Self-Adaptive Low-Rank and Sparse Decomposition for Hyperspectral Anomaly Detection

Qunming Wang, Jiang Zeng, Hao Wu, Jiawen Wang, and Kaipeng Sun

***Abstract*—Hyperspectral anomaly detection is a widely used technique for exploring target of interest in hyperspectral images (HSIs). In recent years, the low-rank and sparse-decomposition-based anomaly detection model has attracted extensive attention. However, these models suffer from two main problems. First, it is difficult for them to completely separate the low-rank background and the sparse anomaly. Moreover, the extracted sparse component is inevitably contaminated by noise. Second, the incorporation of various constraints increases the cost of selecting the optimal parameters. To solve the two key problems, we propose a self-adaptive low-rank and sparse decomposition (SLaSD) method for hyperspectral anomaly detection in this article. The proposed method decomposes the sparse (anomaly) part of the HSI through a novel self-adaptive alternating direction method (S-ADM). The noise of the sparse part is suppressed through a dual strategy of integrating guided filter and the difference between the S-ADM-derived sparse features of pixels. The performance of the proposed method is evaluated by comparing with ten state-of-the-art methods using six real HSIs. It is shown that the proposed SLaSD method can produce more accurate detection results than the ten benchmark methods.**

***Index Terms*—Anomaly detection, hyperspectral images (HSIs), low-rank and sparse decomposition, noise suppression, self-adaptive.**

I. Introduction

HYPERSPECTRAL remote sensing is an important technology for acquiring land surface information [1]. Compared with the images obtained by other platforms, hyperspectral image (HSI) has the characteristics of high spectral resolution, which can obtain nearly continuous spectral curves of ground objects [2]. Therefore, HSI has unique advantages in many application fields, such as in abnormal target detection [3], [4], spectral unmixing [5], [6], land cover mapping [7], [8], and change detection [9], [10].

Anomalous target detection is an important application field for HSI. In practical applications, it can be divided into two types: target detection and anomaly detection according to whether there is prior information available. The former is a target matching recognition technique, whereas the latter is an abnormal target identification technique [11]. In contrast to target detection, anomaly detection does not require any prior knowledge on the target or the background, which focuses on detecting targets with abnormal features (usually in terms of spectral signatures) that are different from the surrounding background. Generally, anomalies are assumed to have more distinguishable characteristics, such as small spatial size, low occurrence probability, and discrete distribution. For situations where it is difficult to obtain prior information (e.g., spectra of targets, especially for those at subpixel-level), anomaly detection has an irreplaceable role. Therefore, HSI-based anomaly detection has important values in a number of scenarios, such as in mineral exploration [12], national security supervision (military aircraft detection) [13], and small-sized crop detection [14]. In recent decades, with the continuous development of hyperspectral remote sensing technology, an increasing number of anomaly detection methods have been proposed [15], which can mainly be grouped into two types: statistics-based and representation-based.

The statistics-based detection methods obey specific statistical distribution assumptions. This type of methods determines whether a pixel is abnormal by calculating the likelihood ratio test value of the pixel. Reed–Xiaoli (RX) proposed in [16] has been regarded as a typical method of this type. The RX method is undertaken based on the premise that the background obeys the multivariate normal distribution. Specifically, it calculates the Mahalanobis distance between the tested pixel and the background pixel to obtain the likelihood ratio test value, and then generates the probability abundance map of detection. There are two main versions of RX: the RX detector (RXD) and the local RXD (LRXD). The difference between them lies in the way the background pixels are selected. The former uses the global image to calculate the average value and covariance matrix, whereas the latter uses a sliding dual window to calculate the average value and covariance matrix. In [17] and [18], the local kernel RX and local cluster kernel RXDs were proposed to project the original HSI into a nonlinear high-dimensional subspace. In addition, some other variants based on the RX method were proposed in [19]–[22]. Meanwhile, some other methods based on different statistical models have also been developed, such as the blocked adaptive computationally efficient outlier nominators [23], the discriminative metric learning based anomaly detector [24], and the probabilistic anomaly

Manuscript received January 3, 2022; revised March 1, 2022, April 1, 2022, and April 23, 2022; accepted April 29, 2022. Date of publication May 3, 2022; date of current version May 18, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 42171345 and Grant 41971297 and in part by the Tongji University under Grant 02502350047. *(Corresponding author: Hao Wu.)*

Qunming Wang and Jiang Zeng are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: wqm11111@126.com; zengj1480@163.com).

Hao Wu is with the College of Urban and Environmental Sciences, Central China Normal University, Wuhan 430079, China (e-mail: haowu@mail.ccnu.edu.cn).

Jiawen Wang and Kaipeng Sun are with the Shanghai Institute of Satellite Engineering, Shanghai 201109, China.

Digital Object Identifier 10.1109/JSTARS.2022.3172120

detector [25]. Generally, the aforementioned statistics-based methods suffer from two main problems. The first is that they involve the inverse computation of the covariance matrix, and the covariance matrix was susceptible to anomalies and noises, which may result in ill-conditioned matrix problems. The second problem is that heterogeneous and complex backgrounds of HSI are ubiquitous, and thus, it is difficult for the background or anomaly to obey the premise of a certain statistical distribution. Therefore, the statistics-based detection methods are sensitive to image quality, whose performances can be compromised when the background is complex.

The second type of methods is representation-model-based, which can further be divided into two main classes. The first class assumes that each background pixel can be represented by a background dictionary, whereas the anomaly pixel cannot. Two common examples are the collaborative representation-based detector (CRD) and kernel CRD (KCRD) proposed in [26]. They are performed based on a sliding dual-window strategy, which selects pixels between the inner and outer windows to constitute the background dictionary. Furthermore, in contrast to the linear representation strategy in CRD, KCRD projects the representation process to a higher dimensional subspace. Different from CRD, a local sparsity divergence-based detector (LSD) was proposed in [27], which considers additionally local sparse and spatial constraints. In [28], the background joint sparse representation detector reconstructs a robust background dictionary by extracting background endmembers from the abundant background information.

The other class of representation-model-based method is the low-rank and sparse-decomposition-based method, which assumes that the background could be projected to a low-dimensional subspace, whereas the anomaly is usually sparsely distributed in the entire image. That is, the background has low-rank characteristics and the anomaly has sparse characteristics. The low-rank and sparse decomposition theory was first used in the field of computer vision, such as in video monitoring, image denoising, and deshading [29]. In recent years, a series of low-rank and sparse-decomposition-based anomaly detection methods has been proposed. For example, in [30], a robust principal component analysis (RPCA) method that uses singular value decomposition (SVD) to construct the low-rank and sparse-decomposition-based model was developed for hyperspectral anomaly detection. In [31], [32], a spectral unmixing and dictionary-based low-rank method was proposed, which considers the use of abundance vectors for low-rank representation. In [33], [34], the low-rank and sparse matrix decomposition-based (LRaSMD) method and the LRaSMD-based Mahalanobis distance (LSMAD) method apply the Go-decomposition (GoDec) algorithm to implement low-rank and sparse decomposition. Different from LRaSMD, LSMAD employs the bilateral random projection and Mahalanobis distance strategies, which increases the iterative convergence speed and enhances the ability in background suppression. In addition to the enthusiasm for building low-rank and sparse models, dictionary construction has also been a prevalent choice. For example, in [35], to reconstruct a background dictionary, the k-means method was used to cluster the original images into K classes and to select reliable background atoms according to the difference between the spectra of pixels. In [36], a dual mean-shift-cluster method was proposed instead, which filters out the possible abnormal pixels by two asynchronous long shift clusters and then uses the pixel nearest to the cluster center to construct the background dictionary. Generally, different HSIs have different spatial distribution characteristics. Thus, it has also been a widespread research concern to add geospatial constraints to enhance the reliability of the low-rank and sparse model. For example, in [37], a graph and total variation regularized low-rank representation method was proposed. Compared with the traditional low-rank and sparse method, it includes a spatial constraint for expressing the surface geometric relationship. In [38], local spatial constraints and total variation were also applied to anomaly detection.

Recently, deep-learning-based methods have also been developed for hyperspectral anomaly detection. These methods are a vital complement to the statistics-based and representation-based methods. For example, in [39], a generative adversarial network (GAN) was used to reconstruct the background, and a statistical method called RXD was used for detection based on the background reconstructed by GAN. The method in [40] uses an adversarial autoencoder for background endmember learning and uses a multiconstrained representation method to complete detection. In [41], the proposed method uses a convolutional neural network to reconstruct anomaly features. Generally, the deep-learning-based strategies can effectively explore feature information in HSI, and thus, have potential to enhance the ability to separate anomalies and backgrounds. However, deep-learning-based methods have to prepare train datasets (usually for background) before detection and the training process always involves laborious parameter tunning, which are significantly different from statistics-based and representation-based methods.

The low-rank and sparse-decomposition-based methods have shown their advantages in anomaly detection (e.g., lower false alarms) over the statistics-based methods in some cases, especially when the background in the image is relatively complex. However, the low-rank and sparse-decomposition-based model also has two limitations. First, when some of the background classes appear with a very small probability in the image, they may show sparseness. In addition, the sparse component decomposed by the model inevitably contains noise, this is, inherent uncertainty in background suppression. Actually, in real applications of anomaly detection, the background suppression ability is an important consideration for detection performance [42]. Second, since different types of geospatial constraints are considered in the model, the computing cost of the model is increased and the added parameters also increase the uncertainty of the model, both of which restrict its ability in practical applications.

To deal with the aforementioned problems, this article proposes a self-adaptive low-rank and sparse decomposition (SLaSD) detector for hyperspectral anomaly detection. This method is also based on the assumption that the background and anomaly obeys low-rank and sparse characteristics, respectively. Specifically, the original image is decomposed into two parts:

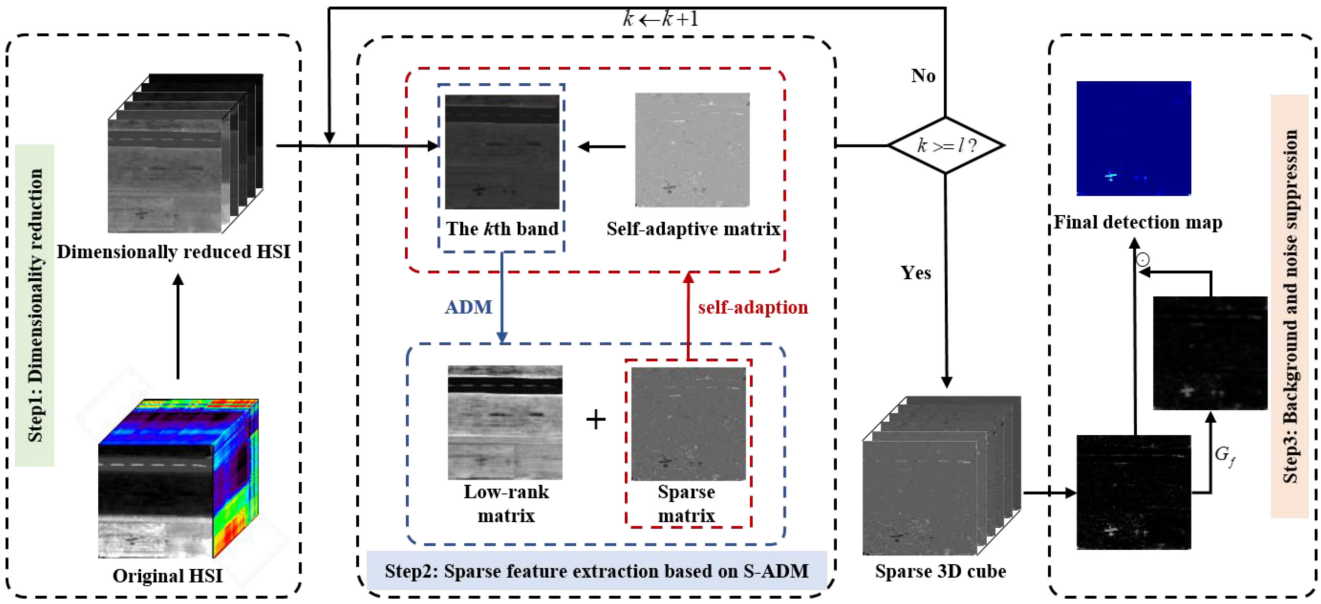


Fig. 1. Schematic of the proposed SLaSD-based anomaly detection method.

low-rank background and sparse anomaly. In contrast to traditional models based on low-rank and sparse decomposition, the proposed SLaSD detector reduces the dimensionality of HSI to a relatively lower dimension for eliminating data redundancy. Moreover, a novel self-adaptive alternating direction method (S-ADM) is developed to obtain the sparse component. Meanwhile, a dual strategy of integrating guided filter [43] and the difference between the spectra (composed of S-ADM-derived sparse feature in each band) of pixels is proposed to solve the problem that the sparse component is inevitably contaminated by the background and noise. The main contributions of this article are summarized as follows.

1) Spectral dimensionality reduction is considered in the proposed SLaSD detector. SLaSD not only reduces the amount of data involved in calculation, but also maintains the geospatial structure of the original HSI data in the calculation process. This is different from traditional low-rank and sparse-decomposition-based model that converts the original three-dimensional data into two-dimensional data for processing, which ignores the spatial structure of HSI and also increases the computation cost in matrix calculation.
2) A self-adaptive alternating iterative optimization strategy is proposed to fully explore the sparse features obtained by decomposition of the low-dimensional data. This can help to make fuller use of the available information presented in low-dimensional data, which suffer inevitable information loss relative to original HSI.
3) For the issue of contamination (noise and background) in the sparse component obtained by decomposition, the dual strategy of integrating guided filter and the difference between the sparse features of pixels is proposed to enhance the suppression in the background and, moreover, the generalizability for dealing with data with complex texture.

The rest of this article is organized as follows. Section II provides a detailed description of the proposed SLaSD detector. In Section III, the experimental results based on six HSI datasets are described and analyzed, including the comparison with ten widely used benchmark methods. Section IV discusses several key issues. Finally, Section V concludes this article.

II. Methods

The proposed SLaSD detector is mainly composed of three steps. First, the high-dimensional HSI data are reduced into low-dimensional HSI data by dimensionality reduction. Second, based on the assumption that the background has low-rank characteristics and the anomaly has sparse characteristics, the augmented Lagrange multiplier (ALM) model is used to construct the functional model of low-rank and sparse decomposition. The sparse component in the functional model is solved by the proposed S-ADM solution. Third, the dual strategy of integrating guided filter and the difference between the S-ADM-derived sparse features of pixels is applied to obtain the final anomaly detection map. A schematic of the proposed SLaSD detector is shown in Fig. 1. The details are described as follows.

A. Dimensionality Reduction

Due to the high spectral resolution of HSI, the difference in ground features can be distinguished using a larger number of features (i.e., by comparing continuous spectral properties). Therefore, in the field of anomaly detection, especially in exploration and recognition of subtle targets, HSI has irreplaceable advantages. However, the redundancy of high-dimensional data increases the computational cost of data processing and,

furthermore, the complexity of the detection model, which may reduce the generalizability of the model in practical applications [44]. Meanwhile, due to the hardware limitations of the existing hyperspectral remote sensing sensors and the interference of the imaging environment, the acquired HSIs are inevitably contaminated by noise. Therefore, in practical applications, it has been a common choice to distil low-dimensional data from the original data to simplify the calculation process and increase the robustness to noise [45].

To cope with the computational complexity caused by data redundancy, this article applies a commonly used band average method [46]. Specifically, all bands of the original HSI (denoted as $\mathbf{X}$) are divided evenly into l subgroups, each of which has N bands. The low-dimensional data are produced as follows:

$$\mathbf{Y}_k = \frac{\sum_{n=1}^{N} \mathbf{X}_k^n}{N}, \ k = 1, \ldots, l \tag{1}$$

where $\mathbf{X}_k^n$ is the nth band in the kth subgroup of the original HSI and $\mathbf{Y}_k$ is the fused single-band data of the kth subgroup.

B. Sparse Feature Extraction Based on S-ADM

RPCA is a low-rank and sparse matrix decomposition model that has been widely applied to image processing. For example, RPCA can be used to remove noises and low signal-to-noise ratio (SNR) shadows in the raw image [47]. It is performed based on the assumption that the noise conforms to discretely sparse characteristics. The raw image can be decomposed into two components: the low-rank component and the sparse component. The sparse component is the noise that needs to be removed, whereas the low-rank component is the restored image. Compared with principal component analysis (PCA), which is more suitable for removing simple noise, RPCA has stronger image denoising ability when the raw image is contaminated by complex noise (e.g., an integration of different types of noises) because most noises conform to a discretely sparse distribution.

In the field of anomaly detection, compared with other low-rank and sparse decomposition models, the RPCA model not only has simple structure of the constraints, but also has satisfactory decomposition ability. Thus, RPCA is used for low-rank and sparse decomposition in this article. The details are introduced as follows.

Suppose the original HSI data $\mathbf{M}$ are a 3-D image cube with a size of $m \times n \times b$, where m and n are the length and width of a single band image, and b is the number of bands. To reduce data redundancy and maintain the spatial structure simultaneously in each band, the original HSI data $\mathbf{M}$ are reduced to the low-dimensional data $\overline{\mathbf{M}}$with a size of $m \times n \times l$ by (1), where l is the number of bands after dimensionality reduction. This is different from the existing schemes that transform each 2-D band image to 1-D vector (i.e., the low-dimensional data are presented in a 2-D matrix with a size of $mn \times l$) to facilitate the calculation. Accordingly, the low-dimensional data $\overline{\mathbf{M}}$ are characterized as follows:

$$\overline{\mathbf{M}} = \mathbf{L} + \mathbf{S} \tag{2}$$

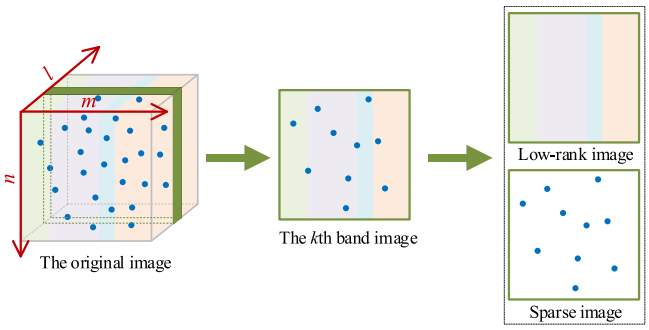


Fig. 2. Low-rank and sparse decomposition model for the kth band in HSI.

where $\mathbf{L}$ and $\mathbf{S}$ are the low-rank and the sparse components, respectively, both with the same size of $m \times n \times l$. The relation between the two parts is displayed in Fig. 2. For (2), the decomposition is transformed into the following optimization problem:

$$\min_{\mathbf{L}_k, \mathbf{S}_k} \text{rank}(\mathbf{L}_k) + \lambda \|\mathbf{S}_k\|_0, \text{s.t.} \ \overline{\mathbf{M}} = \mathbf{L} + \mathbf{S} \tag{3}$$

where $\text{rank}(\cdot)$ is the rank of $\mathbf{L}_k$, $\|\cdot\|_0$ is the sparsity of $\mathbf{S}_k$, k ($k = 1, \ldots, l$) represents the processing of the kth band, and $\lambda > 0$ is the penalty parameter controlling the contribution of $\mathbf{S}_k$ in the optimization process. However, (3) is a nonconvex optimization problem that cannot be solved directly. Thus, its convex optimization is considered, which is shown as follows:

$$\min_{\mathbf{L}_k, \mathbf{S}_k} \|\mathbf{L}_k\|_* + \lambda \|\mathbf{S}_k\|_1, \text{s.t.} \ \overline{\mathbf{M}} = \mathbf{L} + \mathbf{S} \tag{4}$$

where $\|\cdot\|_*$ is the nuclear norm and $\|\cdot\|_1$ is the $l_1-\text{norm}$. (4) is a convex optimization problem that can be solved directly.

To solve (4), the ALM [48] function is constructed accordingly. Then, the low-dimensional data $\overline{\mathbf{M}}$ are decomposed into low-rank and sparse components through the proposed S-ADM solution. The details are described as follows.

Based on ALM, the low-rank and sparse model in (4) is transformed into the following augmented Lagrange function:

$$\begin{aligned} &F(\mathbf{L}_k, \mathbf{S}_k, \mathbf{V}) \\ &= \|\mathbf{L}_k\|_* + \lambda \|\mathbf{S}_k\|_1 + \langle \mathbf{V}, \mathbf{J} - \mathbf{L}_k - \mathbf{S}_k \rangle + \frac{\rho}{2} \|\mathbf{J} - \mathbf{L}_k - \mathbf{S}_k\|_F^2 \end{aligned} \tag{5}$$

where $\mathbf{V}$ is the Lagrange multiplier matrix and ρ is the tradeoff parameter. $\mathbf{J}$ is the weight optimization matrix, which is calculated using a self-adaptive scheme

$$\mathbf{J} = \mathbf{W}_k^{\ln(1+\gamma) - \ln(1-\gamma)} \odot \overline{\mathbf{M}}_k. \tag{6}$$

In (6), $\odot$ is the Hadamard product operator, $\mathbf{W}_k$ is the weight matrix, which is equal to $\mathbf{S}_{k-1}$, $\gamma \in (0, 1)$ is a parameter, and $\ln(1+\gamma) - \ln(1-\gamma)$ is used to expand the range of γ to $(0, +\infty)$. As shown in (6), the extracted sparse component of the (k-1)th band is used adaptively for the next (i.e., the kth) band, which can be used for the self-adaptive matrix of anomalies band by band. An example is provided in Fig. 3 for illustration for the self-adaptive process.

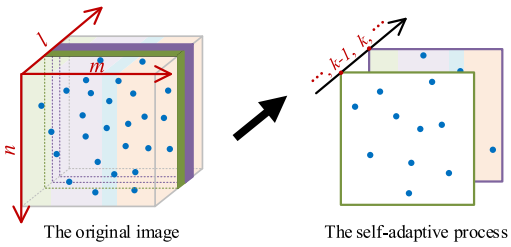


Fig. 3. Self-adaptive process for extracting the sparse component from the bands.

For each band, the minimization problem in (5) is a convex optimization problem. There are several unknown variables $\mathbf{L}$, $\mathbf{S}$, and $\mathbf{V}$, which cannot be predicted directly to obtain the global optimal solution of $F(\mathbf{L}_k,\mathbf{S}_k,\mathbf{V})$. Therefore, the S-ADM is proposed to cope with this issue, which is implemented using an alternate iteration scheme based on the defined self-adaptive model in (6). The global minimization of $F(\mathbf{L}_k,\mathbf{S}_k,\mathbf{V})$ is obtained through the iterative descent of the variables. Specifically, (5) is decomposed into three subproblems, which are solved as follows.

1) The first step is to fix other variables and update $\mathbf{L}$. The prediction for $\mathbf{L}$ can be transformed into the following problem:

$$\begin{aligned}\mathbf{L}_k^{(t+1)} &= \arg\min_{\mathbf{L}_k} \left\|\mathbf{L}_k^{(t)}\right\|_* + \left\langle \mathbf{V}^{(t)}, \mathbf{J}-\mathbf{L}_k^{(t)}-\mathbf{S}_k^{(t)}\right\rangle \\ &\quad + \frac{\rho}{2}\left\|\mathbf{J}-\mathbf{L}_k^{(t)}-\mathbf{S}_k^{(t)}\right\|_F^2 \end{aligned} \tag{7}$$

where t is the number of iterations. The subproblem in (7) can be solved by the following:

$$\mathbf{L}_k^{(t+1)} \leftarrow \Theta_{1/\rho}(\mathbf{J}-\mathbf{S}_k^{(t)}+\rho^{-1}\mathbf{V}^{(t)}) \tag{8}$$

where Θ is the singular value threshold operator and it is calculated as follows:

$$\begin{aligned}&\Theta_\tau(\mathbf{Y}) = \mathbf{U}\Omega_\tau(\mathbf{\Lambda})\mathbf{V}^*, \quad \text{s.t.}\,\mathbf{Y}=\mathbf{U}\mathbf{\Lambda}\mathbf{V}^* \\ &\Omega_\tau(\mathbf{\Lambda}) = \text{sgn}(\mathbf{\Lambda})\max(|\mathbf{\Lambda}|-\tau,0).\end{aligned} \tag{9}$$

In (9), $\mathbf{Y}=\mathbf{U}\mathbf{\Lambda}\mathbf{V}^*$ presents an SVD process and $\text{sgn}(\cdot)$ is a symbolic function.

1) The second step is to fix other variables and update $\mathbf{S}$. The prediction for $\mathbf{S}$ can be transformed into the following problem:

$$\begin{aligned}\mathbf{S}_k^{(t+1)} &= \arg\min_{\mathbf{s}_k} \lambda\left\|\mathbf{S}_k^{(t)}\right\|_1 + \left\langle \mathbf{V}^{(t)}, \mathbf{J}-\mathbf{L}_k^{(t+1)}-\mathbf{S}_k^{(t)}\right\rangle \\ &\quad + \frac{\rho}{2}\left\|\mathbf{J}-\mathbf{L}_k^{(t+1)}-\mathbf{S}_k^{(t)}\right\|_F^2. \end{aligned} \tag{10}$$

The problem in (10) can be calculated by the following:

$$\mathbf{S}_k^{(t+1)} \leftarrow \Omega_{\lambda/\rho}(\mathbf{J}-\mathbf{L}_k^{(t+1)}+\rho^{-1}\mathbf{V}^{(t)}) \tag{11}$$

where Ω is the shrinkage operator and it is calculated by (9).

1) The third step is to fix other variables and update $\mathbf{V}$, which is calculated as follows:

$$\mathbf{V}^{(t+1)} \leftarrow \mathbf{V}^{(t)} + \rho(\mathbf{J}-\mathbf{L}_k^{(t+1)}-\mathbf{S}_k^{(t+1)}). \tag{12}$$

The iteration process of S-ADM is terminated when the predefined maximum number of iterations $t_{\max}$is reached or the convergence condition defined in (13) is satisfied

$$\|\mathbf{J}-\mathbf{L}_k-\mathbf{S}_k\|_F/\|\mathbf{J}\|_F < \varepsilon. \tag{13}$$

Through the proposed S-ADM, the dimensionality reduced HSI data $\overline{\mathbf{M}}$ can be decomposed into the low-rank background component $\mathbf{L}$ and sparse anomaly component $\mathbf{S}$.

C. Background and Noise Suppression

The sparse anomaly component $\mathbf{S}$ does not only contain anomaly information but also inevitable noise and background contamination. For this situation, a dual strategy of integrating guided filter and the difference between the S-ADM-derived sparse features of pixels is proposed to suppress the background and noise.

Although sporadic backgrounds exist inevitably in the S-ADM-derived sparse anomaly component $\mathbf{S}$, their features are closer to the first-order statistics of $\mathbf{S}$ than sparse anomalies and noises. Therefore, the interference of the backgrounds can be removed by the following:

$$\mathbf{E}_{(x,y)} = \left\|\mathbf{S}_{(x,y,:)}-\mathbf{u}_\mathbf{S}\right\|_2^2 \tag{14}$$

where (x,y) is the 2-D spatial location of a pixel, $\mathbf{u}_\mathbf{S}$ is the mean of the S-ADM-derived sparse features for all pixels in $\mathbf{S}$, and $\mathbf{E}$ is the image after background suppression.

Usually, $\mathbf{E}$ contains a number of isolated sparse noises, which could lead to missed detection of anomalies. The spatial size of sparse noises is generally one pixel, which are significantly different from the sparse anomalies (assumed to be larger than one pixel in this article). It is important to remove these sparse noises while retaining the spatial texture of anomalies to the largest extent. The guided filter has shown its advantages in preserving edges of objects in images. Therefore, in this article, the guided filter is used to suppress the sparse noises and, moreover, a dual strategy is proposed as follows:

$$\mathbf{D} = G_f(\mathbf{E},\delta)\odot\mathbf{E} \tag{15}$$

where G_f is the operator of guided filter, δ is the penalty term of the energy constraint, and $\mathbf{D}$ is the final detection result.

The entire flow of the proposed SLaSD detector is summarized as Algorithm 1.

III. Experiments

Six HSI datasets were used in this section to evaluate the effectiveness of the proposed SLaSD method. The SLaSD method was compared with ten state-of-the-art methods, including two statistical model-based methods (RXD, LRXD [16], and LKRXD [17]), two representation-model-based methods (CRD [26] and LSD [27]), three low-rank and sparse decomposition model based methods (LRaSMD [33], LSMAD

Algorithm 1: The SLaSD Algorithm.

Input: a) the original data: $\mathbf{M} \in \mathbb{R}^{m\times n\times b}$
b) parameters: $l \in \mathrm{N}^{+}$, $\gamma \in (0,1)$
c) error threshold for termination: ε

Output: Detection map: $\mathbf{D}$

1. Obtain the dimensionality reduced data:

$$\mathbf{M}_{m\times n\times b} \xrightarrow{(2)} \overline{\mathbf{M}}_{m\times n\times l}$$

2. Initialize conditions:
Set the variables $\mathbf{L}_k^{(0)}$, $\mathbf{S}_k^{(0)}$,$\mathbf{V}^{(0)}$ to zero matrix;
Set the variables $\mathbf{W}$ to one matrix;
Set other parameters to empirical values $T_{\max} = 4e3$, $\lambda = \rho = 1/\sqrt{\max(m,n)}$, $\varepsilon = 1e-4$ and $\delta = 1e-3$.

3. Optimization for the ith band:

a). Obtain the adaptive weighted matrix $\mathbf{J}$ by Eq. (6);

b). Update the low rank matrix $\mathbf{L}_k^{(t)}$ by Eq. (8). Update the sparse matrix $\mathbf{S}_k^{(t)}$ by Eq. (11). Update the Lagrange multiplier $\mathbf{V}^{(t)}$ by Eq. (12). The above iterations are summarized as follows:

$$\mathbf{L}_k^{(0)} \overset{(11)}{\Rightarrow} \mathbf{S}_k^{(0)} \overset{(12)}{\Rightarrow} \mathbf{V}^{(0)} \overset{(8)}{\Rightarrow} \mathbf{L}_k^{(1)} \overset{(11)}{\Rightarrow} \cdots$$
$$\cdots \overset{(12)}{\Rightarrow} \mathbf{V}^{(t-1)} \overset{(8)}{\Rightarrow} \mathbf{L}_k^{(t)} \overset{(11)}{\Rightarrow} \mathbf{S}_k^{(t)} \overset{(12)}{\Rightarrow} \mathbf{V}^{(t)}$$

c). $t \leftarrow t+1$; if $t < T_{\max}$ or Eq. (13) is unsatisfied, go to step (b).

4. $k \leftarrow k+1$; repeat step 3 until $k = l$. Produce the final anomaly detection map by Eqs. (14) and (15).

[34], and LRASR [35]), ADLR [32] that integrate abundance vector and low-rank model, and LRCRD [36] that integrate a low-rank and sparse model with a representation model.

To evaluate the performance of each method, we analyzed visually the anomaly detection maps, the capability of separating anomaly and background, and the detection accuracy. The anomaly-background separation capability of each method was assessed using box-plots presenting the statistical information of the pixel values of anomaly and background classes in the detection map. For quantitative assessment, the receiver operating characteristic (ROC) and the corresponding area under the curve (AUC) were used. In addition, the 3D-ROC [42] was also considered, which represents the relationship between the discrimination threshold τ , the probability of detection P_d, and the false alarm rate P_f. The experimental platform was a computer with a 3.6-GHz CPU, 32-GB RAM, and 64-b Windows 10. All methods were implemented using MATLAB R2020a.

A. Datasets

For the six real HSIs, we eliminated the low SNR bands with noise and water absorption. The details are described as follows.

1) Texas Coast Dataset: The datasets contain two HSIs named Texas Coast 1 and Texas Coast 2, which were acquired by the AVIRIS in 2010. The coverage areas are both located at the Texas Coast, USA, and the spatial resolution is 17.2 m per pixel. After excluding the low SNR bands, 204 and 207 bands are retained for Texas Coast 1 and Texas Coast 2, respectively. Both datasets have a spatial size of 100×100 pixels, as shown in Fig. 4(a) and (b), where the false-color images and references of the anomaly maps are provided.

2) Los Angeles Dataset: The datasets contain two HSIs, that is, Los Angeles 1 and Los Angeles 2, which were acquired by the AVIRIS in 2010. They cover two different areas located at Los Angeles, USA. The spatial resolution is 7.1 m, and the wavelength interval is 0.43–0.86 μm. 205 bands are retained after excluding the low SNR bands. Both datasets have the same spatial size of 100×100 pixels, as shown in Fig. 4(c) and (d).

3) Pavia Dataset: The dataset contains a HSI covering an area in the Pavia city center in northern Italy, which was acquired by the reflective optics system imaging spectrometer sensor. The spatial resolution is 1.3 m. The wavelength interval is 0.43–0.86 μm. After removing the low SNR bands, 102 bands are retained, each with a spatial size of 150×150 pixels. The HSI is shown in Fig. 4(e).

4) Gulfport Dataset: The dataset was acquired by the AVIRIS in 2010. It covers an area in Gulfport, Southern, Mississippi, USA, as shown in Fig. 4(f). The spatial resolution is 3.4 m, and the wavelength interval is 0.40–2.50 μm. 191 bands are selected out after removing the low SNR bands. The spatial size is 100×100 pixels.

B. Anomaly Detection Performance

To ensure the fairness of the comparison, for each benchmark method, the results under the optimal parameter settings were chosen after a number of tests. For the proposed SLaSD method, the band number $l = 6$ was used as the empirical value and the adaptive weight ambiguity parameter γ was selected as 0.1, 0.7, 0.2, 0.9, 0.2, and 0.1 for the six datasets (the parameter selection of the proposed SLaSD method is detailed in Section III-C). Fig. 5 shows the anomaly detection maps of the 11 methods for the six HSI datasets. Generally, four points can be observed. First, the maps of the proposed SLaSD method are the closest to the reference compared to the other ten benchmark methods and have a more satisfactory background suppression effect. Second, LKRXD, LRaSMD, LRASR, and LRCRD outperform the other six benchmark methods, but their weaker background suppression ability lead to relatively larger false alarm rate. For example, the LRaSMD method produces bright background false alarms in the lower left corner of the Los Angeles 2 dataset, and LKRXD produces noticeable false alarms in the Pavia dataset. Third, RX, CRD, and LSMAD show weaker performance than LRaSMD, LRASR, and LRCRD. Their detection results present a number of missed anomalies and false alarms, especially when the background of the dataset is complex. For example, in the Los Angeles 2 and Pavia datasets, the ability of these methods to distinguish between anomaly and background is obviously weaker, and the recognition of anomalies is relatively patchy. Finally, LRX, ADLR, and LSD show the weakest detection performance among all methods. Specifically, a large number of anomalies are undetected by the LRX detector. The main reason for this is that the background is over suppressed.

Fig. 6 shows the box-plots of the statistical information of the anomaly and background in the detection maps for different

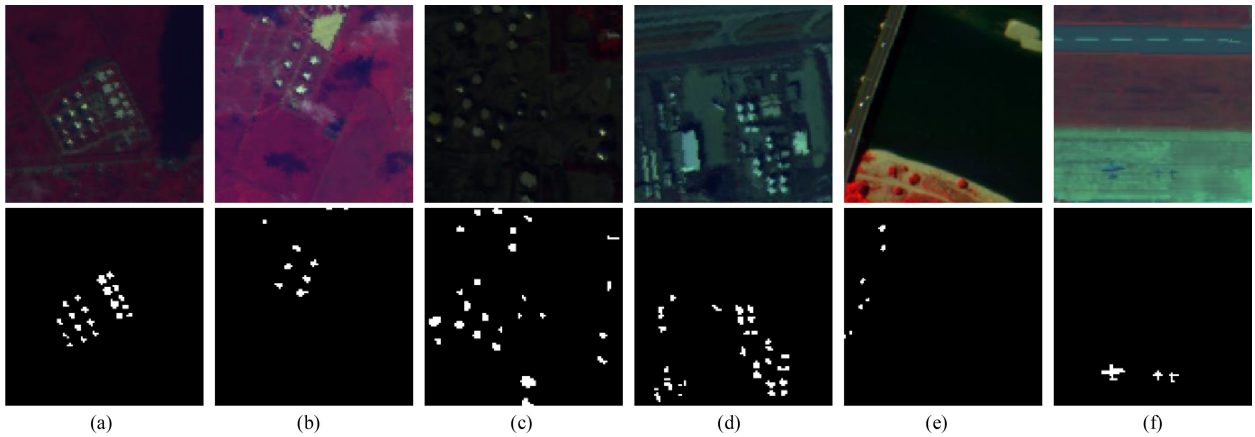


Fig. 4. Six HSIs used in the experiments. (a) Texas Coast 1. (b) Texas Coast 2. (c) Los Angeles 1. (d) Los Angeles 2. (e) Pavia. (f) Gulfport. The first and second rows show the false-color images of the HSIs and the references of anomaly maps.

methods. Overall, the methods show different separation effects for different datasets. For example, the box-plot values of the LRXD for the Texas Coast 1 dataset are very close to zero, and the boxes of the anomaly and background present a large amount of overlap, suggesting that the degree of separation is limited. However, for the Texas Coast 2 dataset, the LRXD suppresses the background to be close to zero. The anomaly and the background overlap only a small amount, and thus, the degree of separation is more satisfactory than that for the Texas Coast 1 dataset. Similar phenomenon can be observed for other methods. For the proposed SLaSD method, however, the background is suppressed to be much closer to zero for all six datasets, which is consistent with the visual check of the detection maps in Fig. 5, indicating that the proposed SLaSD method has superior ability for background suppression than the benchmark methods.

To comprehensively compare the detection accuracy of the methods, Fig. 7 shows the 3D-ROC curves of (P_d, P_f, τ) and the three derived 2D-ROC curves for each dataset. From the 2D-ROC curves of (P_d, P_f) in Fig. 7, it can be seen that in the Texas Coast, Los Angeles 2, and Gulfport datasets, the proposed SLaSD method has a larger probability of detection than the other methods under different false alarm rates. For the Los Angeles 1 dataset, when the false alarm rate is less than 1%, LSMAD has the largest probability of detection, but the proposed SLaSD has the largest probability of detection when the false alarm rate is larger than 1%. From the 2D-ROC curves of (P_d, τ) in Fig. 7, it can be seen that the ADLR method has a larger probability of detection overall. From the 2D-ROC curves of (P_f, τ) in Fig. 7, it can be seen that for the six datasets, the proposed SLaSD method has a very low false alarm rate under different thresholds τ. It should be pointed out that for the Texas Coast 1 and Los Angeles 1 datasets, although the false alarm rate of the LRXD method is lower than that of the proposed SLaSD method, from the 2D-ROC curves of (P_d, P_f) , it is seen clearly that the probability of detection for LRXD was lower correspondingly.

Table I shows the three AUC values derived from the three 2D-ROC curves. The three types of AUC values (i.e., (P_d, P_f) (P_d, τ), and (P_f, τ)) reflect the overall detection accuracy, anomaly detectability, and background suppression accuracy for the methods. As shown in Table I, the AUC values of (P_d, P_f) of the proposed SLaSD method for six datasets are 0.9994, 0.9983, 0.9969, 0.9869, 0.9837, and 0.9981, which are all larger than those of the other methods. The AUC value of (P_f, τ) also indicates that the SLaSD method has the greatest background suppression accuracy overall.

Based on the detection maps, box-plots of anomaly-background, 3D-ROC, and Table I, we can conclude that compared with the ten benchmark methods, the proposed SLaSD method shows more reliable performance in anomaly detection and background suppression.

C. Advantage of the Proposed Self-Adaptive Model

To show the advantage of the proposed self-adaptive model (i.e., S-ADM) for sparse feature extraction in SLaSD, it was compared with the classic ADM model. For the two models, the AUC values of (P_d, P_f) for all six datasets are shown in Table II. It is seen clearly that the AUC values of the proposed S-ADM model is larger than that of the classic ADM model, suggesting that S-ADM is a more accurate solution for sparse feature extraction.

D. Robustness to Noise

To analyze the detection performance of the proposed SLaSD method in relation to noises, four levels of zero-mean Gaussian noises (with SNRs of 30, 25, 20, and 15 dB) were added to the Texas Coast 1 dataset, as shown in Fig. 8. The proposed SLaSD and ten benchmark methods were used for detection based on the four datasets in Fig. 8, and the results are shown in Fig. 9. It can be seen that SLaSD, LRASR, and LRCRD were much more robust to noise.

E. Parameter Analysis

The SLaSD method involves two main parameters: the number of bands after dimensionality reduction (denoted as l) and the

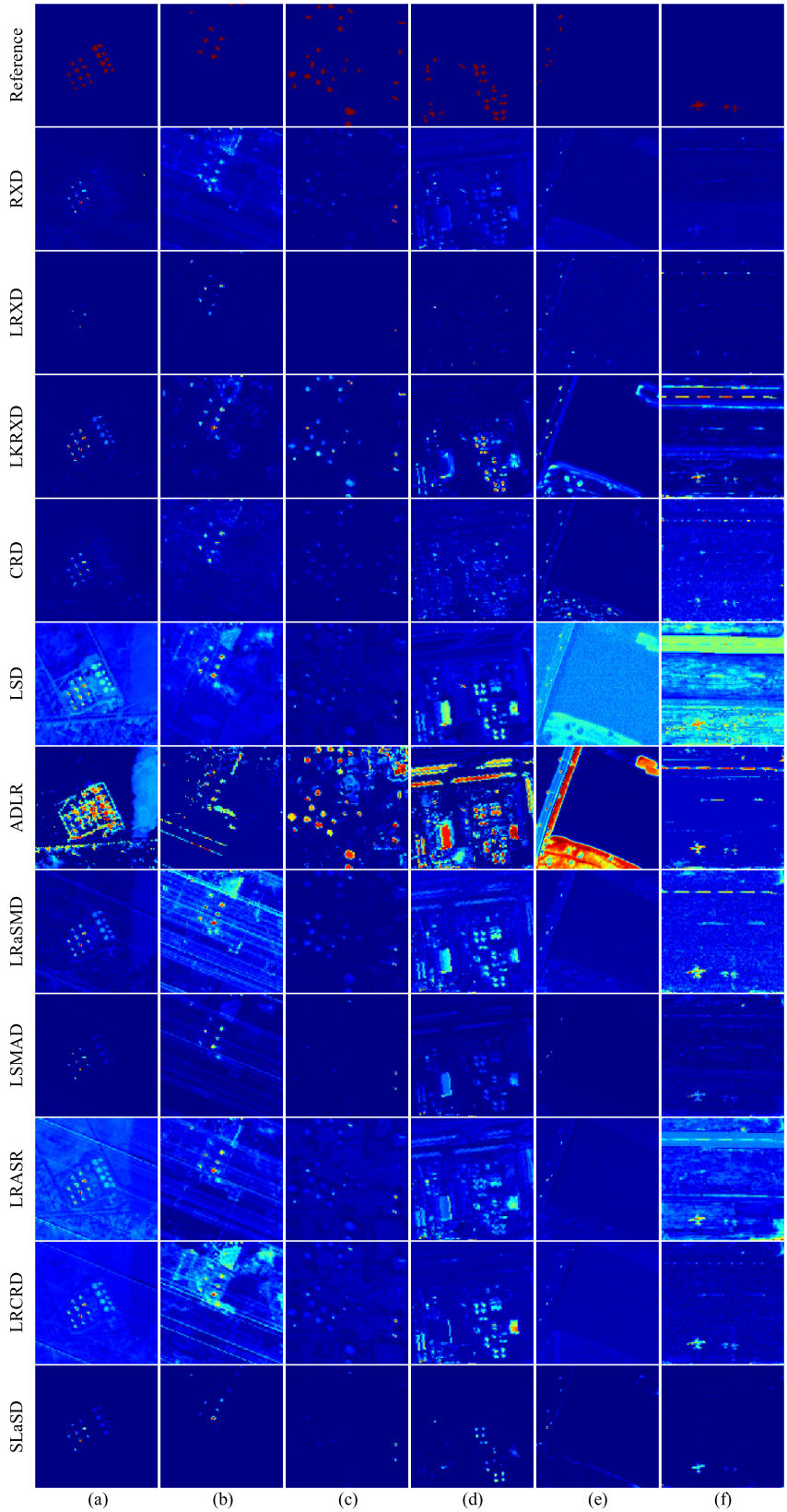


Fig. 5. Anomaly detection maps of different methods for the six HSIs. (a) Texas Coast 1. (b) Texas Coast 2. (c) Los Angeles 1. (d) Los Angeles 2. (e) Pavia. (f) Gulfport.

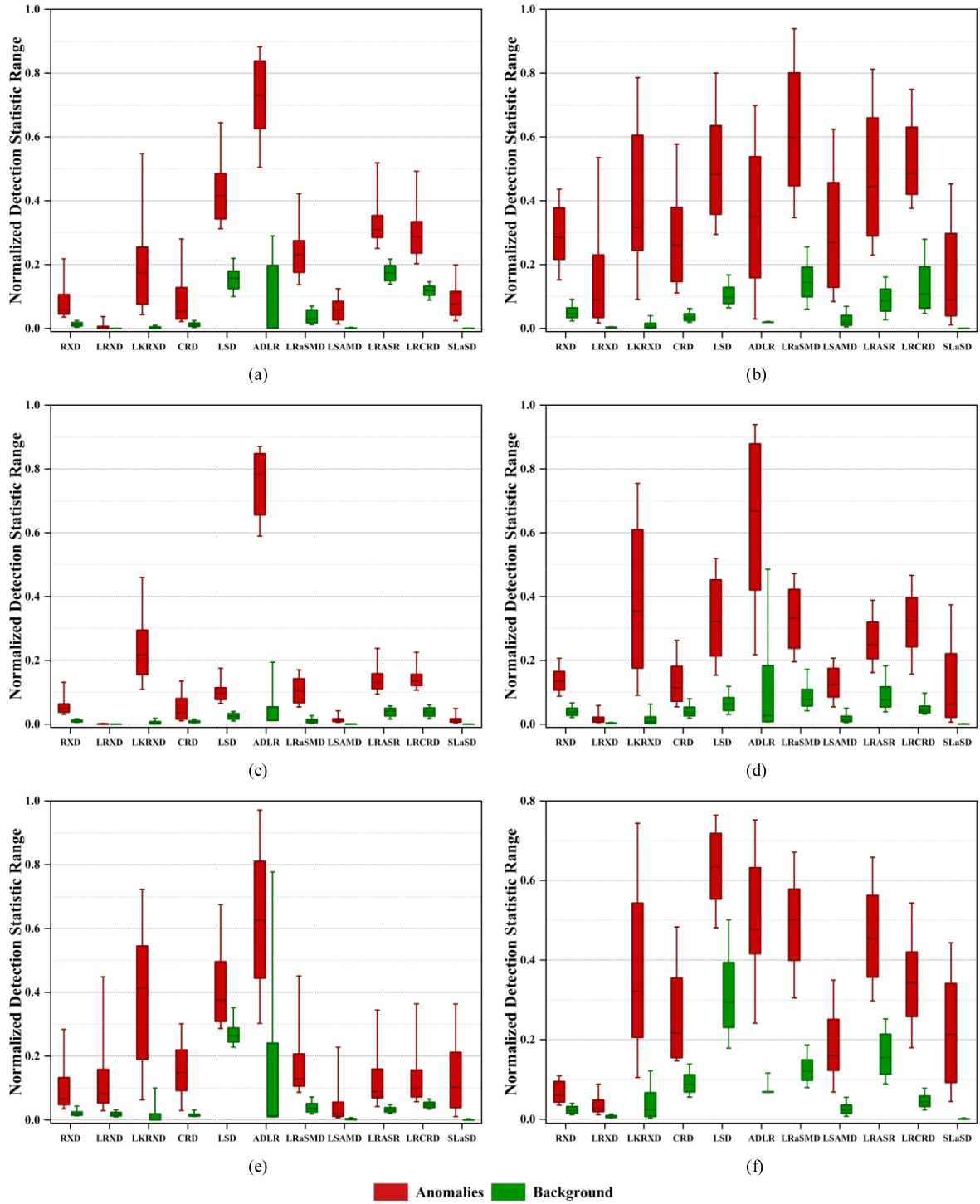


Fig. 6. Box-plots of the statistical information of the anomaly and background in the detection maps for different methods. (a) Texas Coast 1. (b) Texas Coast 2. (c) Los Angeles 1. (d) Los Angeles 2. (e) Pavia. (f) Gulfport.

adaptive weight γ. To analyze the sensitivity to these parameters, in this section, the parameter l was enumerated in the region $(3, 30)$ with a step of 3, and the parameter γ was enumerated in the region $(0, 1)$ with a step of 0.1. For all six datasets, the AUC values of (P_d, P_f) for different combinations of l and γ are shown in Fig. 10. Generally, the performance of the proposed SLaSD method depends on the two parameters to different degrees. Specifically, the first four datasets are much less sensitive to the two parameters, but the accuracy for last two datasets varies obviously with the variation of the parameters. Overall, the proposed SLaSD method generally has satisfactory detection performance when l is low, which can also reduce the computational burden. Therefore, $l \in (3, 12)$ could be suggested as an appropriate choice for the SLaSD method in practical

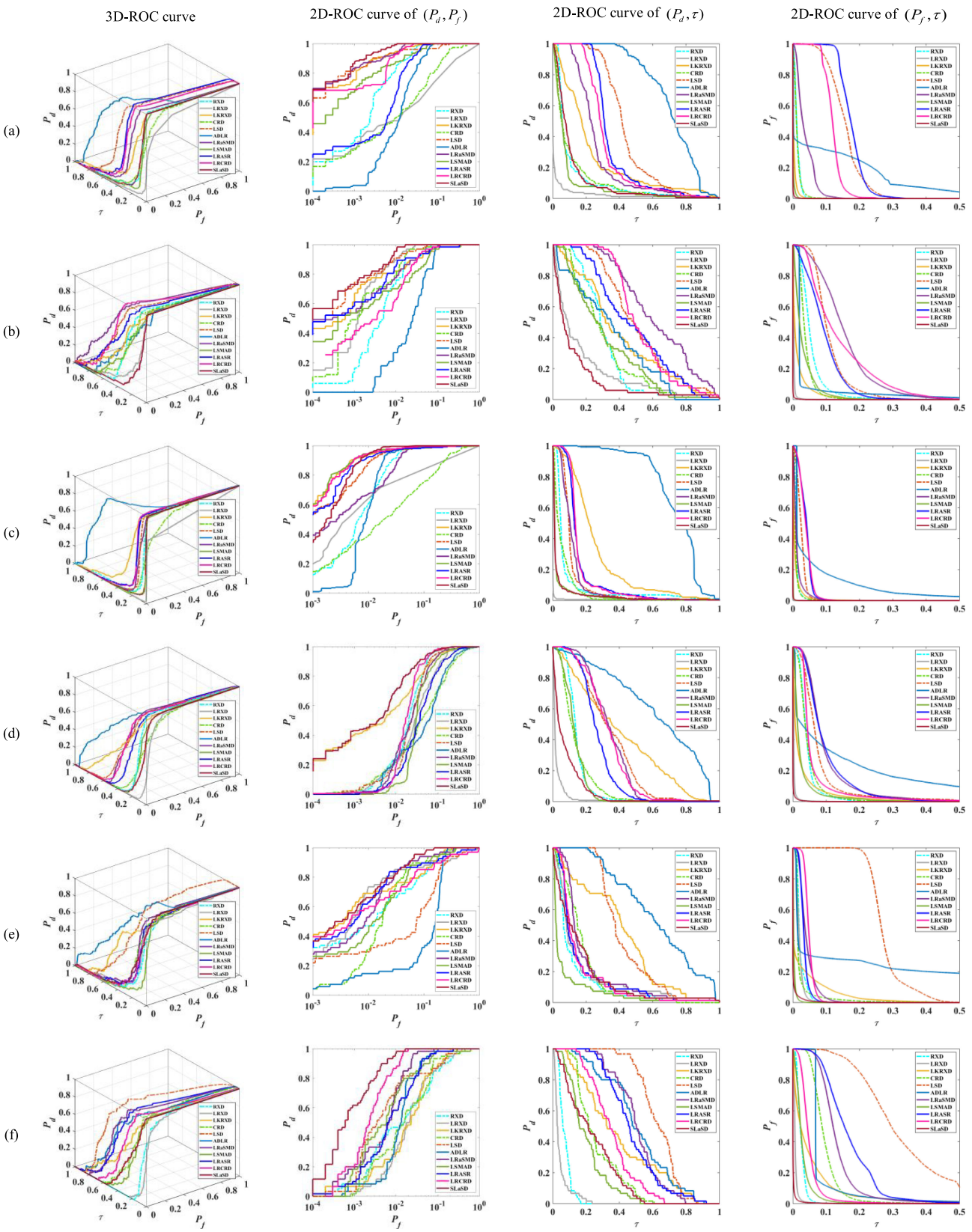


Fig. 7. 3D-ROC curves along with their three derived 2D-ROC curves for different methods. (a) Texas Coast 1. (b) Texas Coast 2. (c) Los Angeles 1. (d) Los Angeles 2. (e) Pavia. (f) Gulfport.

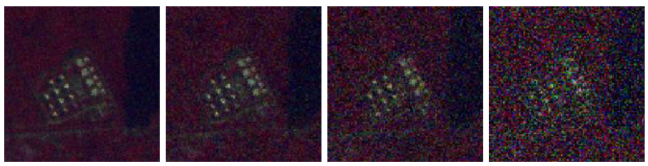
TABLE I
AUC VALUES OF THE METHODS FOR THE SIX DATASETS

Datasets	AUC values of (P_d, P_f)										
	RXD	LRXD	LKRXD	CRD	LSD	ADLR	LRaSMD	LSMAD	LRASR	LRCRD	SLaSD
Texas Coast 1	0.9946	0.9228	0.9979	0.9532	0.9948	0.9829	0.9986	0.9953	0.9908	0.9974	**0.9994**
Texas Coast 2	0.9907	0.9968	0.9911	0.9942	0.9964	0.9678	0.9911	0.9878	0.9855	0.9885	**0.9983**
Los Angeles 1	0.9887	0.9696	0.9941	0.9229	0.9928	0.9882	0.9887	0.9935	0.9930	0.9959	**0.9969**
Los Angeles 2	0.9692	0.9496	0.9704	0.9062	0.9651	0.9103	0.9618	0.9548	0.9395	0.9758	**0.9869**
Pavia	0.9538	0.9391	0.9623	0.9631	0.8905	0.8617	0.9721	0.9602	0.9534	0.9293	**0.9837**
Gulfport	0.9526	0.9538	0.9613	0.9584	0.9705	0.9818	0.9906	0.9778	0.9815	0.9960	**0.9981**
Datasets	AUC values of (P_d, τ)										
	RXD	LRXD	LKRXD	CRD	LSD	ADLR	LRaSMD	LSMAD	LRASR	LRCRD	SLaSD
Texas Coast 1	0.1147	0.0269	0.2353	0.1145	0.4423	**0.7170**	0.2649	0.0862	0.3543	0.3227	0.1118
Texas Coast 2	0.3113	0.1754	0.4012	0.3058	0.5157	0.3478	**0.6254**	0.3117	0.4832	0.5353	0.1311
Los Angeles 1	0.0874	0.0077	0.2617	0.0597	0.1183	**0.7441**	0.1200	0.0333	0.1621	0.1644	0.0340
Los Angeles 2	0.1456	0.0249	0.3964	0.1393	0.3393	**0.6299**	0.3337	0.1310	0.2662	0.3280	0.0764
Pavia	0.1284	0.1616	0.3990	0.1778	0.4220	**0.6276**	0.1962	0.0723	0.1609	0.1580	0.1629
Gulfport	0.0727	0.0438	0.3924	0.2639	**0.6317**	0.5062	0.5036	0.2027	0.4763	0.3554	0.2319
Datasets	AUC values of (P_f, τ)										
	RXD	LRXD	LKRXD	CRD	LSD	ADLR	LRaSMD	LSMAD	LRASR	LRCRD	SLaSD
Texas Coast 1	0.0135	**0.0001**	0.0039	0.0132	0.1569	0.1016	0.0375	0.0015	0.1761	0.1195	0.0004
Texas Coast 2	0.0555	0.0036	0.0154	0.0390	0.1099	0.0388	0.1540	0.0306	0.0928	0.1393	**0.0004**
Los Angeles 1	0.0114	0.0002	0.0076	0.0095	0.0253	0.0673	0.0131	0.0004	0.0390	0.0390	**0.0002**
Los Angeles 2	0.0438	0.0038	0.0257	0.0454	0.0781	0.1431	0.0977	0.0262	0.0975	0.0622	**0.0010**
Pavia	0.0234	0.0202	0.0280	0.0229	0.2763	0.1847	0.0418	0.0035	0.0337	0.0487	**0.0027**
Gulfport	0.0248	0.0085	0.0537	0.0956	0.3163	0.0899	0.1329	0.0297	0.1681	0.0503	**0.0021**

The value in bold means the most accurate result in each case.

TABLE II
AUC VALUES OF THE S-ADM AND ADM MODELS FOR THE SIX DATASETS

Datasets	AUC values of (P_d, P_f)	
	ADM	S-ADM
Texas Coast 1	0.9982	0.9994
Texas Coast 2	0.9939	0.9983
Los Angeles 1	0.9797	0.9969
Los Angeles 2	0.9681	0.9869
Pavia	0.9659	0.9837
Gulfport	0.9720	0.9981



(a) (b) (c) (d)

Fig. 8. False-color images of the Texas Coast 1 dataset with different SNRs. (a) 30 dB. (b) 25 dB. (c) 20 dB. (d) 15 dB.

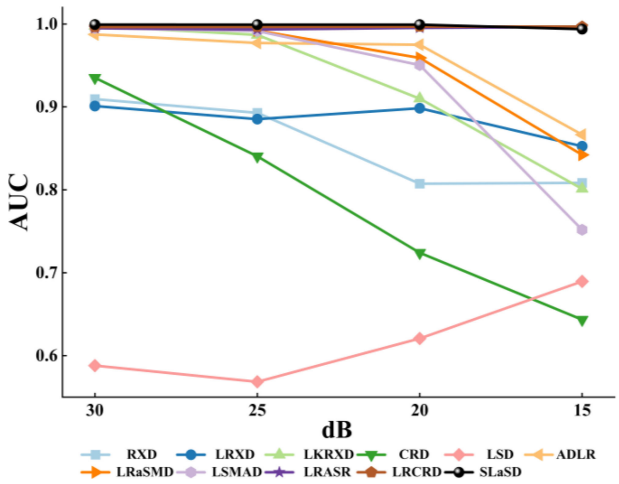


Fig. 9. AUC values of (P_d, P_f) of different methods for the Texas Coast 1 dataset with different SNRs.

applications. In this article, $l = 6$ was used as the empirical value for experiments. In practical applications, however, the value of γ needs to be determined case by case.

IV. DISCUSSION

A. Dimensionality Reduction

To analyze the influence of dimensionality reduction methods on the proposed method, two methods (i.e., the scheme of averaging bands in the divided subspace used in this article and PCA) were compared using the six experimental datasets. The AUC values of (P_d, P_f) are shown in Table III. Generally, the two dimensionality reduction methods show very similar accuracy. Furthermore, no matter which method is used for dimensionality reduction, the proposed SLaSD method consistently

TABLE III
AUC VALUES OF SLASD FOR DIFFERENT DIMENSIONALITY REDUCTION METHODS

Datasets	AUC values of (P_d, P_f)	
	Average	PCA
Texas Coast 1	0.9994	0.9994
Texas Coast 2	0.9983	0.9966
Los Angeles 1	0.9969	0.9970
Los Angeles 2	0.9867	0.9845
Pavia	0.9837	0.9843
Gulfport	0.9981	0.9943

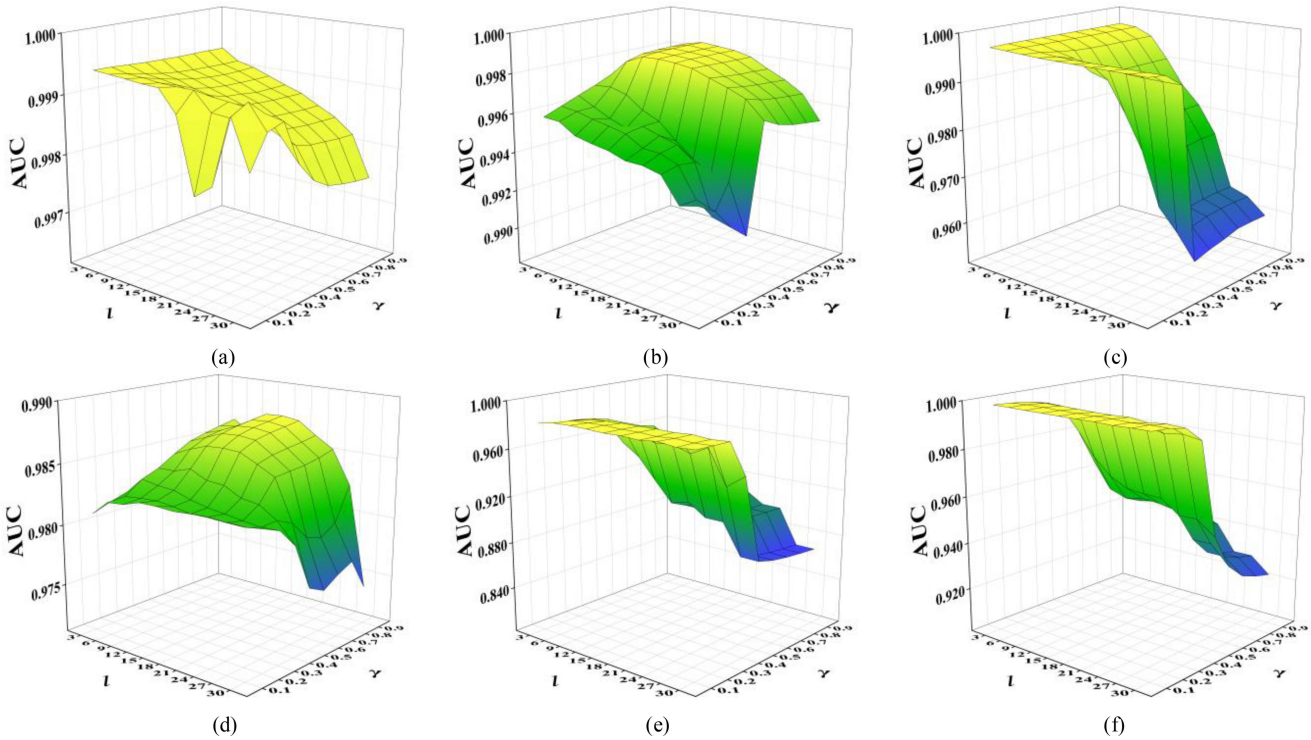


Fig. 10. AUC of SLaSD with the change of the band number l and the weighting coefficient γ. (a) Texas Coast 1. (b) Texas Coast 2. (c) Los Angeles 1. (d) Los Angeles 2. (e) Pavia. (f) Gulfport.

produces greater detection accuracy than the ten benchmark methods.

B. Comparison of Computational Efficiency

The SLaSD method proposed in this article and the LRaSMD, LSMAD, LRASR, and LRCRD methods are detection methods based on the low-rank and sparse model, but the methods for low-rank and sparse decomposition are different. Both LRaSMD and LSMAD use the GoGec algorithm, but their difference lies in the iteration process. Specifically, the LRaSMD method uses SVD, whereas the LSMAD method uses bilateral random projection. For LRASR and LRCRD, they use the linearized alternating direction method with adaptive penalty (LADMAP) for low-rank and sparse decomposition, whereas the proposed SLaSD method uses S-ADM alternatively. Generally, the S-ADM converges slower than LADMAP, but its advantage lies in the robustness and the generalizability. On the other hand, the original HSI data are reduced to lower dimensionality, and S-ADM is used for low-rank and sparse decomposition. Therefore, the computational cost can be reduced greatly for the proposed SLaSD method.

To objectively compare the computational efficiency of the methods based on low-rank and sparse models, the average computing time of the aforementioned five methods for the six datasets is provided. As shown in Table IV, SLaSD has the lowest computational cost. However, considering the slow convergence of the S-ADM process, a fast version for S-ADM could be investigated to further expedite the SLaSD method in future research.

TABLE IV
AVERAGE COMPUTING TIME (IN UNITS OF SECONDS) OF DIFFERENT METHODS FOR THE SIX DATASETS

Method	LRaSMD	LSMAD	LRASR	LRCRD	SLaSD
time	352.14	43.60	304.35	64.88	23.22

C. Model Uncertainty

Each detection method has its own advantages in practical applications. For statistics-based methods, it is difficult for them to satisfy the specific assumption of background distribution in complex HSIs and they are prone to produce false background alarms. Nevertheless, their main advantages lie in simplicity, stability, and sensitivity to anomalies [49]. These characteristics are important in some special application fields, especially where missed detection is much more serious than false detection, such as in small military target detection. On the other hand, the principle of the representation-based methods is that the anomaly cannot be represented by the surrounding background. When the background presents strong homogeneity, the detection performance tends to be more satisfactory than that of the statistics-based methods. Therefore, the representation-based methods have been developed rapidly in recent years. However, due to the limitations of hardware in existing remote sensing sensors, it is difficult to acquire satellite remote sensing images with both fine spatial and spectral resolution, but images with coarse

spatial but fine spectral resolution instead. Generally, coarse spatial resolution images tend to have strong heterogeneity [50], and it is more difficult for them to meet the hypothesis of the representation-based models, leading to inferior performance in anomaly detection.

For the proposed SLaSD and the low-rank and sparse representation-based methods, they can cope with the limitations mentioned earlier to some extent, but still suffer from the uncertainty in two points. First, sparse anomalies are difficult to be decomposed thoroughly from the original HSI. That is, a small number of anomalies (subpixel anomaly) are retained in the low-rank components, leading to leak detection. Second, when the original HSI presents strong heterogeneity, a small amount of false alarms cannot be avoided.

V. Conclusion

The representation-based methods have received increasing attention in hyperspectral anomaly detection. However, this type of methods suffers from the problems of involving a number of controlled parameters, slow rate of convergence, and limited ability in background suppression. To deal with these issues and enhance the performance for hyperspectral anomaly detection, in this article, we proposed the SLaSD method. In the SLaSD method, we proposed the S-ADM scheme for the low-rank and sparse decomposition, which not only increases the iteration speed and the separation between low-rank and sparse components, but also provides a new insight for characterizing the low-rank and sparse model. Meanwhile, to cope with the contamination produced by the sparse noise and background information, a strategy integrating the guided filter and the difference between the sparse features of pixels is proposed for more effective suppression. The experimental results on six HSI datasets showed that the proposed SLaSD method produced greater detection accuracy with more satisfactory background suppression than ten state-of-the-art methods.

References

[1] L. Zhang and L. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geosci. Remote Sens. Mag.*, pp. 2–27, 2022, doi: 10.1109/MGRS.2022.3145854.

[2] L. Deng *et al.*, "M2H-Net: A reconstruction method for hyperspectral remotely sensed imagery," *ISPRS J. Photogramm. Remote Sens.,* vol. 173, pp. 323–348, 2021.

[3] C. Jiao, C. Chen, R. G. McGarvey, S. Bohlman, L. Jiao, and A. Zare, "Multiple instance hybrid estimator for hyperspectral target characterization and sub-pixel target detection," *ISPRS J. Photogramm. Remote Sens.,* vol. 146, pp. 235–250, 2018.

[4] A. Zare, C. Jiao, and T. Glenn, "Discriminative multiple instance hyperspectral target characterization," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 40, no. 10, pp. 2342–2354, Oct. 2018.

[5] Q. Wang, X. Ding, X. Tong, and P. M. Atkinson, "Spatio-temporal spectral unmixing of time-series images," *Remote Sens. Environ.,* vol. 259, 2021, Art. no. 112407.

[6] F. Kowkabi and A. Keshavarz, "Using spectral geodesic and spatial Euclidean weights of neighbourhood pixels for hyperspectral endmember extraction preprocessing," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 201–218, 2019.

[7] L. E. C. La Rosa, C. Sothe, R. Q. Feitosa, C. M. de Almeida, M. B. Schimalski, and D. A. B. Oliveira, "Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data," *ISPRS J. Photogramm. Remote Sens.,* vol. 179, pp. 35–49, 2021.

[8] S. Abbas *et al.*, "Characterizing and classifying urban tree species using bi-monthly terrestrial hyperspectral images in Hong Kong," *ISPRS J. Photogramm. Remote Sens.,* vol. 177, pp. 204–216, 2021.

[9] L. Wang, L. Wang, Q. Wang, and P. M. Atkinson, "SSA-SiamNet: Spectral–spatial-wise attention-based Siamese network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1–18, 2022.

[10] C. Wu, B. Du, and L. Zhang, "Hyperspectral anomalous change detection based on joint sparse representation," *ISPRS J. Photogramm. Remote Sens.,* vol. 146, pp. 137–150, 2018.

[11] C.-I. Chang, "Hyperspectral anomaly detection: A dual theory of hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–20, 2022.

[12] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.,* vol. 19, no. 1, pp. 12–16, Jan. 2002.

[13] Y. Liu, J. Zhao, and Y. Qin, "A novel technique for ship wake detection from optical images," *Remote Sens. Environ.,* vol. 258, 2021, Art. no. 112375.

[14] K. M. Leon-Lopez, F. Mouret, H. Arguello, and J.-Y. Tourneret, "Anomaly detection and classification in multispectral time series based on hidden Markov models," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1–11, 2022.

[15] I. Makki, R. Younes, C. Francis, T. Bianchi, and M. Zucchetti, "A survey of landmine detection using hyperspectral imaging," *ISPRS J. Photogramm. Remote Sens.,* vol. 124, pp. 40–53, 2017.

[16] 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.

[17] R. Zhao, B. Du, and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 7, no. 4, pp. 1227–1234, Apr. 2014.

[18] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.

[19] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 43, no. 2, pp. 374–387, Feb. 2005.

[20] N. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," in *Proc. SPIE Defense Secur. Symp.*, vol. 6966, 2008, Art. no. 696604.

[21] S. Matteoli, M. Diani, and G. Corsini, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Opt. Eng.,* vol. 49, 2010, Art. no. 046201.

[22] 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.

[23] N. Billor, A. S. Hadi, and P. F. Vellema, "BACON blocked adaptive computationally efficient outlier nominators," *Comput. Statist. Data Anal.*, vol. 34, pp. 279–298, 2000.

[24] D. Bo and Z. Liangpei, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 11, pp. 6844–6857, Nov. 2014.

[25] L. Gao, Q. Guo, A. Plaza, J. Li, and B. Zhang, "Probabilistic anomaly detector for remotely sensed hyperspectral data," *J. Appl. Remote Sens.,* vol. 8, no. 1, 2014, Art. no. 083538.

[26] 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.

[27] Z. Yuan, H. Sun, K. Ji, Z. Li, and H. Zou, "Local sparsity divergence for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.,* vol. 11, no. 10, pp. 1697–1701, Oct. 2014.

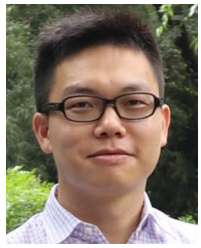
[28] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 8, no. 6, pp. 2523–2533, Jun. 2015.

[29] X. Ye, J. Yang, X. Sun, K. Li, C. Hou, and Y. Wang, "Foreground–background separation from video clips via motion-assisted matrix restoration," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 25, no. 11, pp. 1721–1734, Nov. 2015.

[30] S.-Y. Chen, S. Yang, K. Kalpakis, and C.-I. Chang, "Low-rank decomposition-based anomaly detection," in *Proc. SPIE Defense, Secur., Sens.*, vol. 8743, 2013, Art. no. 87430N.

[31] Y. Qu *et al.*, "Anomaly detection in hyperspectral images through spectral unmixing and low rank decomposition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 1855–1858.

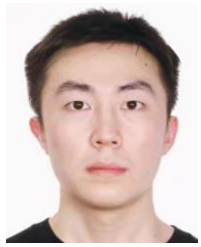
[32] Y. Qu *et al.*, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.
[33] W. Sun, C. Liu, J. Li, Y. M. Lai, and W. Li, "Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery," *J. Appl. Remote Sens.,* vol. 8, no. 1, 2014, Art. no. 083641.
[34] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 54, no. 3, pp. 1376–1389, Mar. 2016.
[35] 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.
[36] H. Su, Z. Wu, A. X. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogramm. Remote Sens.,* vol. 169, pp. 195–211, 2020.
[37] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 58, no. 1, pp. 391–406, Jan. 2020.
[38] R. Feng, H. Li, L. Wang, Y. Zhong, L. Zhang, and T. Zeng, "Local spatial constraint and total variation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1–16, 2022.
[39] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, "GAN-based hyperspectral anomaly detection," in *Proc. 28th Eur. Signal Process. Conf.*, 2021, pp. 1891–1895.
[40] X. Lu, W. Zhang, and J. Huang, "Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 58, no. 3, pp. 1527–1537, Mar. 2020.
[41] W. Xie, J. Lei, S. Fang, Y. Li, X. Jia, and M. Li, "Dual feature extraction network for hyperspectral image analysis," *Pattern Recognit.*, vol. 118, 2021, Art. no. 107992.
[42] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.,* vol. 59, no. 6, pp. 5131–5153, Jun. 2021.
[43] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[44] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.,* vol. 7, no. 2, pp. 118–139, Jun. 2019.
[45] M. Song, F. Li, C. Yu, and C.-I. Chang, "Sequential band fusion for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 60, pp. 1–16, 2022.
[46] Z. Huang, L. Fang, and S. Li, "Subpixel-pixel-superpixel guided fusion for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 58, no. 9, pp. 5998–6007, Sep. 2020.
[47] N. Xue, J. Deng, S. Cheng, Y. Panagakis, and S. Zafeiriou, "Side information for face completion: A robust PCA approach," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 41, no. 10, pp. 2349–2364, Oct. 2019.
[48] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Univ. Illinois, Urbana-Champaign, Champaign, IL, USA, Tech. Rep. UILU-ENG-09-2215, 2010.
[49] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022.
[50] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.



Qunming Wang received the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2015.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. From 2017 to 2018, he was a Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K. His three-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding thesis in the faculty. He has authored or coauthored more than 70 peer-reviewed articles in international journals, such as *Remote Sensing of Environment*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Prof. Wang is an Editorial Board Member of the *Remote Sensing of Environment*, and serves as an Associate Editor for the *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering & Remote Sensing*. From 2017 to 2020, he was an Associate Editor for the *Computers and Geosciences*.



Jiang Zeng received the B.Eng. degree from Xiangtan University, Hunan, China, in 2018. He is currently working toward the M.S. degree with Tongji University, Shanghai, China.

His research interests include hyperspectral image processing and machine learning.

Hao Wu received the Ph.D. degree from Wuhan University, Wuhan, China, in 2007.

He is currently a Professor with the College of Urban and Environmental Sciences, Central China Normal University, Wuhan, China. His research interests include earth observation science, hyperspectral target detection, and urban environmental modeling.

Jiawen Wang, photograph and biography not available at the time of publication.

Kaipeng Sun, photograph and biography not available at the time of publication.