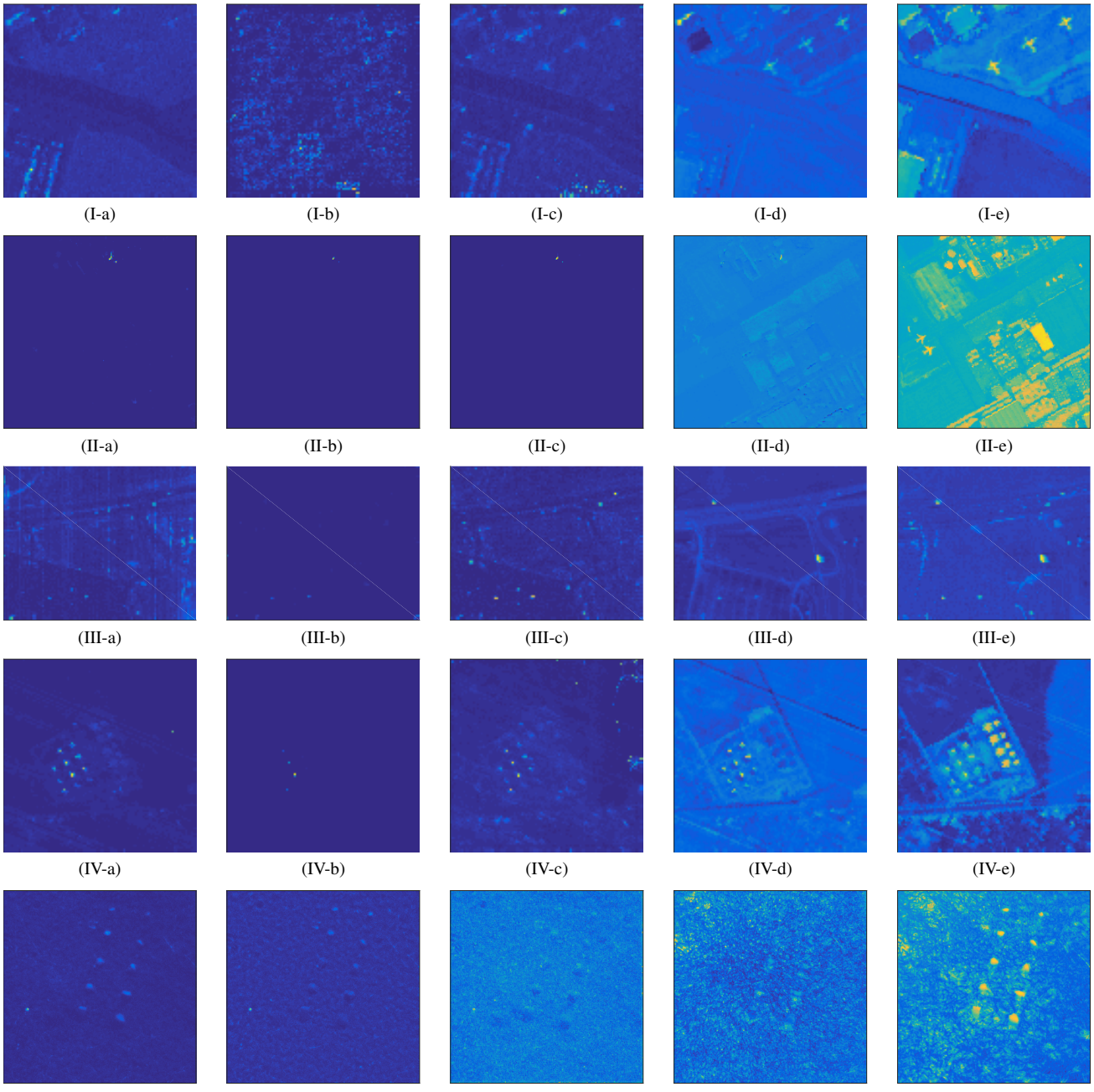

Fig. 4. I is the results using AVIRIS-I dataset; II is the results using AVIRIS-II dataset; III is the results using HYDICE dataset; IV is the results using Urban dataset; V is the results using Cri dataset. (a) Globa-RX (b)local CRX (c) CRD (d) LRSR (e) PAB-DC

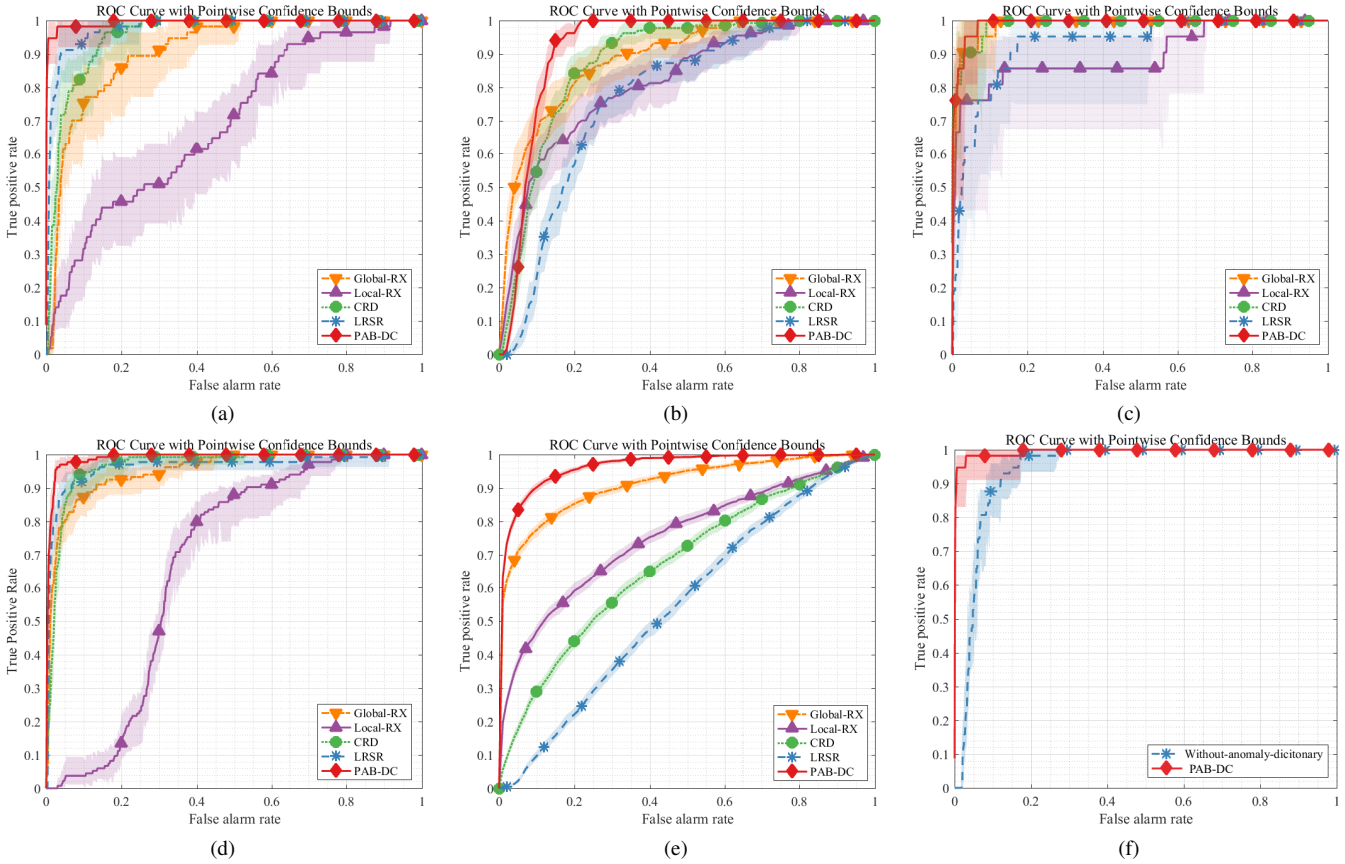

Fig. 5. Experimental results: (a) The ROC curves comparison on the AVIRIS-I dataset; (b) The ROC curves comparison on the AVIRIS-II dataset; (c) The curves ROC comparison on the HYDICE dataset; (d) The ROC curves compaision on Urban dataset; (e) The ROC curves comparison on the Cri dataset; (f) The ROC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

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, 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 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 dataset. The numbers of classes we chose are 5, 10, 15, 20, 25 and the window size is 3, 5, 7, 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. So, in the experiment, a 3×3 windows is used.

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, 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 dataset. The result is shown in Fig. 9, where x-axis and y-axis represent λ and β, respectively.

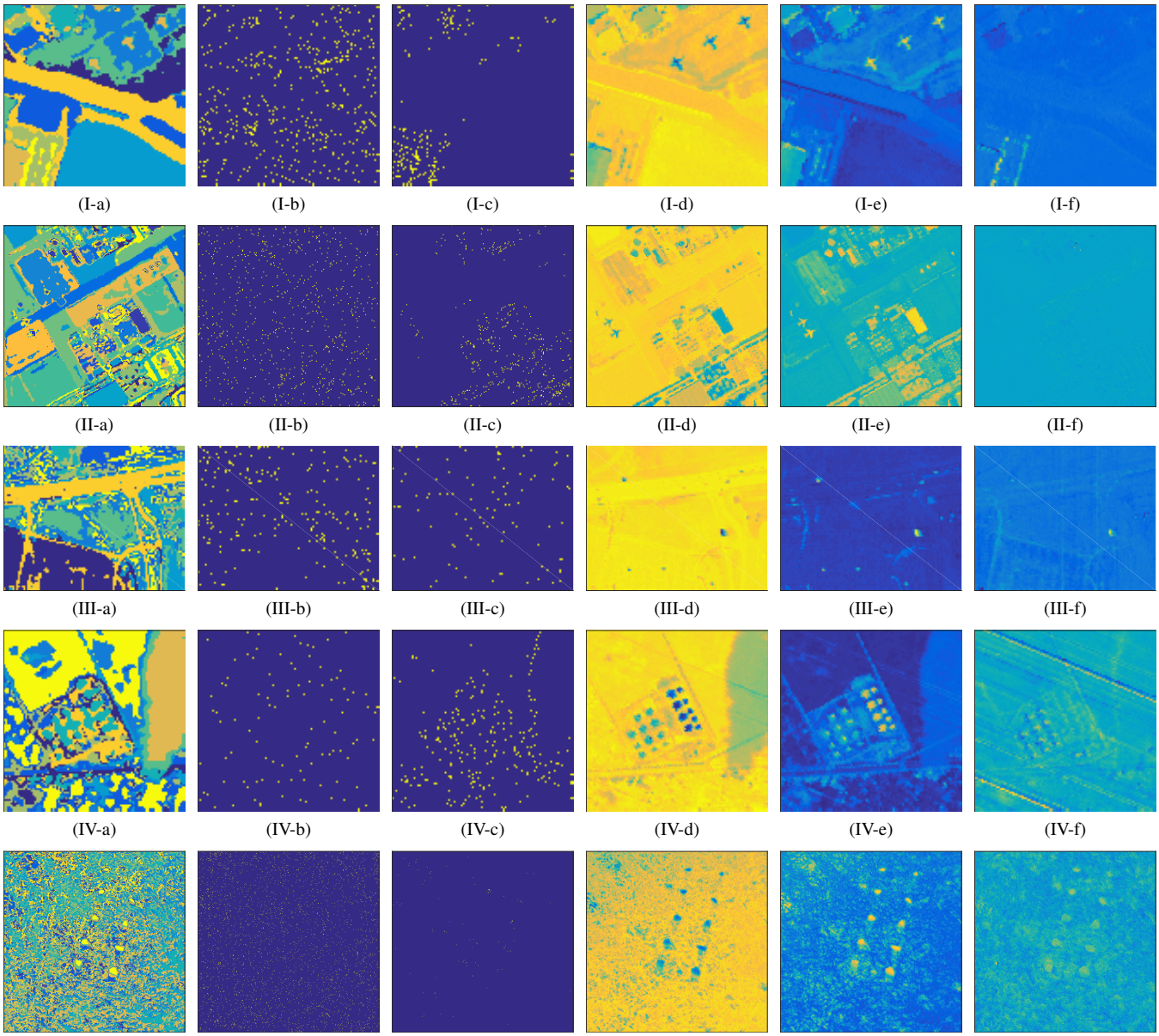

Fig. 6. I is the results using AVIRIS-I dataset; II is the results using AVIRIS-II dataset; III is the results using HYDICE dataset; IV is the results using Urban dataset; V is the results using Cri dataset. (a) The visualization of segmentation map. (b) The visualization of the chosen atoms in the background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

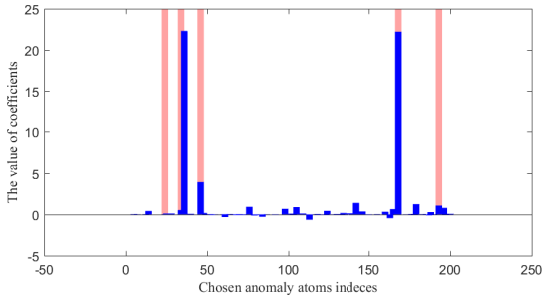

Fig. 7. The coefficients of the chosen anomaly pixels in the potential anomaly dictionary. The red bar reflect the true anomaly pixels chosen in the potential anomaly dictionary. The blue bar presents the value of the coefficient corresponding to each atom in the potential dictionary.

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. So, 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, 0.5 and ρ is chosen from 10, 50, 100, 200, 250, 300, 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

above. 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 fluctuations of the AUC value. With the increase of the number of atoms, the results tend to be stable. However, too many non-anomaly pixels involved in will degrade the effectiveness of the anomaly part in the process. So we set the number of potential dictionary atoms to be 200, which is slightly greater than the number of bands.

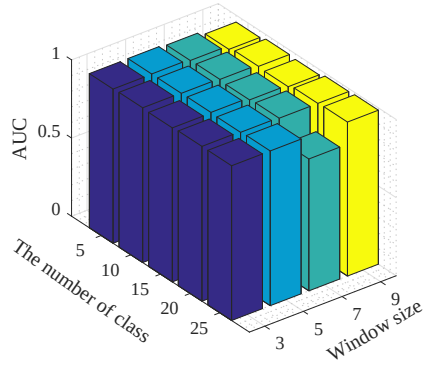

Fig. 8. Analysis of the window size and class number on the AVIRIS-I dataset.

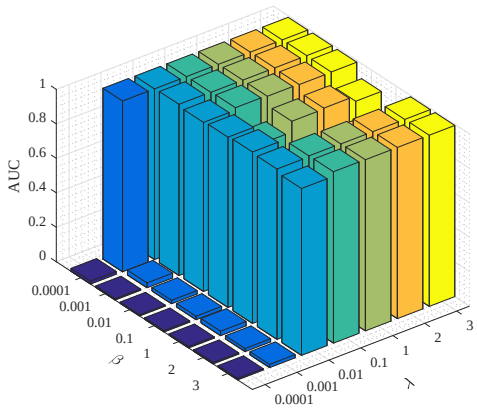

Fig. 9. Analysis of parameters β and λ on the AVIRIS-I dataset.

IV. Conclusion

In this paper, we have presented a new hyperspectral image anomaly detection method based on background and potential anomaly dictionaries utilizing low rank and sparse representation strategy, denoted as PAB-DC, where the original data is decomposed into background, anomaly and noise parts. For the background part, the low rank representation is used to capture underlying subspaces of different background materials. For the anomaly part, the sparse representation is utilized to simulate the property of anomalies. In addition,

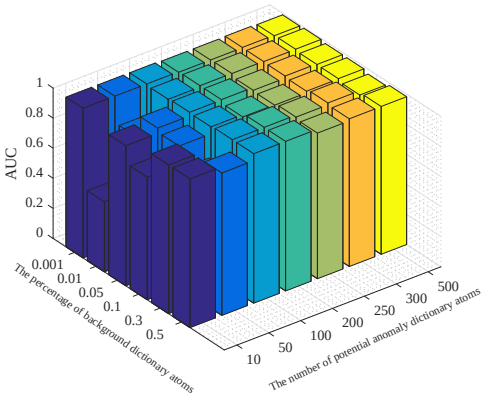

Fig. 10. Analysis of parameters of ρ and η on the AVIRIS-I dataset..

different from other commonly used methods which only consider the background information, we proposed to use both background and estimated anomaly spectral information to enhance the performance. The background dictionary is constructed by introducing the JSR model with the grouped over-complete 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 were demonstrated in five real data experiments through the comparison with four standard methods.

V. Acknowledgment

The authors would like to thank Prof. Qian Du for providing suggestions and discussions. This work was supported in part by the National Natural Science Foundation of China (nos. 61772400, 61501353, 61772399, 91438201, 61573267). H. Zhou is supported by UK EPSRC under Grants EP/N508664/1, EP/R007187/1 and EP/N011074/1, and Royal Society-Newton Advanced Fellowship under Grant NA160342.

References

[1] M. Borengasser, W. S. Hungate, and R. Watkins, "Hyperspectral remote sensing: Principles and applications," *Iasri.res.in*, vol. 31, no. 12, p. 1249C1259, Mar 2007.

[2] D. W. J. Stein, S. G. Beaven, L. E. Hoff, and E. M. Winter, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process Mag.*, vol. 19, no. 1, pp. 58–69, June 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, June 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, June 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. Ertrk, M. D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *IEEE J Sel Top Appl Earth Obs Remote Sens.*, vol. 9, no. 2, pp. 708–719, Sept 2017.

[9] B. Datt, T. R. Mcvicar, T. G. Van Niel, and D. L. B. Jupp, "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, Aug 2003.

[10] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln laboratory journal*, vol. 14, no. 1, pp. 79–116, Jan 2003.

[11] 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, May 2011.

[12] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection." *IEEE Trans Cybern*, vol. 46, no. 12, 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, Nov 2011.

[16] R. Zhao, B. Du, and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J Sel Top Appl Earth Obs 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 Top Appl Earth Obs Remote Sens.*, vol. 6, no. 2, pp. 769–778, Sept 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, Aug 2002.

[19] M. Rossacci, D. Manolakis, J. Cipar, R. Lockwood, T. Cooley, and J. Jacobson, "Effects of dimensionality reduction on the statististical distribution of hyperspectral backgrounds," in *Proc. SPIE*, 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," *Intelligent Engineering Systems through Artificial Neural Networks: Computational Intelligence in Architecting Complex Engineering Systems*, vol. 20, pp. 517–524, 2010.

[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. Def. 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 Top Appl Earth Obs Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Apr 2014.

[25] S. Matteoli and M. Diani, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Opt. Eng.*, vol. 49, no. 4, pp. 258–258, Apr 2010.

[26] A. P. Schaum, "Hyperspectral anomaly detection beyond rx," *Proc. SPIE*, vol. 6565, 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, Jan 2005.

[28] N. M. Nasrabadi, "Regularization for spectral matched filter and rx anomaly detector," *Proc. SPIE*, vol. 6966, pp. 696 604–696 604–12, 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, July 2016.

[30] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer Publishing Company, Incorporated, Jan 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, Apr 2009.

[32] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.(2008)*, Aug 2008, pp. 1–8.

[33] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.(2011)*. IEEE, Aug 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, Dec 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 Top Appl Earth Obs Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, June 2015.

[36] R. Zhao, B. Du, and L. Zhang, "Hyperspectral anomaly detection via a sparsity score estimation framework," *IEEE Trans Geosci Remote Sens.*, Feb 2017.

[37] E. J. Candes and Y. Plan, "Matrix completion with noise," *IEEE Proc.*, vol. 98, no. 6, pp. 925–936, Apr 2010.

[38] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. Assoc. Comput. Mach.*, vol. 58, no. 3, p. 11, Dec 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, Apr 2013.

[40] S. Jia, X. Zhang, and Q. Li, "Spectral–spatial hyperspectral image classification using $l_{1/2}$ regularized low-rank representation and sparse representation-based graph cuts," *IEEE J Sel Top Appl Earth Obs Remote Sens.*, vol. 8, no. 6, pp. 2473–2484, July 2015.

[41] A. Sumarsono and Q. Du, "Low-rank subspace representation for supervised and unsupervised classification of hyperspectral imagery," *IEEE J Sel Top Appl Earth Obs Remote Sens.*, vol. 9, no. 9, pp. 4188–4195, May 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, May 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, Nov 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, Oct 2016.

[45] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, PhD thesis, Stanford University, 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 imgery," in *Proc. IEEE IGARSS (2013)*. IEEE, Jan 2013, pp. 414–417.

[48] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*. Academic press, 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, Oct 2010.

[50] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, Sept 2010.

[51] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *Cs Technion*, vol. 40, Jan 2008.

[52] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans Geosci Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Aug 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, July 2017.

[55] Y. Xing and R. B. Gomez, "Hyperspectral image analysis using envi (environment for visualizing images)," in *Geo-Spatial Image and Data Exploitation II*, vol. 4383. International Society for Optics and Photonics, June 2001, pp. 79–87.