

# Sparse Representation Based Hyperspectral Anomaly Detection via Adaptive Background Sub-Dictionaries

YI LU<sup>1</sup> AND SHUCAI HUANG<sup>2</sup>

<sup>1</sup>Graduate School, Air Force Engineering University, Xi'an 710000, China

<sup>2</sup>Air Defense and Anti-Missile College, Air Force Engineering University, Xi'an 710000, China


Corresponding author: Yi Lu (frankluyitougao@163.com)

**ABSTRACT** Hyperspectral anomaly detection has drawn much attention in recent years. In this paper, in order to effectively extract anomalies in hyperspectral images, a novel sparse-representation based hyperspectral anomaly detection method via adaptive background sub-dictionaries is proposed. Firstly, a background estimation strategy is proposed to provide representative background information. Based on the estimated background, a global dictionary is constructed by utilizing K-means clustering algorithm. Next, Several active atoms are selected from the global dictionary to form a sub-dictionary to adaptively approximate the local region in each dual-window. This sub-dictionary construction strategy can remove potential anomaly contamination in local regions. Finally, a re-weighting strategy is proposed to enhance the performance of sparse-representation-based anomaly detector. Experimental results demonstrate that our method can effectively extract anomalies and suppress background simultaneously.

**INDEX TERMS** Anomaly detection, hyperspectral imagery, sparse representation, background estimation, endmember extraction.

## I. INTRODUCTION

Hyperspectral images (HSIs) are capable to contain abundant spectral characteristics of ground materials [1]. They consist of hundreds or even thousands of continuous and narrow spectral bands ranged from 0.4-2.5  $\mu\text{m}$  and each band is approximately 0.01  $\mu\text{m}$  wide [2]. Owing to the high resolution in spectral dimension, different objects can be recognized and distinguished according to their spectral signatures via hyperspectral images. Upon this basis, HSIs have been employed in various tasks that require identification of objects such as target detection. In terms of whether prior information is required, hyperspectral target detection can be categorized into two types: supervised and unsupervised. Due to small spatial size of targets and unpredictable atmosphere factors, it is usually hard to obtain the spectral information of targets. Therefore, unsupervised target detection, known as anomaly detection, is more commonly researched in practical and has drawn much attention with state-of-the-art techniques, such as compressive sensing [3] and deep learning [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Zhao .

Anomalies in HSIs refer to objects that occupy a few pixels (even subpixels in some situations). They have significantly distinct spectral characteristics from neighboring regions. Over the last few decades, a quantity of anomaly detection methods have been proposed. The well-known Reed-Xiaoli (RX) algorithm [5] exploits the assumption that the background follows a multivariate normal distribution. It measures the Mahalanobis distance between the spectrum vectors of the test pixel and the background pixels as detection results. Local RX and Global RX are studied respectively according to different means of background estimation. However, the performance of RX algorithm is unstable as it essentially depends on the estimated background covariance matrix. Moreover, the assumption of background distribution is not in accordance with the fact that the background in real-world HSI is much more complicated. To address these issues, quite an amount of RX-based algorithms have been developed. The regularized RX algorithm [6] aims to attenuate the ill conditioning of the matrix inversion by regularizing the background covariance matrix. Aiming at decreasing anomaly contamination in background statistics, the weighted-RX algorithm [7] estimates the Gaussian probability as weight vectors. In order to effectively separate

anomaly pixels from the background, the kernel RX algorithm [8] projects the HSI dataset into a higher dimensional feature space. The subspace-based RX is introduced in [9], which explores the background features via the representative eigenvectors of the covariance matrix.

In recent years, with the development of compressed sensing theory, representation based techniques have emerged as a hot topic in many application fields, such as anomaly detection [3], face recognition [10], image classification [11], image denoising [12], and so on. Sparse representation (SR) based HSI anomaly detectors assume that a background pixel can be linearly represented with only a few coefficients over a background dictionary while an anomaly pixel can not. Li *et al.* [13] select the most representative background elements to adaptively approximate local regions, thus false alarm rate can be effectively reduced. Aiming at reducing anomaly contamination in the background dictionary, Zhu *et al.* [14] construct a background dictionary via extracted background endmembers. A sparsity score estimation framework is proposed in [15] to provide a novel view for HSI anomaly detection. The atom usage probability (AUP) score is used to assess reconstruction energy of dictionary atoms, which helps enhancing the discriminative power of the background dictionary. Low-rank representation (LRR) based methods also play a vital role in HSI anomaly detection. In LRR model, HSI data is assumed drawn from multiple subspaces. Based on this assumption, the background part and the anomaly part are able to be separated by a background dictionary. The anomaly detector based on low-rank and sparse representation introduced in [16] employs LRR model to obtain the sparse anomaly component. The  $l_2$ -norm is then applied to columns in the sparse matrix to locate anomaly pixels. Wang *et al.* [17] form a background dictionary with the material signature matrix for the LRR model to extract the background information to identify the anomaly components. As one of the significant characteristic of HSI, Tan *et al.* [18] analyze the spatial similarity among pixels in local regions and impose a spatial constraint to improve the detection performance with LRR model. Collaborative representation (CR) technique also has an outstanding performance in detecting anomalies in HSIs. The collaborative representation based detectors (CRD) adopt a sliding dual-window strategy and consider that the central test pixel lies in the subspace spanned by neighboring pixels in the outer window. The detection criterion is the reconstruction error of the test pixel. Li *et al.* [19] introduce a distance-weighted Tikhonov regularization to the CRD optimization procedure and then project the detector into a higher dimension by the kernel trick. For the aim to eliminate the influence by potential anomalies in the outer window, Li *et al.* [20] develop a principal-component-analysis (PCA) based method to remove the outliers in neighboring regions. Wu *et al.* [21] combine LRR model with CRD to achieve a more effective separation between background component and sparse anomaly component. The aforementioned methods focus on obtaining a promising background estimation, which is further used to

extract anomalies. Although they design various strategies to extract pure background information, the detection results yet suffer from serious false alarms. This is attributed to anomaly contamination and lack of complete background information. Therefore, estimating a pure background without anomaly contamination still remains a challenge. Especially, the accuracy of the representation-based detectors essentially relies on the quality of estimated background, i.e. the quality of the constructed background dictionary. Generally, a desirable background dictionary is expected to be immune from anomaly contamination and to contain as abundant background information as possible.

In this paper, inspired by the work of Zhu *et al.* [14], and from the perspective of dictionary construction for SR, we propose a novel hyperspectral anomaly detection method based on adaptive background sub-dictionaries. The main contributions of this paper can be summarized as follows:

1. An SMACC endmember extraction model based background estimation strategy is proposed so that representative and pure background information can be extracted.
2. Based on the estimated background, a global dictionary is constructed by utilizing K-means clustering algorithm. Several active atoms are selected from this global dictionary to form a sub-dictionary. The local region in each dual-window can be adaptively approximated by this sub-dictionary.
3. With the sub-dictionaries, a re-weighting strategy based on spectral angle distance is proposed to enhance the performance of SR based anomaly detector.

The remainder of this paper is organized as follows. In Section 2, the basic theories of SR based anomaly detector and SMACC endmember extraction model are briefly reviewed. In Section 3, the proposed hyperspectral anomaly detection method is demonstrated in detail. In Section 4, with the experiments on real HSI datasets, the effectiveness of the proposed method is evaluated and the proposed strategies and parameters are further discussed. In Section 5, we draw the conclusions.

## II. RELATED WORKS

### A. SPARSE REPRESENTATION FOR ANOMALY DETECTION

The basic idea of SR based anomaly detection is to represent the test pixel with the linear combination of the background dictionary atoms. It assumes that if a pixel belongs to the background class, it lies in the subspace spanned by the background dictionary atoms. Given a reshaped HSI dataset denoted as  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{B \times N}$  where  $B$  is the number of spectral bands and  $N$  is the number of pixels. The SR model for each pixel  $\mathbf{x}_i$  ( $1 \leq i \leq N$ ) can be expressed as

$$\mathbf{x}_i = \mathbf{D}\alpha_i = \alpha_{i1}\mathbf{d}_1 + \alpha_{i2}\mathbf{d}_2 + \dots + \alpha_{ik}\mathbf{d}_k \quad (1)$$

Here  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K] \in \mathbf{R}^{B \times K}$  ( $B \ll K$ ) is the overcomplete background dictionary with  $K$  atoms,  $\mathbf{d}_i$  ( $1 \leq i \leq K$ ) denotes the  $i$ th atom, and  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$  is the sparse coefficient vector with only a few nonzero entries. This implies that  $\mathbf{x}_i$  can be represented with the linear combination

of  $K_0$  atoms in  $\mathbf{D}$  in which  $K_0$  is far less than  $K$ . The sparse vector can be acquired via solving the following optimization problem

$$\min_{\alpha_i} \|\mathbf{x}_i - \mathbf{D}\alpha_i\|_2^2 \quad \text{s.t. } \|\alpha_i\|_0 \leq K_0 \quad \forall i \quad (2)$$

where  $\|\cdot\|_0$  denotes the  $l_0$ -norm that counts the number of nonzero entries in the vector, and  $K_0$  is the upper bound of the sparsity level for  $\alpha_i$ . This optimization problem can be solved by the orthogonal matching pursuit algorithm (OMP) [22]. Once the estimated coefficient vector  $\hat{\alpha}_i$  is obtained, the detection response of the  $i$  th pixel can be obtained by computing the reconstruction residual

$$r_i = \|\mathbf{x}_i - \mathbf{D}\hat{\alpha}_i\|_2 \quad (3)$$

Here  $r_i$  is the reconstruction residual of the pixel  $\mathbf{x}_i$ . If the residual  $r_i$  is larger than a given threshold, then the test pixel  $\mathbf{x}_i$  is considered to be an anomalous pixel.

### B. SMACC ENDMEMBER EXTRACTION

In an HSI, an endmember refers to the spectral characteristics of certain one type pure component. In order to extract endmember spectra and abundance maps simultaneously, Gruninger *et al.* proposed the sequential maximum angle convex cone (SMACC) endmember extraction model. Given an HSI dataset  $\mathbf{X} \in \mathbf{R}^{B \times N}$ , where  $B$  is the number of spectral bands and  $N$  denotes the number of pixels, the linear spectral mixture model can be written as follows

$$\mathbf{X}_{i,j} = \sum_{h=1}^H \mathbf{M}_{i,h} \mathbf{A}_{h,j} + \mathbf{R}_{i,j} \quad (4)$$

Here  $\mathbf{X}_{i,j}$  denotes the  $i$  th band of the  $j$  th pixel in  $\mathbf{X}$ ,  $H$  is the expansion length.  $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_H] \in \mathbf{R}^{B \times H}$  is the endmember spectral matrix, where each column indicates an endmember spectrum vector.  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_H]^T \in \mathbf{R}^{H \times N}$  is the abundance matrix, where each row contains the abundance map of the corresponding endmember for each pixel. The matrix  $\mathbf{R} \in \mathbf{R}^{B \times N}$  is the residuals. The SMACC model recognizes the endmember spectra via a convex cone model, and a positive constraint is imposed since the spectrum vector represents reflectance. The convex cone is determined by using the extreme points and thus the first endmember is defined. The residuals denote the elements distributed outside the convex cone. The rest endmember is successively derived by implementing a constrained oblique projection on the previous convex cone. Adding new endmembers alternates with updating the convex cone. This process is terminated until a certain error is satisfied. The final result of SMACC contains the endmember spectra set and the abundance images. Additionally, the abundance images demonstrate the contribution of endmembers for each pixel. This extraction process can be performed via the ENVI remote sensing image processing platform [23].

### III. PROPOSED METHOD

In this section, the detailed introduction of the proposed method is illustrated. This section includes four parts. In the first part, the background estimation strategy via the SMACC model is introduced. In the second part, the adaptive background sub-dictionary construction method based upon the atom usage probability (AUP) is described. In the third part, the spectral angle distance (SAD) based adaptive re-weighted SR based anomaly detection method is demonstrated. Finally, the overview of the proposed method is summarized.

#### A. BACKGROUND ESTIMATION STRATEGY

The performance of representation based anomaly detectors highly relies on the background dictionary. By constructing discriminative dictionary to improve detection performance has been a hot topic. Qu *et al.* [24] construct a background dictionary based on the estimated background from the main shift clustering algorithm instead of raw data, which will enhance the separation between anomalies and background. Ma *et al.* [25] divide background into several categories and select a series of representative samples from each categories to build multiple background dictionaries, so that the differences between anomalies and background are enhanced. Yang *et al.* [26] establish a pure background dictionary that excludes possible anomalies and thus providing more reliable detection results based on LRR model.

For the SR based detectors, the quality of background dictionary evidently influences the detection probability. Generally, two options of background dictionaries for unsupervised SR based detectors are available: the global dictionary and the local dictionary. The global one is usually constructed by randomly selecting some pixels from the HSI [27]. As for the local one, a dual-window strategy (shown in Fig. 1) is adopted and the pixels in the outer window are collected to form the dictionary. The local dictionary based SR model is referred as joint sparsity model. In the work by Zhu *et al.* [14], a new global background dictionary is constructed to eliminate the anomalies embedded in the background. The dictionary atoms are randomly selected from the estimated background by using the SMACC model, and the global dictionary is used directly for detection. Different from Zhu's work, we implement K-means clustering algorithm to the estimated

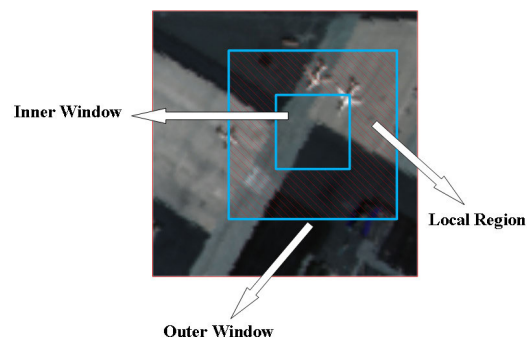


FIGURE 1. Dual-window strategy for anomaly detection.

background and choose several samples from each cluster to ensure that all types of background information can be revealed in this global dictionary. Moreover, we use this global dictionary to eliminate the anomaly contamination in the local regions in the dual-window.

The local region in the outer window can be regarded as a local background dictionary. Given a test pixel  $\mathbf{x}_i \in \mathbf{R}^{B \times 1}$ , the local region pixel set with  $L$  pixels is denoted as  $\mathbf{S} = [s_1, s_2, \dots, s_L] \in \mathbf{R}^{B \times L}$ , the local SR detection model can be expressed as

$$\min_{\alpha_i} \|\mathbf{x}_i - \mathbf{S}\alpha_i\|_2^2 \quad \text{s.t.} \quad \|\alpha_i\|_0 \leq K_0 \quad \forall i \quad (5)$$

However, the possible anomaly contamination in local region pixel set  $\mathbf{S}$  can significantly affect the detection result. In order to solve this problem, inspired by the works in [13] and [28], we consider constructing adaptive dictionaries from the perspective of background estimation. Two common situations of local regions mixed with anomaly spectrum signatures are presented in Fig. 2. The first situation depicted in Fig. 2(a) is caused by two factors: (1) the anomaly components are too close to each other. (2) The outer window size is large enough to include neighboring anomaly components. As for the situation in Fig. 2(b), this phenomenon happens when the size of the inner window is set as small as the anomaly component. As a consequence, the anomalies may exist in both the inner and the outer window.

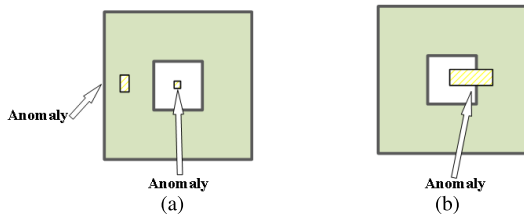


FIGURE 2. Two common situations of local regions mixed with anomaly spectrum signatures. (a) The first situation. (b) The second situation.

Since the background categories in local regions are less than in global scene, it is assumed that a global dictionary can be constructed where all local patches will lie in a low-dimensional subspace spanned by this dictionary. Therefore, each pixel in a local region can be regarded as the linear combination of the global dictionary atoms. In view of this, we consider extracting the most informative and discriminative atom sets in a global background dictionary to adapt the pure background information in local regions. For the local region pixel set  $\mathbf{S}$ , the global dictionary with  $K$  atoms is denoted as  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K] \in \mathbf{R}^{B \times K}$ . The  $j$  th pixel  $s_j$  ( $1 \leq j \leq L$ ) in  $\mathbf{S}$  can be represented as follows

$$s_j = \mathbf{H}\beta_j \quad \forall j \quad (6)$$

where  $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jK}]$  is the sparse coefficient vector. We select  $N_B$  atoms in  $\mathbf{H}$  that make major contribution to the representation. The selection procedure will be detailed in next subsection. We form a sub-dictionary with the selected atoms, denoted as  $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{N_B}] \in \mathbf{R}^{B \times N_B}$ . This sub-dictionary is assumed to contain no anomaly information.

It is also considered to reveal all the background information in  $\mathbf{S}$ . Then, the SR based anomaly detection process in Problem (5) can be replaced as

$$\min_{\alpha_i} \|\mathbf{x}_i - \mathbf{B}\alpha_i\|_2^2 \quad \text{s.t.} \quad \|\alpha_i\|_0 \leq k_0 \quad \forall i \quad (7)$$

After obtaining the sparse coefficient vector, the anomaly response can be calculated via (3).

As demonstrated above, the local regions can be replaced by the constructed sub-dictionaries  $\mathbf{B}$ . Obviously, the quality of the sub-dictionaries is determined by the global dictionary. Inspired by the works in [28], we first estimate the global background via SMACC model, then we utilize the K-means clustering algorithm to build the global dictionary. Yang *et al.* [26] directly perform K-means clustering algorithm to the original HSI to extract the background information and a satisfactory detection performance is achieved. However, the prior information of the HSI is unknown, excessive manual-set parameters may cause background information missing. In this paper, we expect to estimate a pure background that could cover all background categories.

According to the endmember model expressed in (4), all endmembers in an HSI dataset can be divided into background-related endmembers and anomaly-related endmembers [28]. Therefore, the HSI dataset  $\mathbf{X}$  can be modeled as

$$\mathbf{X} = \mathbf{M}_{Bg} \times \mathbf{A}_{Bg} + \mathbf{M}_{An} \times \mathbf{A}_{An} \quad (8)$$

Here  $\mathbf{M}_{Bg}$  denotes the background-related endmembers set and  $\mathbf{M}_{An}$  denotes the anomaly-related endmembers set.  $\mathbf{A}_{Bg}$  and  $\mathbf{A}_{An}$  are the corresponding abundance maps set (also known as abundance images set). Some of the extracted abundance images of the San Diego Airport hyperspectral image are shown as examples in Fig. 3.

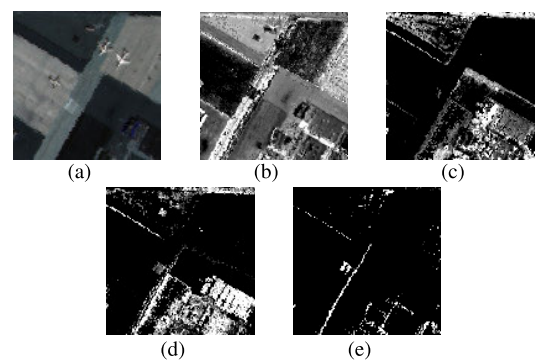


FIGURE 3. The examples of the extracted abundance images by the SMACC model. (a) The AVIRIS hyperspectral image. (b)-(e) Some of the extracted abundance images.

After implementing SMACC endmember extraction to the HSI, the endmember set  $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_H] \in \mathbf{R}^{B \times H}$  and the abundance maps set  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_H]^T \in \mathbf{R}^{H \times N}$  are obtained. For the  $i$  th abundance map  $\mathbf{a}_i$ , ( $1 \leq i \leq H$ ), the  $j$  th coefficients in  $\mathbf{a}_i$  are abundance fractions of the endmember  $\mathbf{m}_i$  for the corresponding pixel  $\mathbf{x}_j$ . If the  $j$  th coefficient in  $\mathbf{a}_i$  is larger than a preset threshold  $t$ , then



the endmember  $\mathbf{m}_i$  is considered to make major contribution to constitute the pixel  $\mathbf{x}_j$ . We define such pixel in the abundance map  $\mathbf{a}_i$  as the Major-Component-Related-Pixel (MCRP). The number of MCRPs in the abundance map  $\mathbf{a}_i$  is denoted as  $nMCRP_i$ . As mentioned before, anomalies usually appear with low probability in the image. This spatial characteristics suggests that if an abundance map contains very few MCRPs, the corresponding endmember can be regarded as an anomaly-related endmember, or else it is a background-related endmember. In each abundance map, the proportion of MCRPs  $nMCRP_i/N$  is calculated and then it is used to sort all abundance maps in descending order. The first  $s$  abundance maps are selected and the corresponding endmembers are determined as background-related endmembers. In this way, the first part of the model in (8) is obtained, which means the pure background estimation  $\mathbf{X}_{Bg}$  can be extracted as

$$\mathbf{X}_{Bg} = \mathbf{M}_{Bg} \times \mathbf{A}_{Bg} \quad (9)$$

where  $\mathbf{M}_{Bg} = [\mathbf{m}_{Bg1}, \mathbf{m}_{Bg2}, \dots, \mathbf{m}_{Bgs}] \in \mathbf{R}^{B \times s}$ , ( $s < H$ ).

Zhou *et al.* [29] perform clustering on the background to generate several cluster centers and then directly apply them in anomaly detection based on kernel RX algorithm. In this paper, after the pure background information is extracted, the K-means clustering algorithm is used to divide the background dataset  $\mathbf{X}_{Bg}$  into  $K_B$  clusters. The number of clusters  $K_B$  is estimated by the HySime algorithm [30]. Then, we randomly choose  $P$  percent samples in each cluster to construct the global background dictionary  $\mathbf{H}$ .

### B. ADAPTIVE BACKGROUND SUB-DICTIONARY CONSTRUCTION METHOD

As mentioned previously, the local region pixel set in outer window can be represented as a linear combination of atoms in the global background dictionary  $\mathbf{H}$ . In order to extract the most informative atom set, the atom usage probability (AUP) [15] method is adopted in this subsection. It measures the reconstruction strength of atoms by computing the normalized norms of the rows in the sparse coefficient matrix. For the local region pixel set  $\mathbf{S}$ , the  $j$  th column  $\beta_j$  ( $1 \leq j \leq L$ ) in the sparse coefficient matrix can be derived by solving the following optimization problem:

$$\min_{\beta_j} \|\mathbf{s}_j - \mathbf{H}\beta_j\|_2^2 + \lambda \|\beta_j\|_1 \quad (10)$$

where  $\lambda$  is a regularization parameter, and the corresponding sparse coefficient matrix is denoted as  $\beta = [\beta_1, \beta_2, \dots, \beta_L]$ . The regularization term  $\|\cdot\|_1$  is the convex relaxation of the  $l_0$ -norm. Then the AUP value of the  $k$  th atom  $\mathbf{h}_k$  can be calculated as

$$AUP_k = \frac{\sum_{j=1}^L |\beta_{kj}|}{\sum_{j=1}^L \sum_{g=1}^K |\beta_{gj}|} \quad (11)$$

A higher AUP value suggests a stronger role the atom plays in the reconstruction. The greater the AUP value diverges between atoms, the more conducive is for the selection of

active atoms. Aiming to increase the divergence of the AUP values, in this paper, the  $l_{1/2}$ -norm is used to replace the regularization term in Problem (10) to enhance the sparsity in matrix  $\beta$ . The objective function can be rewritten as

$$\min_{\beta_j} \|\mathbf{s}_j - \mathbf{H}\beta_j\|_2^2 + \lambda \|\beta_j\|_{1/2}^{1/2} \quad (12)$$

where  $\|\cdot\|_{1/2}^{1/2}$  denotes the  $l_{1/2}$ -norm. Compared with the  $l_1$ -norm, the  $l_q$ -norm regularization can yield more sparse solutions when  $q \in (0, 1)$  [31]. Especially, when  $q \in [1/2, 1)$ , the sparsity of the solution drops remarkably as the value of  $q$  decreases, while the sparsity shows no prominent difference when  $q$  varies in  $(0, 1/2)$ . Hence, the  $l_{1/2}$ -norm regularization is adopted to enhance the sparsity of the coefficient matrix. The solution of Problem (12) can be obtained by the iterative half-thresholding algorithm [31]. Subsequently, the AUP values of atoms in  $\mathbf{H}$  are acquired and the atoms are sorted by the AUP values in descending order. The first  $N_B$  atoms are selected to form the adaptive background sub-dictionary  $\mathbf{B}$  instead of the local region pixel set  $\mathbf{S}$ . In this paper, the value of  $N_B$  is decided by  $N_B = M \cdot B$ , where  $M$  is a proportion parameter within  $(0,1]$ , and  $B$  is the number of band. In this way, the anomaly contamination in the local regions can be eliminated and thus the detection results can be further improved.

### C. RE-WEIGHTED SR BASED ANOMALY DETECTION

For each test pixel  $\mathbf{x}_i$ , with the corresponding sub-dictionary  $\mathbf{B}$  acquired above, the objective function of the SR based anomaly detection can be illustrated as follows

$$\hat{\alpha}_i = \operatorname{argmin}_{\alpha_i} \|\mathbf{x}_i - \mathbf{B}\alpha_i\|_2^2 + \gamma \|\alpha_i\|_1 \quad (13)$$

where  $\|\cdot\|_1$  is the convex relaxation of the  $l_0$ -norm, and the  $\gamma$  is a regularization parameter. To further improve the detection performance, for center pixels that are quite different from the atoms in background sub-dictionary  $\mathbf{B}$ , the coefficients should be suppressed, i.e. the penalty in regularization term for yielding large coefficients should be heavy. In this paper, a spectral angle distance (SAD) based re-weighting strategy is proposed to adjust the coefficient vectors. The SAD depicts the similarity between two spectrum vectors  $\mathbf{x}$  and  $\mathbf{y}$ . It is defined as

$$SAD(\mathbf{x}, \mathbf{y}) = \arccos\left(\frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right) \quad (14)$$

If two spectrum vectors are significantly different from each other, then their SAD score is large. We consider enhancing the penalty by calculating the SAD between the input pixel and all the atoms in the sub-dictionary. For the input test pixel  $\mathbf{x}_i$  with the corresponding sub-dictionary  $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M]$ , we design an adaptive divergence measurement operator (ADMO) defined as

$$ADMO_{ij} = \frac{\exp(\pi - SAD_{ij})}{\sigma} \quad (15)$$

Here  $ADMO_{ij}$  measures the divergence between the  $i$ th pixel  $\mathbf{x}_i$  and the  $j$ th atom  $\mathbf{b}_j$  in  $\mathbf{B}$ , and  $\sigma$  is a positive adjusting factor. Obviously, a larger value of  $ADMO$  indicates a greater difference between the input pixel and the atoms. Finally, the anomaly detection based on adaptive re-weighted sparse representation can be formulated as

$$\hat{\alpha}_i = \operatorname{argmin}_{\alpha_i} \|\mathbf{x}_i - \mathbf{B}\alpha_i\|_2^2 + \gamma \|\mathbf{W}\alpha_i\|_1 \quad (16)$$

where  $\mathbf{W}$  is the weight matrix with  $ADMO$  values on the diagonal and other elements are zero. It is depicted as

$$\mathbf{W} = \begin{bmatrix} \frac{\exp(\pi - SAD_{i1})}{\sigma} & & 0 \\ & \ddots & \\ 0 & & \frac{\exp(\pi - SAD_{iM})}{\sigma} \end{bmatrix} \quad (17)$$

The coefficient vectors can be obtained by solving the optimization problem (16) with the SPAM toolbox [32]. Finally, the anomaly pixels can be extracted by the following response value

$$r_i = \|\mathbf{x}_i - \mathbf{B}\hat{\alpha}_i\|_2 \quad (18)$$

If the response value of the test pixel is greater than a preset threshold, it can be determined as an anomalous pixel.

#### D. OVERVIEW OF THE PROPOSED METHOD

The overview of the proposed method is illustrated in Fig. 4. In this paper, we propose an anomaly detection method based on sparse representation via adaptive background sub-dictionaries. The main idea of this method is to estimate the global background via SMACC endmember extraction model, and then the K-means clustering algorithm is used to form a global background dictionary. With the dual-window strategy, the local region pixel set in the outer window are approximated by active atoms in the global background dictionary. Finally, for each local regions, these atoms are selected to form a sub-dictionary for the SR based anomaly detection. Additionally, a spectral angle distance (SAD) based re-weighting strategy is proposed to improve the detection performance. The detailed steps of the proposed method are described in 1.

#### IV. EXPERIMENTS AND ANALYSIS

In this section, in order to evaluate the effectiveness of our method on HSI anomaly detection, we conduct experiments on one synthetic dataset and five real-world HSI datasets and three real-world HSI datasets are used for experiments to analyze the improvement of our method and the parameters settings. The assessment criteria used includes color detection map, receiver operating characteristics (ROC) curve, area under curve (AUC) value, and background-anomaly separability map. The experiments are implemented via MATLAB 2018a on a laptop with an Intel i5-7300HQ 2.50 GHz CPU, 16 GB memory, and 64-bit Windows 10 operating system. The constituent parts of this section are described as follows:

#### Algorithm 1 Sparse Representation Based Hyperspectral Anomaly Detection via Adaptively Estimated Background Sub-Dictionaries

**Input:** The HSI dataset  $\mathbf{X}$ , the two dual window sizes  $w_{in}$  and  $w_{out}$ , the number of sub-dictionary atoms  $N_B$ , the threshold  $t$  and  $s$ , the parameter  $P$  in K-means clustering, the regularization parameter  $\lambda$  and  $\gamma$ .

1. Re-arrange  $\mathbf{X}$  into a 2-D matrix.
2. Adopt SMACC model to extract the endmembers spectral set  $\mathbf{M}$  and the corresponding abundance maps set  $\mathbf{A}$ .
3. With the method in Section 3.1, select  $s$  background-related endmembers and obtain the estimated background  $\mathbf{X}_{bg}$ .
4. Use K-means clustering algorithm for  $\mathbf{X}_{bg}$ , then in each cluster randomly select  $P$  percent spectral vectors to form the global background dictionary  $\mathbf{H}$ .
2. **for**  $i = 1 : N$  **do**
3. Use dual-window strategy to extract the local region  $\mathbf{S}$  around pixel  $\mathbf{x}_i$ .
4. Calculate the sparse matrix  $\beta$  by optimizing the Problem (12) with the iterative half-thresholding algorithm.
5. Calculate the AUP values of all atoms by (11), and select  $N_B$  atoms to form the sub-dictionary  $\mathbf{B}$ .
6. Obtain the anomaly response of pixel  $\mathbf{x}_i$  via (13) - (18).
7. **end for**

**Output:** The anomaly detection map  $\mathbf{R}$ .

1. The detail information of the used HSI datasets is described.

2. Detection performance of the proposed method on six HSI datasets are evaluated. Other six anomaly detection methods are used for comparison at the same time. The evaluation criteria include color detection map, receiver operating characteristics (ROC) curve, and background-anomaly separability map.

3. The advantages of the background estimation strategy,  $l_{1/2}$ -norm constraint and the re-weighting strategy are discussed via receiver operating characteristics (ROC) curve, area under curve (AUC) value, and background-anomaly separability map. The experiments are conducted on three HSI datasets.

4. The effect of several significant parameters in our method are analyzed via area under curve (AUC) value. Then, the optimal parameters settings are analyzed with the experimental results.

#### A. DATA DESCRIPTION

The synthetic dataset is generated by embedding simulated anomaly pixels in a real-world hyperspectral image from the San Diego Airport hyperspectral dataset captured by the Airborne Visible /Infrared Imaging Spectrometer (AVIRIS). A sub-region with a size of  $100 \times 100$  from this dataset is selected as background and it is consist of 224 spectral bands. After removing bands with low signal-to-noise-ration

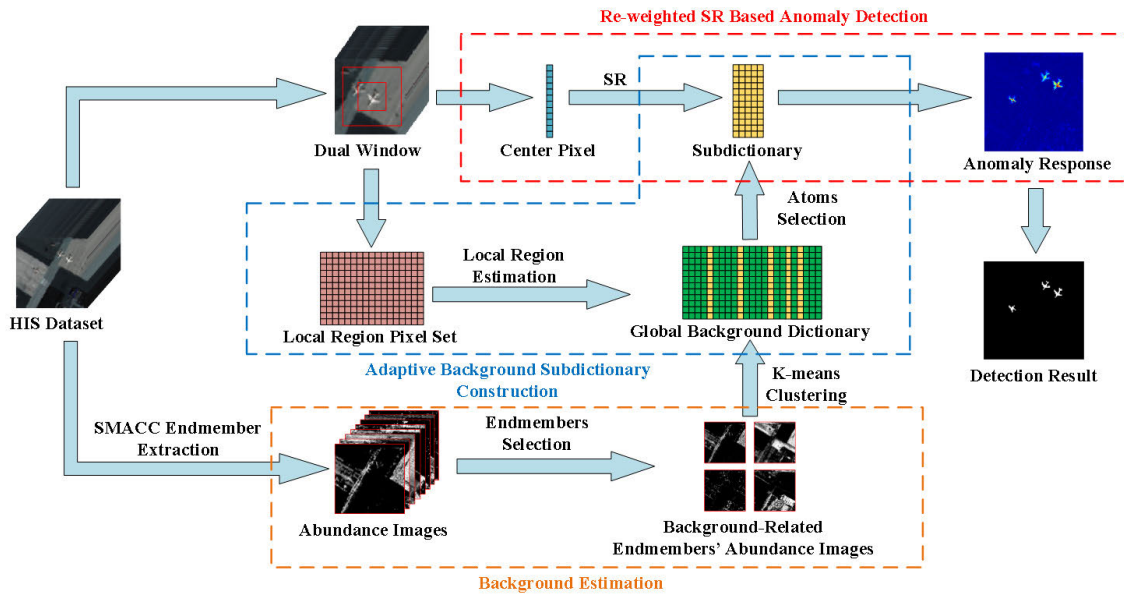


FIGURE 4. Overview of the proposed method.

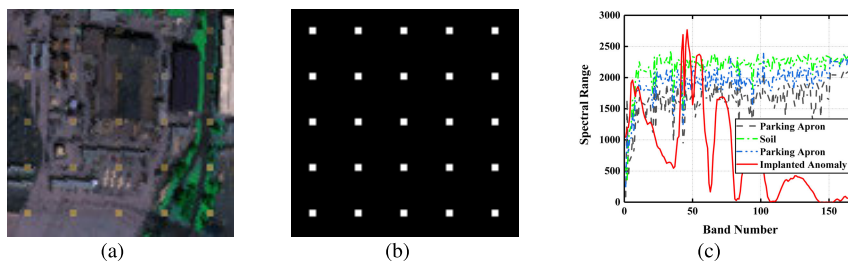


FIGURE 5. The pseudo color images, ground truth maps and the spectral curves of the synthetic dataset. (a) The pseudo color images. (b) The ground truth maps. (c) The spectral curves.

(SNR) and water vapor absorption (1-6, 33-35, 97, 107-113, 153-166 and 221-224), 189 bands are retained. A group of synthetic anomalies are implanted based on the linear mixing model with the desired target spectrum  $\mathbf{t}$ , background spectrum  $\mathbf{b}$  and the specified abundance fraction  $\alpha$  as follows:

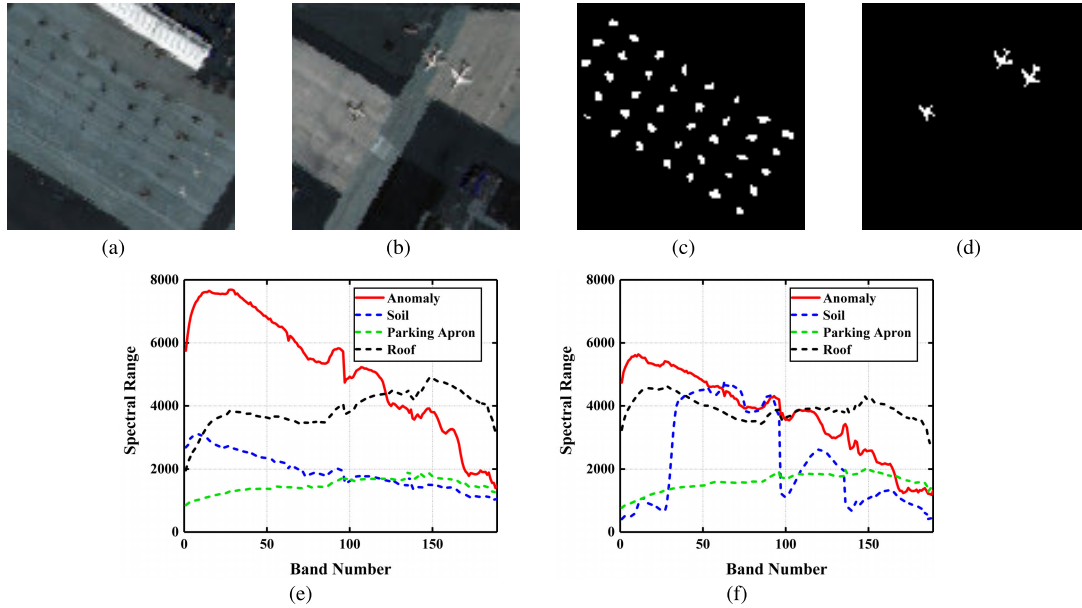
$$\mathbf{x} = \alpha \cdot \mathbf{t} + (1 - \alpha) \cdot \mathbf{b} \quad (19)$$

The target spectrum corresponds to a real-world aircraft. 25 targets (150 pixels) are implanted and distributed in 5 rows and 5 columns. The abundance fractions  $\alpha$  for the first row to the last row are 0.2, 0.35, 0.5, 0.65, 0.8 respectively. The pseudo-color image, the ground-truth map, and the spectral curves of different components are shown in 5(a), 5(b) and 5(c) respectively.

The first two real-world HSI datasets are from the hyperspectral image of the San Diego airport area captured by the Airborne Visible /Infrared Imaging Spectrometer (AVIRIS). The raw dataset consists of 224 spectral bands ranging from 370-2510 nm. In our experiment, the bands with low signal-to-noise-ratio (SNR) and water vapor absorption (1-6, 33-35, 97, 107-113, 153-166 and 221-224) are removed. The spatial resolution is 3.5 m per pixel. We select two

subregions in this HSI in our experiment as shown in Fig. 6(a) and Fig. 6(b). Their corresponding ground truth maps and spectral curves are shown in Fig. 6(c)-6(d) and 6(e)-6(f) respectively. The sizes of selected areas are both  $100 \times 100$  and the background includes roof, soil, concrete parking apron and shadows. The anomaly targets in these images are airplanes lying on the ground.

The last three real-world HSI datasets are from the Airport-Beach-Urban (ABU) database. The HSIs in this database are mostly captured by the AVIRIS sensor, and others are from the Reflective Optics System Imaging Spectrometer (ROSIS) sensor. Three images that contain  $100 \times 100$  pixels from this database are picked out for our experiment. The pseudo color images are depicted in Fig. 7(a) - Fig. 7(c). Their corresponding ground truth maps and spectral curves are presented in Fig. 7(d) - Fig. 7(f) and Fig. 7(g) - Fig. 7(i) respectively. As shown in Fig. 7(a) - Fig. 7(b), the airport dataset and the urban dataset are scenes with several airplanes as anomalies. The backgrounds in these two HSIs are mainly soil, asphalt road, parking apron and roofs. The beach dataset in Fig. 7(c) contains sea, asphalt road, soil and sand as background. The embedded anomalies are vehicles in the asphalt road.



**FIGURE 6.** The pseudo color images, ground truth maps and the spectral curves of the AVIRIS datasets for experiments. (a)-(b) The pseudo color images. (c)-(d) The ground truth maps. (e)-(f) The spectral curves.

**TABLE 1.** Parameters settings for comparison methods.

Method	Synthetic Dataset	AVIRIS Dataset 1	AVIRIS Dataset 2	ABU-Airport	ABU-Urban	ABU-Beach
LRX	$win_{in} = 5$ $win_{out} = 11$	$win_{in} = 5$ $win_{out} = 9$	$win_{in} = 5$ $win_{out} = 9$	$win_{in} = 9$ $win_{out} = 13$	$win_{in} = 5$ $win_{out} = 9$	$win_{in} = 5$ $win_{out} = 9$
LKRX	$win_{in} = 7$ $win_{out} = 11$ $\sigma = 40$	$win_{in} = 5$ $win_{out} = 9$ $\sigma = 40$	$win_{in} = 5$ $win_{out} = 9$ $\sigma = 40$	$win_{in} = 9$ $win_{out} = 13$ $\sigma = 40$	$win_{in} = 5$ $win_{out} = 9$ $\sigma = 40$	$win_{in} = 5$ $win_{out} = 9$ $\sigma = 40$
CRD	$win_{in} = 5$ $win_{out} = 9$ $\lambda = 0.01$	$win_{in} = 5$ $win_{out} = 9$ $\lambda = 0.01$	$win_{in} = 7$ $win_{out} = 15$ $\lambda = 0.01$	$win_{in} = 9$ $win_{out} = 13$ $\lambda = 0.01$	$win_{in} = 7$ $win_{out} = 13$ $\lambda = 0.01$	$win_{in} = 7$ $win_{out} = 11$ $\lambda = 0.01$
KCRD	$win_{in} = 5$ $win_{out} = 9$ $\lambda = 0.01$ $\sigma = 16$	$win_{in} = 5$ $win_{out} = 9$ $\lambda = 0.01$ $\sigma = 16$	$win_{in} = 7$ $win_{out} = 15$ $\lambda = 0.01$ $\sigma = 16$	$win_{in} = 9$ $win_{out} = 13$ $\lambda = 0.01$ $\sigma = 16$	$win_{in} = 7$ $win_{out} = 13$ $\lambda = 0.01$ $\sigma = 16$	$win_{in} = 7$ $win_{out} = 11$ $\lambda = 0.01$ $\sigma = 16$
BJSRD	$win_{in} = 5$ $win_{out} = 9$ $win_{search} = 11$	$win_{in} = 5$ $win_{out} = 9$ $win_{search} = 11$	$win_{in} = 11$ $win_{out} = 15$ $win_{search} = 17$	$win_{in} = 9$ $win_{out} = 13$ $win_{search} = 17$	$win_{in} = 7$ $win_{out} = 11$ $win_{search} = 13$	$win_{in} = 9$ $win_{out} = 15$ $win_{search} = 17$
BEAWSR	$L_0 = 0.15$ $L_1 = 0.3$ $W_{SE} = 9 \times 9$ $\lambda = 10, K_0 = 6$	$L_0 = 0.15$ $L_1 = 0.3$ $W_{SE} = 9 \times 9$ $\lambda = 10, K_0 = 6$	$L_0 = 0.20$ $L_1 = 0.3$ $W_{SE} = 13 \times 13$ $\lambda = 12, K_0 = 6$	$L_0 = 0.20$ $L_1 = 0.3$ $W_{SE} = 9 \times 9$ $\lambda = 10, K_0 = 6$	$L_0 = 0.15$ $L_1 = 0.3$ $W_{SE} = 11 \times 11$ $\lambda = 12, K_0 = 6$	$L_0 = 0.30$ $L_1 = 0.3$ $W_{SE} = 7 \times 7$ $\lambda = 15, K_0 = 6$

The Airport and the Urban datasets contain 205 bands and the Beach dataset contains 105 bands.

**B. DETECTION PERFORMANCE**

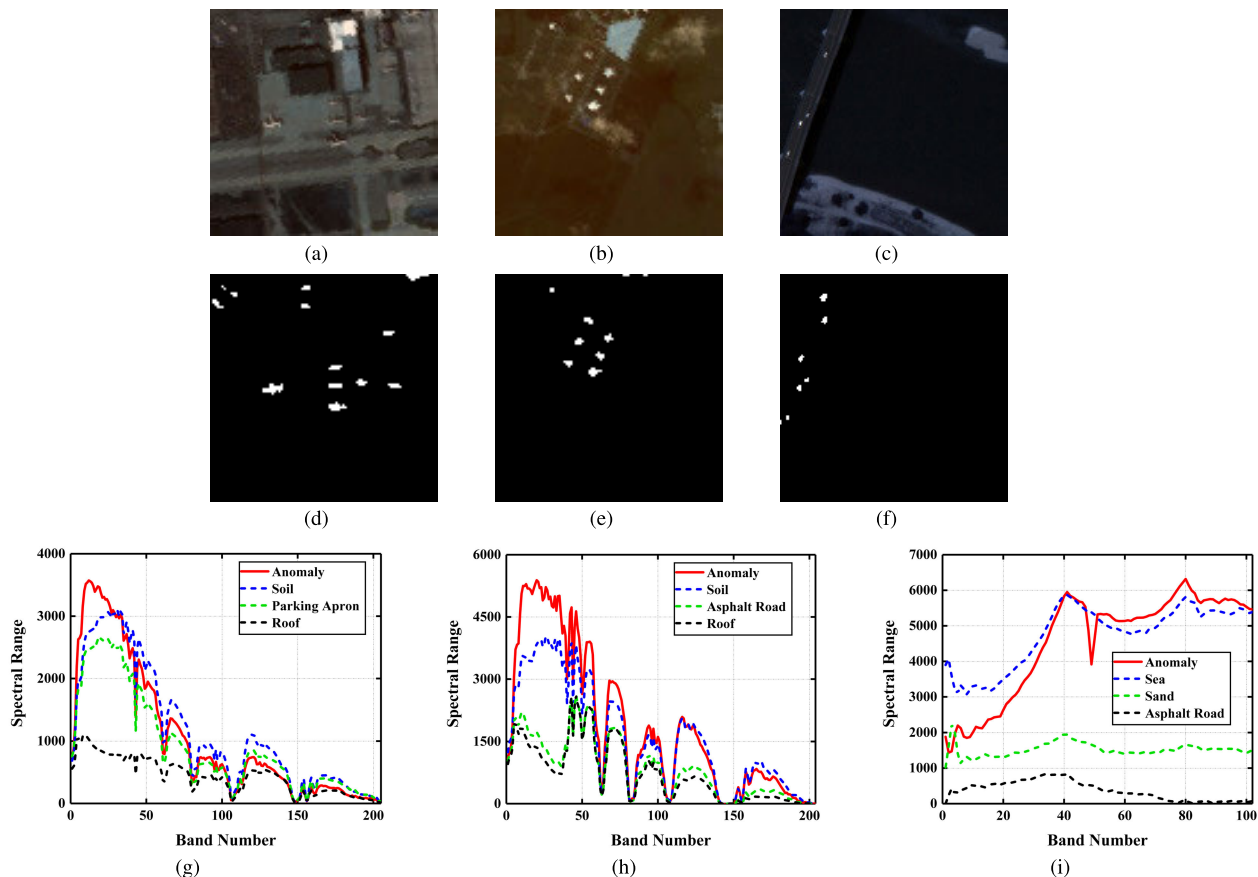
In this section, the performance of the proposed method is evaluated on the aforementioned HSI datasets. Six other detectors are introduced for comparison: LRX [5], Local KRX (LKRX) [8], CRD [19], KCRD [19], BJSRD [13], and BEAWSR [14], where first five methods are all local methods. The assessment criteria used in this section are color detection map, receiver operating characteristics (ROC)

curve and background-anomaly separability map. The ROC curve is a quantitative criterion for detection performance assessment. It plots the relationship between the probability of detection (PD) and the false alarm rate (FAR). The PD and the FAR are defined as

$$PD = \frac{N_{cd}}{N_t}, \quad FAR = \frac{N_{fd}}{N} \quad (20)$$

Here  $N_{cd}$  is the number of correctly detected anomaly pixels,  $N_t$  denotes the number of real anomaly pixels,  $N_{fd}$  is the number of falsely detected anomaly pixels, and  $N$  is the total number of pixels in the HSI. If an anomaly detector





**FIGURE 7.** The pseudo color images, ground truth maps and the spectral curves of the ABU datasets for experiments. (a)-(c) The pseudo color images. (d)-(f) The ground truth maps. (g)-(i) The spectral curves.

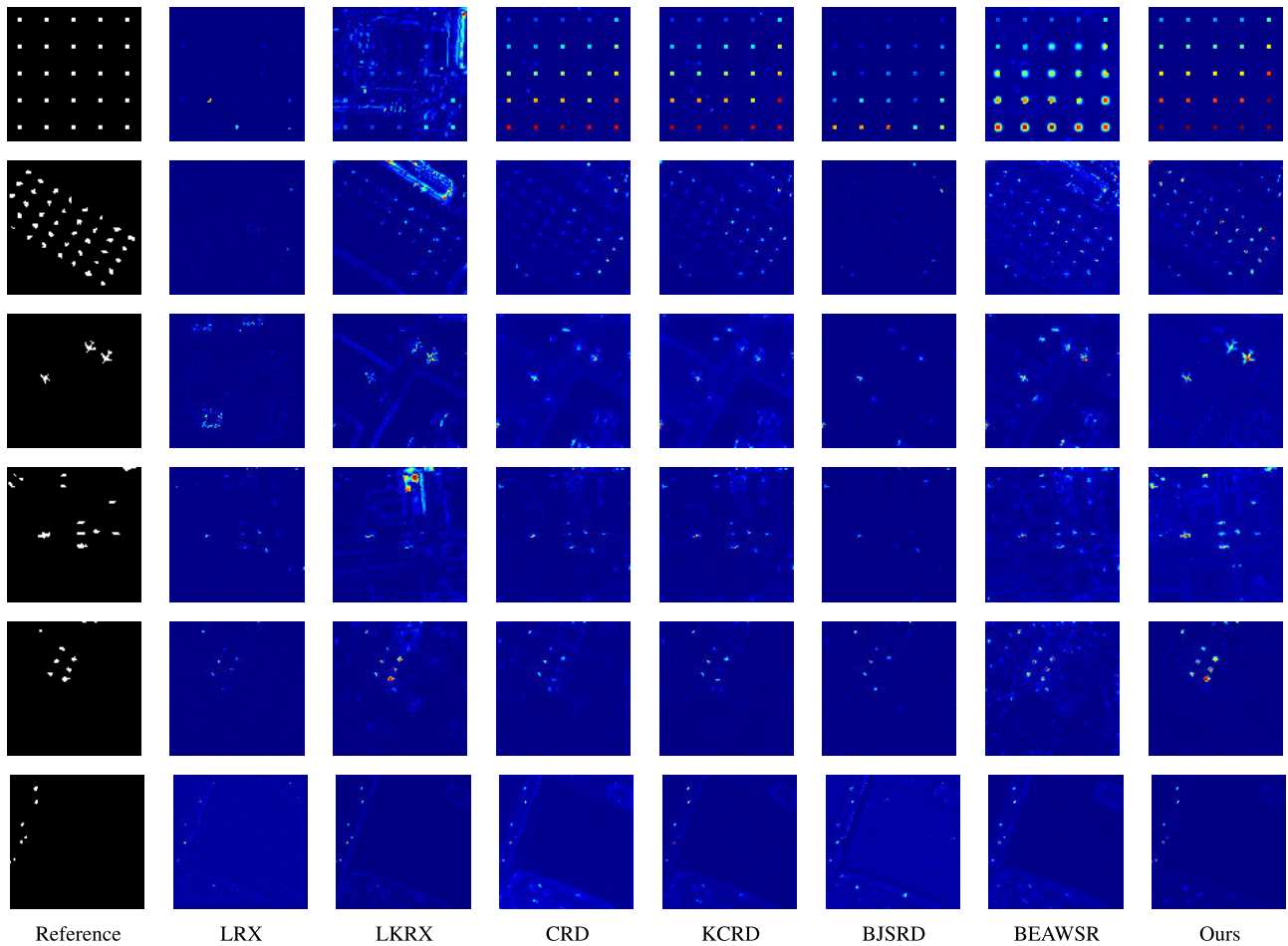
outperforms the others, its ROC curve possesses the upper left position, which indicates that it could achieve a higher probability of detection with the same false alarm rate.

The parameters for the proposed method are set as follows: the dual window sizes are set  $(5 \times 5, 11 \times 11)$  for the synthetic dataset and the first AVIRIS dataset,  $(7 \times 7, 13 \times 13)$  for the second AVIRIS dataset,  $(9 \times 9, 13 \times 13)$  for the ABU-Airport dataset,  $(7 \times 7, 11 \times 11)$  for the ABU-Airport dataset, and  $(5 \times 5, 9 \times 9)$  for the ABU-Airport dataset. The proportion  $M$  are set 0.5 for all six datasets. The threshold  $t$  and  $s$  are set 0.7 and 4 for the synthetic dataset and the two AVIRIS datasets, 0.8 and 4 for the ABU airport dataset and the ABU-Urban dataset, and 0.8 and 6 for the ABU beach dataset. The parameter  $P$  in the K-means clustering is set 35. The regularization parameters  $\lambda$  and  $\gamma$  are set 0.01 and 0.05 for all six HSI datasets. The parameters for LRX, LKRX, CRD, KCRD, BJSRD and BEAWSR are set after cross-validation. The parameters settings are listed in Table 1.

The color detection maps of all methods on one synthetic dataset and five real HSI datasets are depicted in Fig. 8. The first column presents the ground truth maps of the corresponding datasets as references. As shown in the first row, our method can simultaneously effectively identify the anomalies and suppress the background. The anomalies in the detection

maps of CRD, KCRD, BJSRD, BEAWSR and the proposed method have significantly greater responses than the background. For LRX, only several anomalies with abundance fraction larger than 0.35 are barely detected. For the kernel version of LRX, the detection result is severely interfered by the background. Although CRD, KCRD and BEAWSR can well identify all implanted anomalies, the background components in the middle left area of the scene are not effectively suppressed. Compared to BJSRD, the higher brightness of the anomaly pixels in the detection map of our method indicates that their detection responses are stronger than those of BJSRD. This implies that our method obtains a better detection performance than BJSRD. Generally, throughout all the detection maps in the first row, the ability for anomaly detection and background suppression of our method is relatively better than all the other methods.

For the first AVIRIS dataset, LRX can hardly identify the anomalies in the scenario and the anomaly responses of BJSRD are evidently weak. It can be seen in the results of LKRX, CRD, KCRD and BEAWSR that the anomaly pixels are identified at different levels while the background components at the top right of the scene are also highlighted by these four detectors. As for the detection result of our method, most of the anomaly pixels have evident detection responses



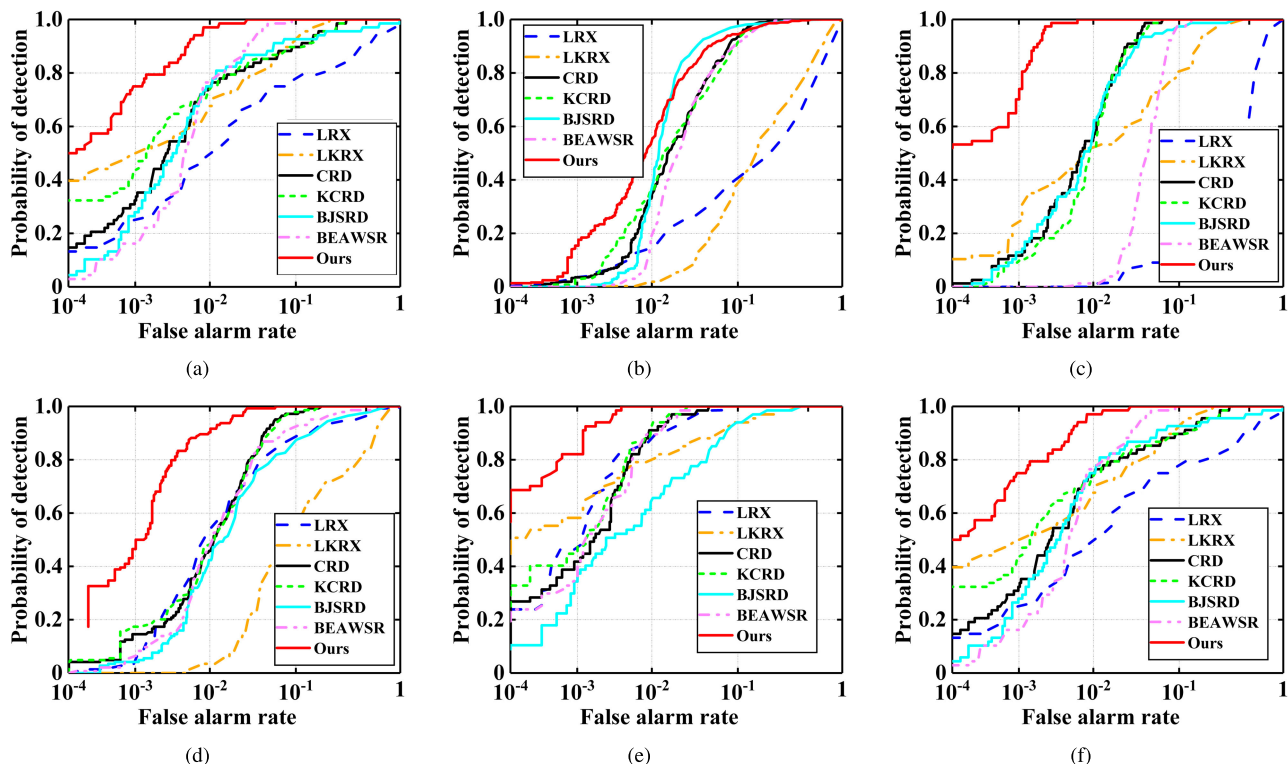
**FIGURE 8.** The ground truth maps and the color detection maps of the HSI datasets. The datasets in first row to last row are synthetic dataset, AVIRIS San Diego Airport 1, AVIRIS San Diego Airport 2, ABU-Airport, ABU-Urban and ABU-Beach.

and the background interference at the top right are more effectively suppressed compared to LKRX, CRD, KCRD and BEAWSR. For the second AVIRIS dataset, as shown in second row in Fig. 8, LRX fails to identify any anomalies in the image. Meanwhile, other five comparison methods can only extract a few anomaly pixels with heavy false alarms at the top of the scene. Our method not only identifies the anomaly targets with clear shape but also suppresses most of the background. These imply that our method can outperform all the other comparison methods on two AVIRIS datasets.

As depicted in fourth row to last row, anomalies in ABU datasets exist in complex background with various constituent parts. Under this situation, our method can prominently enhance the anomalies and effectively suppress the background components at the same time. As for the other methods, CRD, KCRD and BJSRD have comparable performances on ABU-Urban. On ABU-Beach, all the comparison methods expect LRX can successfully identify all the anomalies. However, they have weaker background suppression compared to our method. For the ABU-Airport dataset, there are a number of undetected anomaly pixels in the results of six comparison methods. Our method has much stronger

responses of anomalies while there are obvious false alarms at the top of the scene. In general, these observations show that our method achieves a more stable detection performance under complex backgrounds.

The ROC curves of all methods on the synthetic dataset and the five real-world datasets are illustrated in Fig. 9 as quantitative comparisons. In Fig. 9(a), it is observed that the comparison methods obtain similar detection performances except LRX, while our method possesses a prominently higher position. It can be observed from Fig. 9(b) that for the AVIRIS dataset 1, the curves of our method, BJSRD and BEAWSR are close to each other. The probability of detection for BJSRD reaches 1 with even a lower false alarm rate than our method. However, the area under the curve of our method is the largest among all methods. For the assessment result in Fig. 9(c), the detection probability of our method achieves 1 with the false alarm rate less than  $10^{-2}$ , while the detection probabilities of other methods are barely over 0.4. As for the comparison results in Fig. 9(d), Fig. 9(e) and Fig. 9(f), the curves of our method are much closer to the top-left corner than the other methods. In these three figures, the positions of the curves for CRD, KCRD, BJSRD and BEAWSR indicate



**FIGURE 9.** The ROC curves of all methods on six HSI datasets. (a) Synthetic Dataset. (b) San Diego Airport 1. (c) San Diego Airport 2. (d) ABU-Airport. (e) ABU-Urban. (f) ABU-Beach.

that they obtain similar detection performances. The above observations illustrate that our method achieves more superior detection results than the other methods on all hyperspectral datasets in this experiment.

To further quantitatively validate the superiority of our method, the normalized background-anomaly separability maps are illustrated in Fig. 10. The red boxes and the green boxes represent the statistics distributions of background and anomalies, respectively. It can be seen in Fig. 10(a) that for the synthetic dataset, CRD, KCRD, BJSRD, BEAWSR and our method can evidently separate anomalies and backgrounds. Our method achieves the best separation since the gap between the two boxes is the largest. For two AVIRIS datasets, as shown in Fig. 10(b) and Fig. 10(c), the red boxes and the green boxes of our method obtain the largest gaps with no overlaps among all methods, which reveal strong discrimination power between anomalies and backgrounds. In Fig. 10(d)-Fig. 10(f), the background-anomaly separation performance on three ABU datasets are described. These three figures depict prominent superiority of our method on anomaly extraction ability. Furthermore, from all the separability maps in Fig. 10, it can be observed that the background distribution boxes of our method are all suppressed to the most narrow ones, which demonstrates that our method can effectively suppress background components. All the analyses above correspond with the observations in the color detection maps.

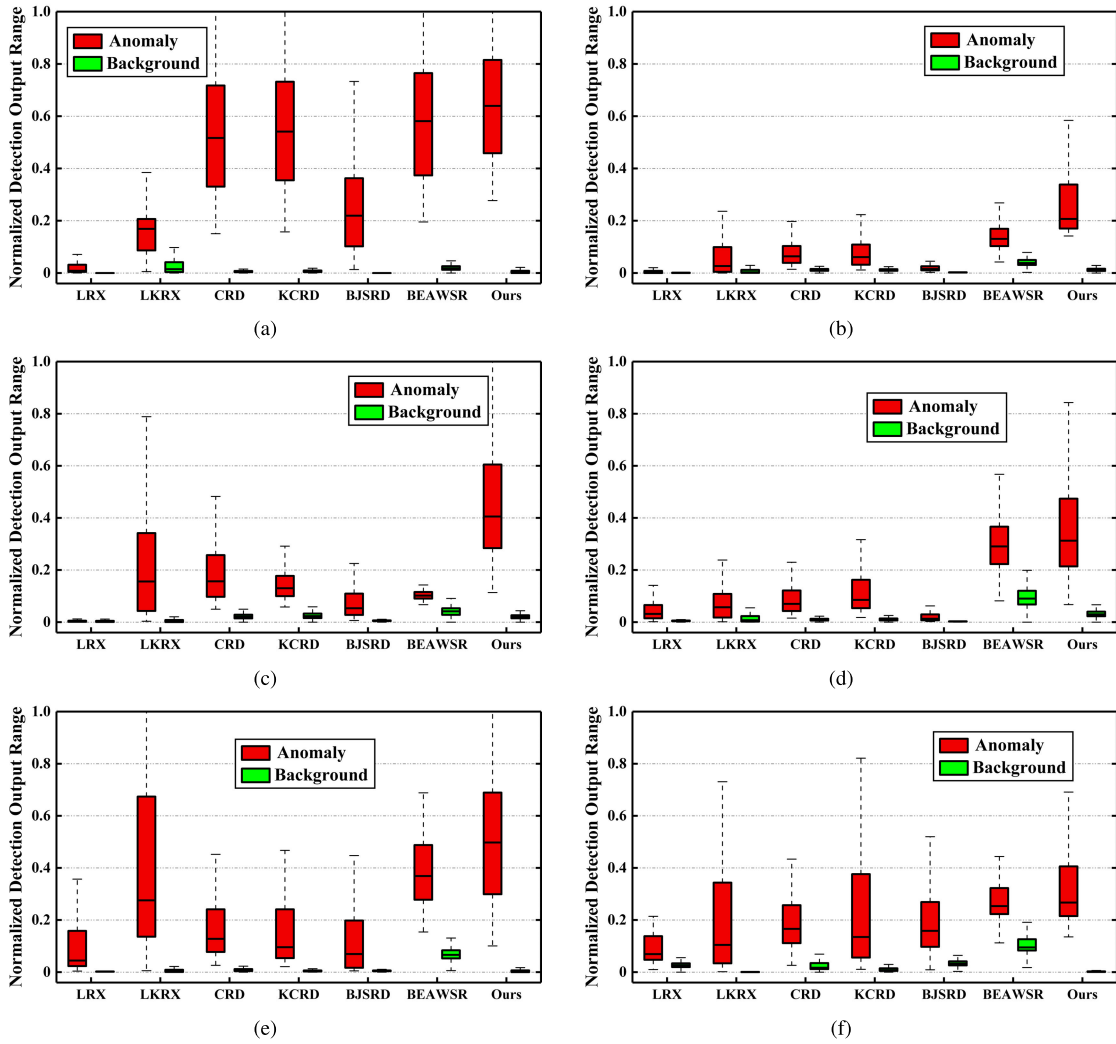
Additionally, we compare the computational time of every method on all six datasets. The results depicted in Table 2 show that CRD costs least computational time compared with other six methods. As for our method, the time cost is relatively less than LKRX and BJSRD on average but evidently more than other three local methods. This is due to the sub-dictionary construction for each test pixel. However, since the sub-dictionary construction process for each test pixel is independent, the heavy computational burden can be improved by employing parallel computation.

### C. DISCUSSION

In this section, the advantages of the proposed background estimation strategy, the  $l_{1/2}$ -norm regularization and the proposed re-weighting strategy are discussed in the following. The related experiments are implemented on three HSI datasets: AVIRIS San Diego Airport 2, ABU-Airport and ABU-Beach.

#### (1) The proposed background estimation strategy

We first conduct experiments to validate the effectiveness of the background estimation strategy proposed in Section 3.1. In order to highlight the advantages, two comparison methods are designed by modifying the means of global dictionary construction: (a) directly applying K-means clustering to the original HSI and then the global dictionary is formed by selecting total  $K$  samples from all background-related clusters (the judging criteria of



**FIGURE 10.** The background-anomaly separability maps of all methods on six HSI datasets. (a) Synthetic Dataset. (b) San Diego Airport 1. (c) San Diego Airport 2. (d) ABU-Airport. (e) ABU-Urban. (f) ABU-Beach.

**TABLE 2.** Computational time for different methods.

Time(s)	Synthetic Dataset	AVIRIS Dataset 1	AVIRIS Dataset 2	ABU-Airport	ABU-Urban	ABU-Beach
LRX	25.19	38.31	42.43	44.09	38.44	25.39
LKRX	140.60	60.91	56.01	206.93	90.52	100.03
CRD	<b>8.19</b>	<b>4.79</b>	<b>27.50</b>	<b>9.53</b>	<b>16.56</b>	<b>8.57</b>
KCRD	18.98	19.12	63.54	32.27	54.00	52.60
BJSRD	69.23	78.55	107.31	105.99	69.81	288.71
BEAWSR	48.12	54.25	55.13	52.39	51.92	47.76
Ours	62.55	71.35	70.19	67.24	67.19	80.62

background-related cluster is detailed in [26]) with the same proportion, denoted as Comparison A. (b) the global dictionary is formed by randomly selecting samples in the HSI, denoted as Comparison B. The parameters settings remain the same with the values in Section 4.2.

The experimental results are demonstrated by ROC curves in Fig. 11 and separability maps in Fig. 12. As can be seen from Fig. 11(a), the detection performance of the original proposed method is significantly superior than the

comparison methods. The performances of two comparison methods are close to each other. Better performance of Comparison A benefits from the K-means clustering algorithm so that the potential anomaly contamination in dictionary is removed. However, some background information may also be removed by the background cluster selection procedure in Comparison B. The results in Fig. 11(b) and Fig. 11(c) also draw the same conclusions from the above observations. Additionally, according to Fig. 11(c), the performance of



Comparison A is close to the original proposed method. The reason is that there are few scattered background components in ABU-beach dataset, the chances that background classes are falsely removed by Comparison A are slim. The conclusions drawn from the above are consistent with the observations in Fig. 12: (1) the original method achieves the best background-anomaly separation results and background suppression performance simultaneously. (2) the background suppression capability of Comparison A is slightly better than that of Comparison B. These experimental results illustrate that the proposed background estimation strategy can provide more representative and pure background information for global dictionary construction.

#### (2) The $l_{1/2}$ -norm regularization

In order to illustrate superiority of the  $l_{1/2}$ -norm regularization based local region approximation, the  $l_1$ -norm regularization for Problem (10) is adopted for comparison. The parameters settings remain the same as in Section 4.2. The results are presented as ROC curves in Fig. 13. It can be observed that in Fig. 13(a) and Fig. 13(c), the curves of the  $l_{1/2}$ -norm regularization based method are much closer to the upper left corner which indicates a significantly better detection performance. Moreover, as presented in Fig. 11(b), the PD of the  $l_{1/2}$ -norm regularization achieves 1 with the FAR less than  $10^{-1}$  while the  $l_1$ -norm regularization based method detects all anomalies when the FAR rises to 1. This advantage benefits from the sparser solutions yielded by the  $l_{1/2}$ -norm regularization based optimization model. It will lead to greater divergence for the AUP values of atoms in global dictionary, in which more representative atoms will be selected to form the sub-dictionary.

#### (3) The proposed re-weighting strategy

With the respect of the proposed re-weighting strategy, a comparison method is designed by removing the re-weighting strategy. The parameters settings remain the same as in Section 4.2. ROC curve is used for quantitative evaluation, as presented in Fig. 14. From all three figures, we can see that with the proposed background estimation strategy and the adaptive sub-dictionary, the comparison method can achieve comparable detection performances even without re-weighting strategy. After implementing the re-weighting strategy, the performances dramatically improve as the probabilities of detection rapidly reach 1 with the false alarm rates smaller than  $10^{-1}$ . The above observations confirm that the proposed SAD based re-weighting strategy can effectively enhance the detection results.

### D. PARAMETER ANALYSIS

In this section, the effectiveness of several significant parameters in our method are analyzed via experiments on three datasets used in Section 4.3. The threshold  $t$  and the number of abundance images  $s$  in background estimation, the regularization parameter  $\lambda$  in optimization problem (10), the regularization parameter  $\gamma$  in optimization problem (16), the proportion  $M$  of atoms in the global dictionary to build the

sub-dictionary, the percentage  $P$  of spectral vectors from each cluster to build the global dictionary and the dual-window size are analyzed in the follows.

The joint assessment of parameter  $t$  and  $s$  are demonstrated by AUC values as shown in Fig. 15. Both the increase of  $t$  and decrease of  $s$  will cause more true endmembers recognized as background-related endmembers. If  $t$  is too small and  $s$  is over certain value, anomaly-related endmembers will be falsely regarded as background-related endmembers so that the detection performance will deteriorate. This can be confirmed from the dotted-lines in three figures where both  $t$  and  $s$  are large. The lines at the top represent the best detection performance, which implies the corresponding values of  $s$  are the number of true background-related endmembers. As shown in Fig. 15(a), the number of real background-related endmembers is 4 since the green line is at the top position. It can be observed that for a certain  $t$ : (1)  $s$  greater than true number of background-related endmembers results in a low AUC value, which means anomalies are brought into estimated background. (2)  $s$  smaller than true number of background-related endmembers causes deterioration in detection performance, meaning that insufficient background information is included. For a certain  $s$ , two situations are observed: (1) if  $s$  is large, increase of  $t$  brings continuous improvement of detection performance. This phenomenon can be explained that a large  $s$  introduces anomalies in background estimation while increase of  $t$  helps removing anomaly-related endmembers. (2) if  $s$  is small, when  $t$  is over certain value, the criterion for determination of background-related endmembers is too strict. In this situation, some background information will be lost in the estimation, which will lead to a high false alarm rate. The above analysis corresponds with the observations in Fig. 15(b) and Fig. 15(c). According to the analysis, to achieve a satisfying performance, the optimal combination of  $(t, s)$  for three datasets are  $(0.8, 4)$ ,  $(0.8, 4)$ , and  $(0.8, 6)$ , respectively.

The quantitative evaluations of parameter  $\lambda$  and parameter  $\gamma$  are presented in Fig. 16(a) and Fig. 16(b), respectively. We can see that for parameter  $\lambda$  within the range of  $[0.001, 0.01]$ , the performances slightly improve as  $\lambda$  increases. However, when  $\lambda$  is greater than 0.01, the AUC values of all three datasets dramatically fall as  $\lambda$  increases. Therefore,  $\lambda$  better falls in the range of  $[0.001, 0.01]$ .

As can be observed from Fig. 16(b), the detection performances of San Diego Airport 2 and ABU-Airport improve slowly as  $\gamma$  increases from 0.001 and begin to deteriorate when  $\gamma$  is over 0.05. As for the experimental results for ABU-beach, the AUC value decreases as  $\gamma$  increases from 0.001 to 0.01. When  $\gamma$  falls in the range of  $[0.01, 0.05]$ , the AUC value begins to increase, and the performance begins to deteriorate at  $\gamma = 0.05$ . The above observations give the conclusion that it is optimal when  $\gamma$  is within the range of  $[0.001, 0.05]$  for San Diego Airport 2 and ABU-Airport. For ABU-beach dataset,  $\gamma$  should be set within the range of  $[0.01, 0.05]$ . Additionally, the AUC values in Fig. 16(b) changes within

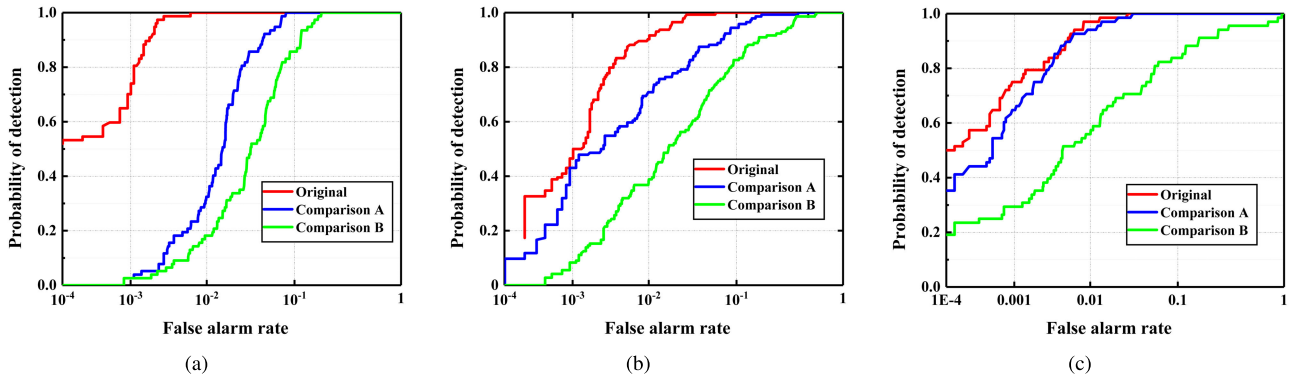


FIGURE 11. The ROC curves of background estimation strategy analysis experiments on three HSI datasets. (a) San Diego Airport 2. (b) ABU-Airport. (c) ABU-Beach.

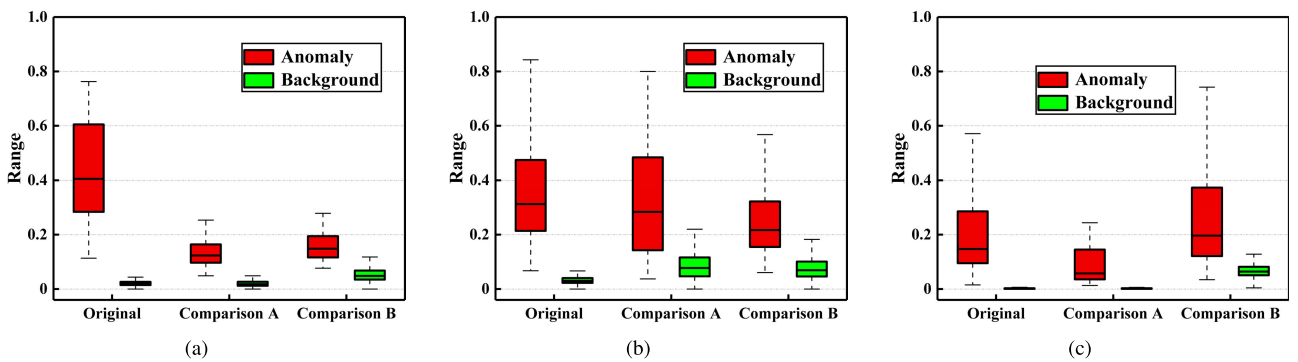


FIGURE 12. The separability maps of background estimation strategy analysis experiments on three HSI datasets. (a) San Diego Airport 2. (b) ABU-Airport. (c) ABU-Beach.

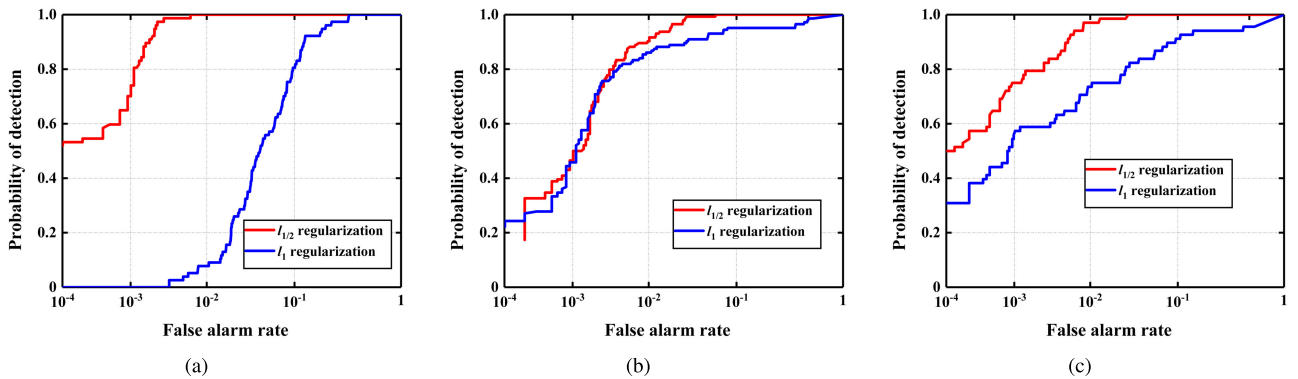


FIGURE 13. The ROC curves of the  $l_{1/2}$ -norm analysis experiments on three HSI datasets. (a) San Diego Airport 2. (b) ABU-Airport. (c) ABU-Beach.

the range of 0.03. This suggests that our method is robust to the change of parameter  $\gamma$ .

The effect of parameter  $M$  are evaluated by AUC as shown in Fig. 17.  $M$  is the proportion parameter in Section III.B, and the number of atoms  $N_B$  in the sub-dictionary  $\mathbf{B}$  is calculated by  $N_B = M \cdot B$ , where  $B$  is the band number of the input dataset. It can be observed from Fig. 17 that for three HSI datasets, when  $M$  is smaller than 0.5, the detection performance improves evidently as  $M$  increases. The improvement of detection performance dramatically slows down when

$M > 0.6$ . This indicates that when the number of atoms in sub-dictionary is over half of the band number, the improvement of detection performance is slight yet the computation cost increases significantly. In other words, when the number of sub-dictionary atoms is half the band number, it could provide most background information. Therefore, the parameter  $M$  is better set within the range of [0.5, 0.6].

The parameter  $P$  is the percentage of samples chosen from each cluster after the implementing K-means to the estimated background. The quantitative assessment of how the

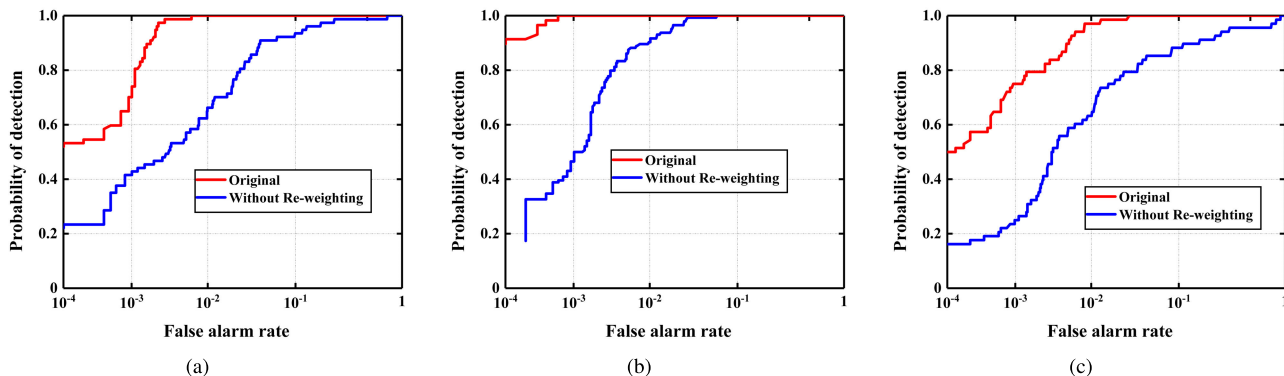


FIGURE 14. The ROC curves of the re-weighting strategy analysis experiments on three HSI datasets. (a) San Diego Airport 2. (b) ABU-Airport. (c) ABU-Beach.

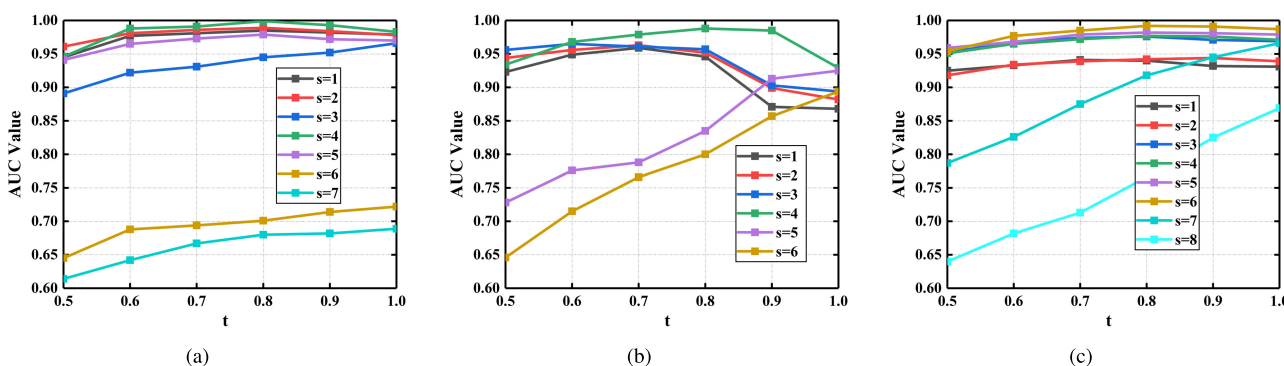


FIGURE 15. The joint consideration of the parameter effect of  $t$  and  $s$  on three HSI datasets via AUC values. (a) San Diego Airport 2. (b) ABU-Airport. (c) ABU-Beach.

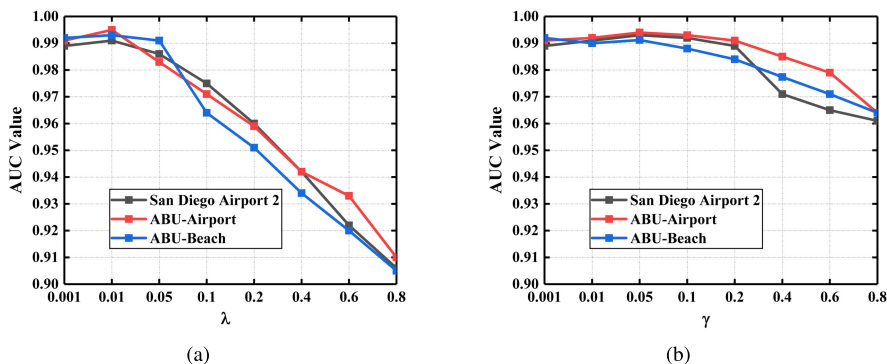


FIGURE 16. The illustration of the parameter effect of  $\lambda$  and  $\gamma$  on three HSI datasets via AUC values.

parameter  $P$  affects the detection performance is illustrated in Fig. 18. The experimental results in the figure show that for the AVIRIS San Diego Airport 2 dataset, the AUC value increases rapidly as  $P$  increases within  $[10,30]$ . However, when  $P$  is over 30, the increasement of AUC value significantly slows down. For two ABU datasets, the change of increasement rate happens when  $P = 40$ . Therefore, aiming to achieve better detection performance and save computational resource simultaneously, the parameter  $P$  is better set 30 for the San Diego Airport 2, and 40 for two ABU datasets.

The experimental results for the effect of the dual-window size are listed in Table 3 - Table 5, and the largest AUC values have been highlighted. As can be seen from Table 3, for the San Diego Airport 2 dataset, when the inner window size is set  $7 \times 7$  and the outer window size is set  $13 \times 13$ , the AUC value becomes the largest, which indicates that  $(7 \times 7, 13 \times 13)$  is the optimal dual-window size setting for the San Diego Airport 2. From the results in Table 4 and Table 5, it can be inferred by the AUC values that the optimal dual-window sizes for the ABU-Airport and the ABU-Beach are  $(7 \times 7, 13 \times 13)$  and

TABLE 3. The effect of dual-window size for the San Diego Airport Dataset 2.

$win_{in} \backslash win_{out}$	$5 \times 5$	$7 \times 7$	$9 \times 9$	$11 \times 11$	$13 \times 13$	$15 \times 15$	$17 \times 17$	$19 \times 19$
$3 \times 3$	0.8546	0.8789	0.8954	0.8981	0.9015	0.8884	0.8736	0.8526
$5 \times 5$		0.8995	0.9165	0.9284	0.9201	0.9087	0.8983	0.8632
$7 \times 7$			0.9820	0.9901	<b>0.9946</b>	0.9814	0.9745	0.9618
$9 \times 9$				0.9626	0.9710	0.9510	0.9212	0.9007
$11 \times 11$					0.9215	0.9026	0.8895	0.8465
$13 \times 13$						0.8993	0.8674	0.8458

TABLE 4. The effect of dual-window size for the ABU-Airport.

$win_{in} \backslash win_{out}$	$5 \times 5$	$7 \times 7$	$9 \times 9$	$11 \times 11$	$13 \times 13$	$15 \times 15$	$17 \times 17$	$19 \times 19$
$3 \times 3$	0.8989	0.9025	0.9126	0.9231	0.9126	0.9012	0.8945	0.8856
$5 \times 5$		0.9258	0.9346	0.9485	0.9498	0.9156	0.9052	0.8879
$7 \times 7$			0.9405	0.9452	0.9564	0.9764	0.9712	0.9412
$9 \times 9$				0.9854	<b>0.9914</b>	0.9781	0.9713	0.9648
$11 \times 11$					0.9650	0.9564	0.9264	0.8994
$13 \times 13$						0.9333	0.9135	0.8847

TABLE 5. The effect of dual-window size for the ABU-Beach.

$win_{in} \backslash win_{out}$	$5 \times 5$	$7 \times 7$	$9 \times 9$	$11 \times 11$	$13 \times 13$	$15 \times 15$	$17 \times 17$	$19 \times 19$
$3 \times 3$	0.9727	0.9784	0.9798	0.9654	0.9565	0.9364	0.9215	0.9005
$5 \times 5$		0.9813	0.9926	<b>0.9894</b>	0.9745	0.9712	0.9681	0.9594
$7 \times 7$			0.9715	0.9703	0.9612	0.9510	0.9423	0.9216
$9 \times 9$				0.9645	0.9515	0.9345	0.9210	0.9125
$11 \times 11$					0.9512	0.9326	0.9154	0.9002
$13 \times 13$						0.9313	0.9131	0.8897

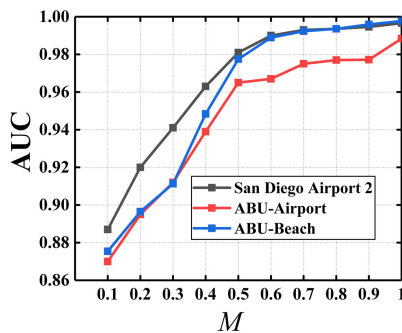


FIGURE 17. The illustration of the parameter effect of  $M$  on three HSI datasets via AUC values.

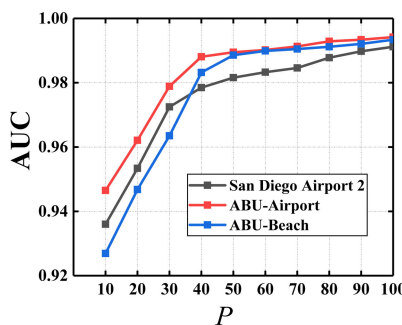


FIGURE 18. The illustration of the parameter effect of  $P$  on three HSI datasets via AUC values.

( $7 \times 7$ ,  $13 \times 13$ ), respectively. Furthermore, it can be observed that as long as the inner window sizes are set slightly larger than the sizes of the anomalies, the detection performance is not very sensitive to the change of the outer window sizes.

This is owing to the proposed sub-dictionary construction method, which can effectively eliminate the interference from the anomaly contamination in the local neighbor regions.

### V. CONCLUSION

In this paper, a novel SR based hyperspectral anomaly detection method via adaptive background sub-dictionaries is proposed. Firstly, an SMACC endmember extraction model based background estimation strategy is proposed to extract a representative and pure estimated background. Then, based on the estimated background, a global dictionary is constructed by utilizing the K-means clustering algorithm. Next, several active atoms are selected from this global dictionary to form a sub-dictionary to adaptively approximate local region in each dual-window. This strategy can help remove potential anomaly contamination in local regions. Finally, with the sub-dictionaries, a re-weighting strategy is proposed to enhance the performance of SR based anomaly detector. Experiments on one synthetic HSI dataset and five real-world HSI datasets are implemented with the proposed method and six comparison methods. The experimental results demonstrate that our method can accurately detect anomalies and effectively suppress background simultaneously. Additionally, experiments conducted on three real-world HSI datasets validate the superiority of the strategies proposed in our method, and testify the effectiveness of several significant parameters.

### ACKNOWLEDGMENT

The authors would like to thank the precious suggestions from the reviewers and the real-world hyperspectral datasets



provided by the researchers online. Specially, the first author would like to thank Miss X. Che for all the love, company and the support she has been giving.

## REFERENCES

- [1] C. Bo, H. Lu, and D. Wang, "Weighted generalized nearest neighbor for hyperspectral image classification," *IEEE Access*, vol. 5, pp. 1496–1509, 2017.
- [2] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, p. 67, Jan. 2017.
- [3] Y. Yang, S. Song, D. Liu, J. C.-W. Chan, J. Li, and J. Zhang, "Hyperspectral anomaly detection through sparse representation with tensor decomposition-based dictionary construction and adaptive weighting," *IEEE Access*, vol. 8, pp. 72121–72137, 2020.
- [4] Y. Qu, H. Qi, B. Ayhan, C. Kwan, and R. Kidd, "DOES multispectral/hyperspectral pansharpening improve the performance of anomaly detection?" in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 6130–6133.
- [5] 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, 1990.
- [6] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," *Proc. SPIE*, vol. 6966, Apr. 2008, Art. no. 696604.
- [7] 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.
- [8] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [9] A. P. Schaum, "Hyperspectral anomaly detection beyond RX," *Proc. SPIE*, vol. 6565, May 2007, Art. no. 656502.
- [10] Y. Zhang, Q. Wang, L. Xiao, and Z. Cui, "An improved two-step face recognition algorithm based on sparse representation," *IEEE Access*, vol. 7, pp. 131830–131838, 2019.
- [11] W.-Y. Xie, B.-D. Liu, S. Shao, Y. Li, and Y.-J. Wang, "Sparse representation and collaborative representation? Both help image classification," *IEEE Access*, vol. 7, pp. 76061–76070, 2019.
- [12] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3450–3459, Dec. 2012.
- [13] 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.
- [14] L. Zhu and G. Wen, "Hyperspectral anomaly detection via background estimation and adaptive weighted sparse representation," *Remote Sens.*, vol. 10, no. 2, p. 272, Feb. 2018.
- [15] 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.
- [16] 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.
- [17] W. Wang, S. Li, H. Qi, B. Ayhan, C. Kwan, and S. Vance, "Identify anomaly component by sparsity and low rank," Tech. Rep., 2015.
- [18] K. Tan, Z. Hou, D. Ma, Y. Chen, and Q. Du, "Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint," *Remote Sens.*, vol. 11, no. 13, p. 1578, Jul. 2019.
- [19] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [20] M. Vafadar and H. Ghassemian, "Hyperspectral anomaly detection using outlier removal from collaborative representation," in *Proc. 3rd Int. Conf. Pattern Recognit. Image Anal. (IPRIA)*, Apr. 2017, pp. 13–19.
- [21] Z. Wu, H. Su, and Q. Du, "Low-rank and collaborative representation for hyperspectral anomaly detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 1394–1397.
- [22] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [23] M. Hsueh and C.-I. Chang, "Field programmable gate arrays (FPGA) for pixel purity index using blocks of skewers for endmember extraction in hyperspectral imagery," *Int. J. High Perform. Comput. Appl.*, vol. 22, no. 4, pp. 408–423, Nov. 2008.
- [24] Y. Qu, W. Wang, R. Guo, B. Ayhan, C. Kwan, S. Vance, and H. Qi, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.
- [25] D. Ma, Y. Yuan, and Q. Wang, "Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation," *Remote Sens.*, vol. 10, no. 5, p. 745, May 2018.
- [26] Y. Yang, J. Zhang, S. Song, and D. Liu, "Hyperspectral anomaly detection via dictionary construction-based low-rank representation and adaptive weighting," *Remote Sens.*, vol. 11, no. 2, p. 192, Jan. 2019.
- [27] D. Ma, Y. Yuan, and Q. Wang, "A sparse dictionary learning method for hyperspectral anomaly detection with capped norm," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 648–651.
- [28] Y. Yang, J. Zhang, D. Liu, and X. Wu, "Low-rank and sparse matrix decomposition with background position estimation for hyperspectral anomaly detection," *Infr. Phys. Technol.*, vol. 96, pp. 213–227, Jan. 2019.
- [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. 65, no. 11, pp. 6497–6504, Nov. 2016.
- [30] C. Gonzalez, S. Lopez, D. Mozos, and R. Sarmiento, "FPGA implementation of the HySime algorithm for the determination of the number of endmembers in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2870–2883, May 2015.
- [31] Z. Xu, X. Chang, F. Xu, and H. Zhang, " $L_{1/2}$  regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.
- [32] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.



**YI LU** received the B.Eng. degree from the School of Optoelectronics, Beijing Institute of Technology, Beijing, China, and the M.E. degree from the School of Air Defense and Anti-Missile, Air Force Engineering University, Xi'an, China, where he is currently pursuing the master's degree with the School of Graduate. His research interests include target detection, machine learning, computer vision, and compressed sensing.



**SHUCAI HUANG** received the B.Eng., M.E., and Ph.D. degrees from the School of Air Defense and Anti-Missile, Air Force Engineering University, Xi'an, China. His research interests include computer vision, pattern recognition, and artificial intelligence.

...