Graph and Total Variation Regularized Low-Rank Representation for Hyperspectral Anomaly Detection

Tongkai Cheng and Bin Wang[ID] , *Senior Member, IEEE*

Abstract—Anomaly detection is of great importance among hyperspectral applications, which aims at locating targets that are spectrally different from their surrounding background. A variety of anomaly detection methods have been proposed in the past. However, most of them fail to take the high spectral correlations of all the pixels into consideration. Low-rank representation (LRR) has drawn a great deal of interest in recent years, as a promising model to exploit the intrinsic low-rank property of hyperspectral images. Nevertheless, the original LRR model only analyzes the spectral signatures without taking advantage of the valuable spatial information in hyperspectral images. Furthermore, it has been shown that the local geometrical information of the hyperspectral data is also important for discrimination between the anomalies and background pixels. In this article, we incorporate the graph regularization and total variation (TV) regularization into the LRR formulation and propose a novel anomaly detection method based on graph and TV regularized LRR (GTVLRR) model, to preserve the local geometrical structure and spatial relationships in hyperspectral images. Extensive experiments have been conducted on both simulated and real hyperspectral data sets. The experimental results demonstrate the superiority of the proposed method over conventional and state-of-the-art anomaly detection methods.

Index Terms—Anomaly detection, graph regularization, hyperspectral images, low-rank representation (LRR), total variation (TV).

I. INTRODUCTION

HYPERSPECTRAL images, acquired by high spectral-resolution sensors, reveal abundant information about the spectral characteristics of the ground surface. Compared with multispectral images, hyperspectral images with hundreds of bands can provide nearly contiguous spectral curves of the materials [1]. This enables the discrimination of different materials based on their distinct spectral signatures [2]. Hyperspectral images, therefore, possess unique advantages in many applications, such as classification and target detection [1]–[4].

Target detection in hyperspectral images has drawn lots of attention for many years. Based on the availability of target information, target detection can be generally divided into two classes: supervised and unsupervised. When the prior spectral information about the target is unavailable, the target detection problem is approached in an unsupervised fashion, also referred to as anomaly detection. The main purpose of anomaly detection in hyperspectral images is to identify uncommon targets that occur as few pixels, such as man-made objects, whose spectra deviate from the natural background significantly. In essence, anomaly detection is a binary hypothesis testing problem, where each pixel under test is labeled as anomaly or background [3]. Due to the difficulty of obtaining accurate spectral signatures of target beforehand, anomaly detection is, therefore, more consistent with practical situations and has been applied in various application fields, such as mineral detection, environmental surveillance, search and rescue, safety control, and food quality [5]–[8]. In addition, it is worth noting that hyperspectral target detection with inaccurately labeled training data has also drawn increasing interest recently. Multiple-instance learning (MIL) is a variation on supervised learning that attempts to target spectra learning in the presence of noise or uncertain target labels and has been recently used for hyperspectral target characterization and detection [9]–[13]. In this article, we focus on anomaly detection.

Over the last two decades, hyperspectral anomaly detection has been researched a lot, and many methods have been proposed [3], [14]. Among these, the Reed–Xiaoli (RX) algorithm [15] is regarded as the benchmark anomaly detection method for hyperspectral images, which is built on the assumption that the background follows a multivariate normal distribution. The decision rule derived from the generalized likelihood ratio test is simply the Mahalanobis distance between the pixel under test and the surrounding background. Two typical versions have been extensively studied: the global RX (GRX), which estimates the background statistics (i.e., mean vector and covariance matrix) using the full image scene, and local RX (LRX), which uses local neighborhood. However, two main limitations exist in these RX methods. On the one hand, a single normal distribution is inadequate to describe the intricate background in real hyperspectral scenes, and on the other hand, the estimated background statistics suffer from the contamination issue due to the existence of noisy pixels and anomalies. In order to cope with

these limitations, a number of improved methods have been presented. Kernel-RX detector [16] employs kernel theory to characterize non-Gaussian distributions in high-dimensional feature space. Support vector data description (SVDD) [17] is another kernel-based detector established on the assumption that the background pixels are enclosed by a hypersphere and those falling outside the hypersphere are regarded as anomalies. However, the computational burden of kernel-based methods is very heavy. In addition, some robust-to-outlier detectors were developed to mitigate the impact of the contamination issue, such as minimum covariance determinant RX (MCD-RX) [18], weighted-RX [19], and blocked adaptive computationally efficient outlier nominator (BACON) [20]. Nevertheless, the intrinsic problem of these detectors is not yet resolved essentially since the naive assumption of RX has not been removed.

In addition to RX detector and its modified versions, representation-based methods without any statistical assumptions have been receiving much attention. These methods are built on the philosophy that the background pixels can be represented by a constructed dictionary, whereas anomalies cannot. Sparse representation has been introduced into various applications of hyperspectral images, such as classification [21], target detection [22], and anomaly detection [23], [24]. The background joint sparse representation detector (BJSRD) [23] adaptively selects the most representative background bases for a local region to achieve robust background estimation. In [24], sparse representation and a linear mixture model were combined to provide physically meaningful solutions. Furthermore, an anomaly removal strategy is proposed to achieve robust detection. Li and Du proposed a collaborative-representation-based detector (CRD) for hyperspectral anomaly detection [25]. This method is developed based on the concept that background pixels can be well represented by its spatial surroundings, while anomalies cannot. Collaboration is achieved through a linear combination of surrounding neighbors of each test sample and is reinforced by ℓ_2-norm minimization of the weight vector.

In general, most of the existing methods perform anomaly detection in a pixel-by-pixel manner and have not taken advantage of the high spectral correlations in the hyperspectral images. More recently, robust principal component analysis (RPCA) [26], as a powerful subspace recovery model in exploring the low-dimensional structure of data, has been applied in hyperspectral anomaly detection [27], in which the RPCA is employed to decompose the hyperspectral data into low-rank and sparse components. To distinguish the noise from the anomaly part, RPCA was extended to model the anomaly and noise separately in [28]. Later, a novel technique called low-rank representation (LRR) [29] was established. This work extends the treatment of data containing outliers from a single subspace (RPCA) to multiple subspaces [29], and thus LRR is more appropriate to recover the intrinsic subspace structure of complex data. The LRR model was first adopted to tackle hyperspectral anomaly detection problem in low rank and sparse representation (LRASR) method [30], where the LRR is used to find the LRR of all pixels jointly with a background dictionary, and a sparsity-inducing regularization

term is added to characterize the local structure of the data. The underlying idea of sparsity criterion is that the region covered by one pixel includes limited materials (endmembers) and the low rankness criterion arises from the fact that the distinct materials (endmembers) contained in a scene are far less than the band number. Thus, these two criteria adopted simultaneously in LRASR method should be the repeated expression to some extent in the physical meaning, and their combination seems to be redundant. In the experimental part of this article, the performance of LRASR is found to be inferior to that of LRR. Additionally, Niu and Wang [31] developed a novel hyperspectral anomaly detector based on LRR and learned dictionary in order to improve the robustness of LRR to the tradeoff parameter.

Unfortunately, these LRR-based methods work simply with the spectral signature and the potentially valuable spatial information which concerns the relationships between each pixel and its spatial neighbors [32] is ignored. Several works on hyperspectral anomaly detection have been developed to utilize spatial information. In [33]–[35], the Markov random field (MRF) modeling was adopted with the purpose of introducing spatial context into the detection process and showed improved anomaly detection performance. In practice, hyperspectral images are usually considered piecewise smooth in the sense that spatial adjacent pixels have similar spectral signatures. Thus, it is highly possible that a pixel and its spatial neighbors have similar representation coefficients in the framework of LRR. Furthermore, the original LRR model characterizes the relationship of all pixels from a global point of view but fails to preserve the local geometrical structure [36]. Recent research has pointed out that the high-dimensional data can be regarded as an embedding on or near to a submanifold of the ambient space [37]. Many manifold learning algorithms, such as Laplacian Eigenmap [37], locally linear embedding (LLE) [38], and ISOMAP [39] have been developed to detect the intrinsic manifold structure. These algorithms are based on the philosophy that nearby points in the original Euclidean space are likely to have similar representations [36], [40]–[42], which is the so-called locally invariant assumption [43]. Many previous studies have shown that both the global and local structures are essential to the representation process [36], [42]. Therefore, it can be expected that the learning performance can be further enhanced if the manifold structures in the data space is preserved.

On the basis of the above analysis, in this article, we present a novel hyperspectral anomaly detection method based on graph and total variation regularized LRR (GTVLRR) model. The total variation (TV) [44], [45], which can promote piecewise smoothness and keep sharp edges effectively is incorporated into the LRR formulation to preserve the spatial relationships in hyperspectral images. Moreover, the proposed method builds a k-nearest neighbor graph to encode the local manifold information in hyperspectral data. The graph Laplacian that derives from the spectral graph theory [46] is utilized to preserve the local manifold structure. Specifically, the graph regularization term is incorporated into the LRR objective function. It is noteworthy that GTVLRR characterizes the properties

of hyperspectral images from completely different aspects, unlike the existing works, such as LRASR with the repeated expression and redundancy. For this reason, GTVLRR could learn a more accurate representation, thereby facilitating the separation of anomalies from natural background. Once the separation process is finished, some basic anomaly detectors can be applied to accomplish the detection process.

The main contributions of this work can be concisely summarized as follows.

1) The proposed method adopts the LRR model to exploit the high spectral correlations in hyperspectral images. Moreover, the TV regularizer serving as a smoothing penalty term is incorporated into the objective function of LRR in order to exploit the spatial correlation between neighboring pixels. Meanwhile, the graph regularization is also imposed on the representations of data points in order that the local geometrical structure on the data manifold can be preserved. The proposed model can, therefore, provide a more accurate characterization of the hyperspectral images, leading to enhanced discriminating power.

2) The alternating direction method of multiplier (ADMM) [47] is utilized to solve the proposed optimization problem. The experimental results on both simulated and real hyperspectral data sets validate the higher detection accuracy of the proposed method, in comparison with conventional and state-of-the-art anomaly detectors.

The remainder of this article is organized as follows. Section II describes the related works. In Section III, the GTVLRR model is first derived, followed by the optimization scheme and the proposed algorithm. Experiments on simulated and real hyperspectral data sets are conducted in Section IV. Finally, the conclusions are presented in Section V.

II. RELATED WORKS

A. RPCA

Traditional principal component analysis (PCA) seeks the best LRR of high-dimensional data in an ℓ_2-sense. It can achieve an optimal result under the condition that the data are corrupted by small Gaussian noise. However, the performance of PCA decreases significantly when the data are grossly corrupted or contain outliers. Comparatively, RPCA [26] was proposed to recover the intrinsic low-rank structure from highly corrupted data. The goal of RPCA is described as follows: $\mathbf{Y}$ is the known observed data and assumed to be a superposition of a low-rank term $\mathbf{L}$ and a sparse term $\mathbf{S}$; the objective is to recover $\mathbf{L}$ and $\mathbf{S}$ from $\mathbf{Y}$. The formulation of the optimization problem [26] is

$$\min_{\mathbf{L},\mathbf{S}} \ \mathrm{rank}(\mathbf{L}) + \lambda \|\mathbf{S}\|_0, \quad \text{s.t. } \mathbf{Y} = \mathbf{L} + \mathbf{S} \tag{1}$$

where $\lambda > 0$ is the regularization parameter. Unfortunately, problem (1) is highly nonconvex due to the discrete nature of the rank function. The following problem is a good convex surrogate [26] for the problem (1):

$$\min_{\mathbf{L},\mathbf{S}} \ \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1, \quad \text{s.t. } \mathbf{Y} = \mathbf{L} + \mathbf{S} \tag{2}$$

where $\|\cdot\|_*$ is the nuclear norm (i.e., the sum of singular values) [48] and $\|\cdot\|_1$ denotes the matrix ℓ_1-norm.

When it comes to hyperspectral anomaly detection, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$ can be regarded as the resized hyperspectral data with N pixels, $\mathbf{y}_i \in \mathbb{R}^B$ denotes the spectrum of the ith pixel with B spectral bands. Consequently, $\mathbf{L}$ is the recovery of the background component, and $\mathbf{S}$ represents the sparse error indicating the anomalies.

B. LRR

The aforementioned RPCA implicitly assumes that the underlying data are drawn from a single subspace [29]. However, mixed pixels exist widely in hyperspectral images, and each material (endmember) corresponds to a single subspace. Thus, the underlying data can be drawn from a union of subspaces. LRR has been established recently to handle the mixed data. The goal of LRR is to find the LRR of $\mathbf{Y}$ with respect to a dictionary $\mathbf{A}$. The optimization problem of LRR [29] can be written as

$$\min_{\mathbf{X},\mathbf{S}} \ \|\mathbf{X}\|_* + \lambda \|\mathbf{S}\|_{2,1}, \quad \text{s.t. } \mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{S} \tag{3}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m]$ is the background dictionary matrix that linearly spans the data space (m is the number of atoms), and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ is the representation coefficient matrix. The $\ell_{2,1}$-norm is adopted to model the sample-specific corruptions [29], and is defined as the sum of ℓ_2-norm of each column in a matrix. One can find RPCA is actually a special case of LRR by setting $\mathbf{A} = \mathbf{I}$. By choosing an appropriate dictionary, LRR can reveal the multiple underlying subspaces and detect anomalies as well.

Once the optimization for problem (3) is accomplished, anomaly detection is then performed on the minimizer $\mathbf{S}^*$. This can be achieved by simply searching the nonzero columns of $\mathbf{S}^*$ under ideal conditions. However, the minimizer $\mathbf{S}^*$ has only approximately sparse column supports. The magnitude of the columns in $\mathbf{S}^*$ can be employed to characterize the anomalous degree [49]. More specifically, the ith pixel is claimed to be an anomaly if the following condition is satisfied:

$$\|[\mathbf{S}^*]_{:,i}\|_2 > \eta \tag{4}$$

where η is a predefined threshold.

III. PROPOSED METHOD

A. Graph and TV Regularized LRR

As seen in [30] and [36], the low rankness captures the high spectral correlations and can reveal the global structure of the observed data $\mathbf{Y}$. However, the basic LRR model focuses solely on the spectral features, while the spatial information concerning the relationships of each test pixel and its spatial neighbors has not been considered. One may naturally hope that if two pixels $\mathbf{y}_i$ and $\mathbf{y}_j$ are adjacent in the spatial dimensions, their corresponding representation coefficients are also similar. The TV regularizer is powerful in preserving the edge information and promoting piecewise smoothness [45]. In order to preserve the spatial relationships in the hyperspectral images, the TV regularizer is included in

the LRR formulation. The cost function of the TV regularized LRR (TVLRR) can be written as follows:

$$\min_{\mathbf{X},\mathbf{S}} \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_{2,1} + \beta \mathrm{TV}(\mathbf{X}), \quad \text{s.t. } \mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{S} \quad (5)$$

where

$$\mathrm{TV}(\mathbf{X}) = \sum_{\{i,j\}\in\varepsilon} \|\mathbf{x}_i - \mathbf{x}_j\|_1 \quad (6)$$

is the expression of TV term [32] and ε contains the pairs of each pixel and its spatial adjacent neighbors (in the horizontal and vertical directions). Minimizing (6) can enhance the piecewise smoothness in spatial dimensions and maintain the sharp edges between the anomalies and background pixels in the coefficient space. The intuitive form of TV term can be further simplified into a matrix form: Let $\mathbf{H}_h$ denote the linear operator calculating the differences of representation coefficients between each pixel and its horizontal neighboring pixel, i.e., $\mathbf{H}_h\mathbf{X} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N]$, where $\mathbf{d}_i = \mathbf{x}_i - \mathbf{x}_{i_h}$ is the difference between the ith pixel and its horizontal neighboring pixel. Here, the periodic boundaries are assumed. Similarly, the linear operator $\mathbf{H}_v$ that counts the vertical differences is defined as $\mathbf{H}_v\mathbf{X} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N]$, in which $\mathbf{v}_i = \mathbf{x}_i - \mathbf{x}_{i_v}$ is the difference between the ith pixel and its vertical neighboring pixel. Thus, the initial form of TV term in (6) can be rewritten in a compact form as follows [32]:

$$\mathrm{TV}(\mathbf{X}) = \left\| \begin{bmatrix} \mathbf{H_h X} \\ \mathbf{H_v X} \end{bmatrix} \right\|_{1,1} = \|\mathbf{HX}\|_{1,1} \quad (7)$$

where $\|\cdot\|_{1,1}$ is defined as the sum of ℓ_1-norm of each column in a matrix. The formulation (5) has the following equivalent form:

$$\min_{\mathbf{X},\mathbf{S}} \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_{2,1} + \beta\|\mathbf{HX}\|_{1,1}, \quad \text{s.t. } \mathbf{Y} = \mathbf{AX} + \mathbf{S}. \quad (8)$$

The TVLRR model is capable of capturing the spatial and spectral information simultaneously. However, both the LRR and the TVLRR models only consider the global Euclidean structure in the data space. It has been shown that the local geometrical structure is also important for discrimination. In order to preserve the local geometrical structure in the data space, one may further hope that the local invariance can be retained during the representation process, i.e., nearby data points in the ambient Euclidean space are likely to have similar representations in the new space [40]. Specifically, if two pixels $\mathbf{y}_i$ and $\mathbf{y}_j$ have similar spectral signatures, their representation coefficients $\mathbf{x}_i$ and $\mathbf{x}_j$ are expected to be close too. This relationship is termed as a manifold assumption. Fig. 1 shows the scatter plot of all the pixels in Hyperspectral Digital Imagery Collection Experiment (HYDICE) data set after applying PCA. The detailed information on this hyperspectral data set will be described in Section IV. From Fig. 1, it is evident that the anomalous pixels and background pixels reside on distinct manifolds. Therefore, if we take the local geometrical information into account and preserve the intrinsic manifold structure, the new representation will be enhanced, and hence, the separability between the anomalies and background pixels should be further enlarged.
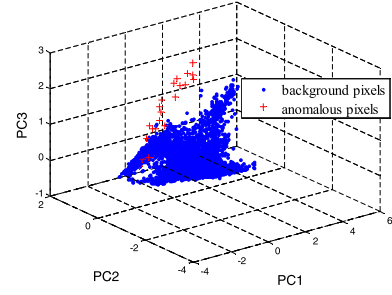

Fig. 1.   Scatter plot of all the pixels in HYDICE data set.

Given a set of pixels $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$, a nearest neighbor graph G is first constructed by regarding each pixel $\mathbf{y}_i$ as its node. Let $\mathbf{W}$ be the weight matrix of the graph G. If $\mathbf{y}_i$ is among the k-nearest neighbors of $\mathbf{y}_j$ or $\mathbf{y}_j$ is among the k-nearest neighbors of $\mathbf{y}_i$, then

$$W_{ij} = \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|_2^2}{\sigma}\right) \quad (9)$$

where the heat kernel [37] is adopted and σ is a scalar parameter; otherwise, the weight is assigned as 0. The aforementioned manifold assumption is equivalent to minimizing the following objective function [37]:

$$\frac{1}{2}\sum_{i,j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 W_{ij} = \mathrm{Tr}(\mathbf{XLX^T}) \quad (10)$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian matrix, $\mathbf{D}$ is a diagonal matrix with entries measuring the degree of each node (pixel), and $D_{ii} = \sum_{j=1}^N W_{ij}$. Incorporating the graph regularization term in (10) into the objective function of TVLRR in (8) gives rise to the optimization problem of GTVLRR as follows:

$$\min_{\mathbf{X},\mathbf{S}} \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_{2,1} + \beta\|\mathbf{HX}\|_{1,1} + \gamma\,\mathrm{Tr}(\mathbf{XLX^T})$$
$$\text{s.t. } \mathbf{Y} = \mathbf{AX} + \mathbf{S} \quad (11)$$

where λ, β, and γ are tradeoff parameters. The model (11) integrates the original LRR model, TV, and graph regularizers into a unified framework, where the high spectral correlations, spatial relationships, and intrinsic geometrical structure in hyperspectral images are taken into consideration simultaneously. As a consequence, the GTVLRR model could lead to a more accurate representation of background, and the residual error should, therefore, provide a finer description for the anomalies.

Additionally, we present a toy example to justify the need for the incorporation of graph and TV regularizers. Similar to the data in [42], a set of data points are constructed in two moons pattern with spatial relationships, as shown in Fig. 2(a) and (b). The result acquired by performing the conventional LRR on the constructed data failed to distinguish the intrinsic manifold structure, as shown in Fig. 3(a). Fig. 3(b) and (c) show the results acquired by performing LRR with graph and TV regularizers, respectively, and both exhibit a significant improvement over that in Fig. 3(a). These results clearly show that the local geometrical information and the
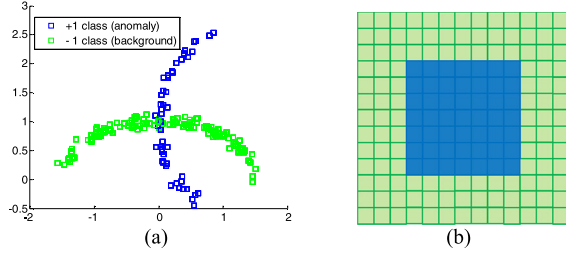
Fig. 2. Illustration of a toy example. (a) Toy data with two intersecting half moons. (b) Spatial relationships between the data points displayed in (a).

spatial relationships encoded in graph and TV regularizers respectively are indeed helpful for discrimination.

B. Optimization Procedure

The problem (11) is first converted to the following equivalent problem by introducing auxiliary variables $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4$:

$$
\min_{\mathbf{V}_1, \mathbf{V}_3, \mathbf{V}_4, \mathbf{S}} \|\mathbf{V}_1\|_* + \lambda \|\mathbf{S}\|_{2,1} + \beta \|\mathbf{V}_3\|_{1,1} + \gamma \operatorname{Tr}\left(\mathbf{V}_4 \mathbf{L} \mathbf{V}_4^T\right),
$$
$$
\text{s.t. } \mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{S}, \quad \mathbf{V}_1 = \mathbf{X}, \quad \mathbf{V}_2 = \mathbf{X}, \quad \mathbf{V}_3 = \mathbf{H}\mathbf{V}_2, \quad \mathbf{V}_4 = \mathbf{X}.
$$
(12)

The objective function in (12) is now separable, and the optimization problem can be efficiently solved by the popular ADMM [47]. The augmented Lagrangian function of problem (12) is formulated as follows:

$$
\mathcal{L}(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4, \mathbf{X}, \mathbf{S}, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4, \mathbf{D}_5, \mu)
$$
$$
= \|\mathbf{V}_1\|_* + \lambda \|\mathbf{S}\|_{2,1} + \beta \|\mathbf{V}_3\|_{1,1} + \gamma \operatorname{Tr}\left(\mathbf{V}_4 \mathbf{L} \mathbf{V}_4^T\right)
$$
$$
+ \frac{\mu}{2} \left(\|\mathbf{Y} - \mathbf{A}\mathbf{X} - \mathbf{S} - \mathbf{D}_1\|_F^2 + \|\mathbf{V}_1 - \mathbf{X} - \mathbf{D}_2\|_F^2 \right.
$$
$$
+ \|\mathbf{V}_2 - \mathbf{X} - \mathbf{D}_3\|_F^2 + \|\mathbf{V}_3 - \mathbf{H}\mathbf{V}_2 - \mathbf{D}_4\|_F^2
$$
$$
\left. + \|\mathbf{V}_4 - \mathbf{X} - \mathbf{D}_5\|_F^2 \right)
$$
(13)

where $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4, \mathbf{D}_5$ are the Lagrange multipliers, and μ is the positive penalty parameter. The minimization of (13) can be achieved by iteratively optimizing the objective function with respect to one variable while fixing the others. The augmented Lagrange function $\mathcal{L}$ is convex, and thus, the global optimality of solution can be guaranteed. The problem can be divided into several subproblems as follows.

1) Update $\mathbf{X}$ while fixing other variables; the subproblem can be formulated as follows:

$$
\mathbf{X}^{(t+1)}
$$
$$
= \arg \min_{\mathbf{X}} \left\|\mathbf{Y} - \mathbf{A}\mathbf{X} - \mathbf{S}^{(t)} - \mathbf{D}_1^{(t)}\right\|_F^2 + \left\|\mathbf{V}_1^{(t)} - \mathbf{X} - \mathbf{D}_2^{(t)}\right\|_F^2
$$
$$
+ \left\|\mathbf{V}_2^{(t)} - \mathbf{X} - \mathbf{D}_3^{(t)}\right\|_F^2 + \left\|\mathbf{V}_4^{(t)} - \mathbf{X} - \mathbf{D}_5^{(t)}\right\|_F^2 \quad (14)
$$

where t is the index of iteration. By differentiating the objective function in (14) with respect to $\mathbf{X}$ and setting the resultant to zero, we can obtain the following result of (14):

$$
\mathbf{X}^{(t+1)}
$$
$$
= (\mathbf{A}^T \mathbf{A} + 3\mathbf{I})^{-1} \left[\mathbf{A}^T \left(\mathbf{Y} - \mathbf{S}^{(t)} - \mathbf{D}_1^{(t)} \right) + \left(\mathbf{V}_1^{(t)} - \mathbf{D}_2^{(t)} \right) \right.
$$
$$
\left. + \left(\mathbf{V}_2^{(t)} - \mathbf{D}_3^{(t)} \right) + \left(\mathbf{V}_4^{(t)} - \mathbf{D}_5^{(t)} \right) \right]. \quad (15)
$$

2) Update $\mathbf{V}_1$ while fixing other variables; the subproblem has the following form:

$$
\mathbf{V}_1^{(t+1)} = \arg \min_{\mathbf{V}_1} \|\mathbf{V}_1\|_* + \frac{\mu^{(t)}}{2} \left\|\mathbf{V}_1 - \mathbf{X}^{(t+1)} - \mathbf{D}_2^{(t)}\right\|_F^2.
$$
(16)

The subproblem (16) can be solved by the singular value thresholding (SVT) operator [50] as follows:

$$
\mathbf{V}_1^{(t+1)} = \Theta_{1/\mu^{(t)}}\left(\mathbf{X}^{(t+1)} + \mathbf{D}_2^{(t)}\right).
$$
(17)

In (17), Θ is the SVT operator.

3) Update $\mathbf{V}_2$ while fixing other variables; the subproblem $\mathbf{V}_2^{(t+1)}$

$$
= \arg \min_{\mathbf{V}_2} \left\|\mathbf{V}_2 - \mathbf{X}^{(t+1)} - \mathbf{D}_3^{(t)}\right\|_F^2 + \left\|\mathbf{V}_3^{(t)} - \mathbf{H}\mathbf{V}_2 - \mathbf{D}_4^{(t)}\right\|_F^2
$$
(18)

having the solution

$$
\mathbf{V}_2^{(t+1)} = (\mathbf{H}^T\mathbf{H} + \mathbf{I})^{-1}\left(\mathbf{X}^{(t+1)} + \mathbf{D}_3^{(t)} + \mathbf{H}^T\left(\mathbf{V}_3^{(t)} - \mathbf{D}_4^{(t)}\right)\right)
$$
(19)

which can be computed efficiently via the discrete Fourier transform diagonalization [32].

4) Update $\mathbf{V}_3$ while fixing other variables; $\mathbf{V}_3$ is computed by solving the following optimization problem:

$$
\mathbf{V}_3^{(t+1)} = \arg \min_{\mathbf{V}_3} \beta \|\mathbf{V}_3\|_{1,1} + \frac{\mu^{(t)}}{2} \left\|\mathbf{V}_3 - \mathbf{H}\mathbf{V}_2^{(t+1)} - \mathbf{D}_4^{(t)}\right\|_F^2
$$
(20)

whose solution has the following analytical form:

$$
\mathbf{V}_3^{(t+1)} = \operatorname{soft}\left(\mathbf{H}\mathbf{V}_2^{(t+1)} + \mathbf{D}_4^{(t)}, \frac{\mu^{(t)}}{\beta}\right)
$$
(21)

where soft$(\cdot, \cdot)$ is the well-known soft thresholding operator [51].

5) Update $\mathbf{V}_4$ while fixing other variables; the subproblem to be solved is

$$
\mathbf{V}_4^{(t+1)} = \arg \min_{\mathbf{V}_4} \gamma \operatorname{Tr}\left(\mathbf{V}_4 \mathbf{L} \mathbf{V}_4^T\right)
$$
$$
+ \frac{\mu^{(t)}}{2} \left\|\mathbf{V}_4 - \mathbf{X}^{(t+1)} - \mathbf{D}_5^{(t)}\right\|_F^2. \quad (22)
$$

Similar to that of the problem (14), the minimization of (22) can be achieved through differentiation. The solution has the following equivalent form:

$$
\mathbf{V}_4^{(t+1)} \cdot \left(2\gamma \mathbf{L} + \mu^{(t)}\mathbf{I}\right) = \mu^{(t)}\left(\mathbf{X}^{(t+1)} + \mathbf{D}_5^{(t)}\right) \quad (23)
$$

where $2\gamma \mathbf{L} + \mu^{(t)}\mathbf{I}$ is a sparse matrix with the size of $N \times N$. Equation (23) is actually a large-scale sparse system of equations that can be solved efficiently through many established algorithms, for instance, preconditioned conjugate gradient (PCG).

6) Update $\mathbf{S}$ while fixing other variables; the optimization problem can be written as follows:

$$
\mathbf{S}^{(t+1)} = \arg \min_{\mathbf{S}} \lambda \|\mathbf{S}\|_{2,1}
$$
$$
+ \frac{\mu^{(t)}}{2} \left\|\mathbf{Y} - \mathbf{A}\mathbf{X}^{(t+1)} - \mathbf{S} - \mathbf{D}_1^{(t)}\right\|_F^2. \quad (24)
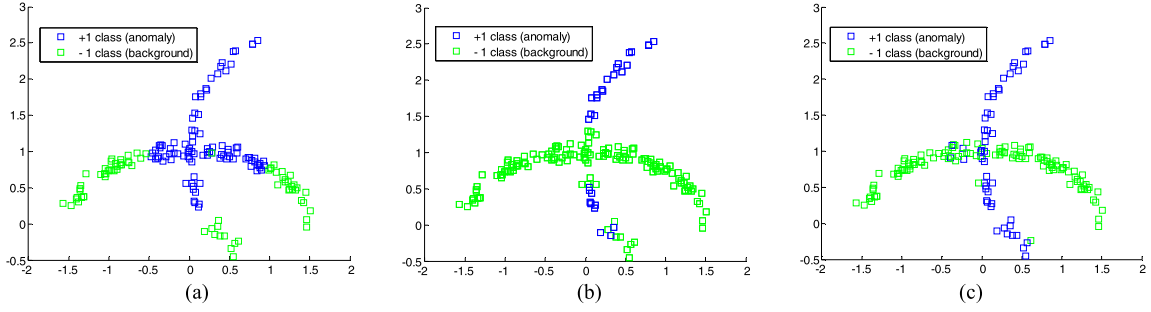$$

Fig. 3.   Classification results on the constructed toy data acquired by performing (a) LRR, whose accuracy is 71%. (b) LRR with graph regularization, whose accuracy is 87.6%. (c) LRR with TV regularization, whose accuracy is 93.5%.

Problem (24) can be solved via the $\ell_{2,1}$-minimization operator [29] and has a closed-form solution. Thus, the computation of this step is time-saving

$$\mathbf{S}^{(t+1)} = \mathbf{\Omega}_{\lambda/\mu^{(t)}}\left(\mathbf{Y} - \mathbf{AX}^{(t+1)} - \mathbf{D}_1^{(t)}\right) \quad (25)$$

where $\mathbf{\Omega}$ is the $\ell_{2,1}$-minimization operator.

Note that the order of updating the above six variables can be changed. Finally, the Lagrange multipliers $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4, \mathbf{D}_5$ and the penalty parameter μ are updated according to (26) and (27), respectively,

$$\begin{aligned}
\mathbf{D}_1^{(t+1)} &= \mathbf{D}_1^{(t)} - \left(\mathbf{Y} - \mathbf{AX}^{(t+1)} - \mathbf{S}^{(t+1)}\right) \\
\mathbf{D}_2^{(t+1)} &= \mathbf{D}_2^{(t)} - \left(\mathbf{V}_1^{(t+1)} - \mathbf{X}^{(t+1)}\right) \\
\mathbf{D}_3^{(t+1)} &= \mathbf{D}_3^{(t)} - \left(\mathbf{V}_2^{(t+1)} - \mathbf{X}^{(t+1)}\right) \\
\mathbf{D}_4^{(t+1)} &= \mathbf{D}_4^{(t)} - \left(\mathbf{V}_3^{(t+1)} - \mathbf{HV}_2^{(t+1)}\right) \\
\mathbf{D}_5^{(t+1)} &= \mathbf{D}_5^{(t)} - \left(\mathbf{V}_4^{(t+1)} - \mathbf{X}^{(t+1)}\right) \\
\end{aligned} \quad (26)$$

$$\mu^{(t+1)} = \min\left(\rho\mu^{(t)}, \mu_{\max}\right). \quad (27)$$

The alternatively iterative process continues until some specific stopping criteria are satisfied. Here, we use both the maximum number of iterations $T_{\max}$ and the residual in (28) to build the stopping conditions. The proposed GTVLRR model is outlined in **Algorithm 1**

$$\begin{aligned}
\|\mathbf{Y} - \mathbf{AX} - \mathbf{S}\|_F + \|\mathbf{V}_1 - \mathbf{X}\|_F + \|\mathbf{V}_2 - \mathbf{X}\|_F \\
+ \|\mathbf{V}_3 - \mathbf{HV}_2\|_F + \|\mathbf{V}_4 - \mathbf{X}\|_F \leq \varepsilon. \quad (28)
\end{aligned}$$

IV. EXPERIMENTAL RESULTS

In this section, we will give a comprehensive analysis of the proposed GTVLRR. Five hyperspectral data sets are used to conduct the experiments; the first one is the simulated hyperspectral data set used for analyzing the performance of the proposed GTVLRR by utilizing the definite ground-truth map of anomalies, and then GTVLRR is applied to the four real hyperspectral images to estimate its practical effectiveness.

We compare the detection performance of the proposed detector with two conventional detectors: GRX and LRX, and three state-of-the-art detectors: CRD [25], RPCA [27], and LRASR [30]. The GRX and LRX are two versions of the benchmark RX detector [14], which are commonly used as the comparison methods in the literature. CRD is

---

Algorithm 1 ADMM Algorithm for GTVLRR

Input: data matrix $\mathbf{Y} \in \mathbb{R}^{B \times N}$, dictionary $\mathbf{A}$, parameters $\lambda > 0$, $\beta > 0$, $\gamma > 0$, the number of nearest neighbors k and the scalar parameter σ in (9).
Initialize: set all the variables $\mathbf{X}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4, \mathbf{S}, \mathbf{D}_1,$ $\mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4, \mathbf{D}_5$ as zero matrices, $\mu_0 = 1e4$, $\mu_{\max} = 1e10$, $\rho = 1.5$, $T_{\max} = 400$, $\varepsilon = 1e - 4$, $t = 0$.
While stopping conditions are not met, **do**
a) Update $\mathbf{X}$ with (15).
b) Update $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4$ with (17), (19), (21), and (23), respectively.
c) Update $\mathbf{S}$ with (25).
d) Update $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4, \mathbf{D}_5$ and μ with (26) and (27), respectively.
e) $t \leftarrow t + 1$.
end while
Output: an optimal solution $(\mathbf{X}^*, \mathbf{S}^*)$.

---

a representation-based detection method, in which the representation is assumed to be the linear combination of surrounding pixels. RPCA is a classical low-rank-based method, and the anomalous degree is also characterized via the l_2-norm of the columns in $\mathbf{S}^*$ after decomposition. Since the proposed GTVLRR can be regarded as the generalization of LRR, we also compare the proposed detector with LRASR. The parameters of the compared detectors are optimally tuned or automatically set as the recommended values.

The receiver operating characteristic (ROC) curve [52] and the area under the ROC curve (AUC) [53] are employed to assess the detection accuracy. The ROC curve describes detection performance with P_D (probability of detection) plotted as a function of P_{FA} (probability of false alarm). When a specific threshold η is applied to the test statistic of a detector, these two probabilities are estimated as follows:

$$P_D = \frac{N_{\text{TT}}}{N_T}, \quad P_{\text{FA}} = \frac{N_{\text{BT}}}{N_B} \quad (29)$$

where N_{TT} counts the number of true targets labeled as a target, N_T denotes the total number of targets (anomalies), N_{BT} counts the number of true background pixel labeled as target, and N_B is the total number of background pixels. The threshold η can be changed across the range of the test statistic to obtain a group of pairs (P_D, P_{FA}), which can be used to
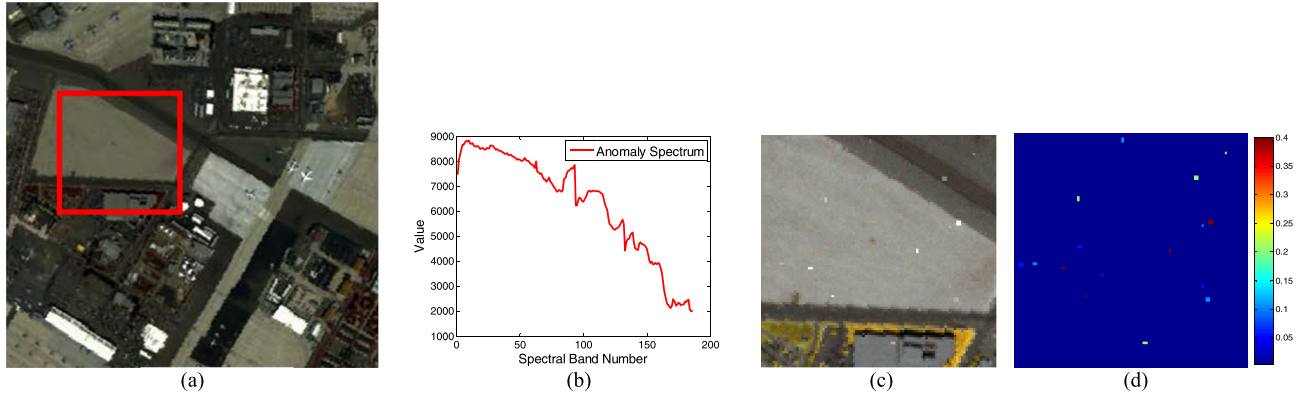
Fig. 4. Hyperspectral data set. (a) Original San Diego image. (b) Anomaly spectrum. (c) False-color image of the simulated data set. (d) Ground-truth map.

generate the ROC curve. Furthermore, the ROC curves with a point-wise confidence interval of P_{FA} [52] are also adopted to make statistical significance comparison. That is, if there is no overlap between the confidence intervals of two different ROC curves, it can be declared that a statistical difference presents between the results. Otherwise, there is no statistical difference. The confidence level is set to 95% in our experiments.

The experimental platform is a computer with an octa-core Intel CPU Xeon CPU 3.07- and 3.06-GHz processors, 48-GB memory, and 64-bit Windows 7. All the experiments are implemented in MATLAB R2013b.

A. Simulated Data Experiments

The first data set is the simulated one, which is generated based on a real scene. It was collected by the airborne visible/infrared imaging spectrometer (AVIRIS) from San Diego airport region, CA, USA. The spatial resolution is 3.5 m, and it contains 224 spectral bands in wavelengths ranging from 370 to 2510 nm. After eliminating the water absorption and low-signal-to-noise ratio (SNR) bands (1–6, 33–35, 94–97, 107–113, 153–166, and 221–224), 186 bands are left in our experiments. The original image has a size of 400×400. In this article, a subimage of size 100×100 is selected to generate the simulated data. The original image and the selected region (in the red square) are displayed in Fig. 4(a). The anomalous pixels are generated by the target implantation method [54]. A synthetic anomalous pixel with spectral vector $\mathbf{z}$ and a specified mixing fraction f $(0 \leq f \leq 1)$ is generated by fractionally implanting a specified anomaly with spectral signature $\mathbf{t}$ into a given background pixel with spectral signature $\mathbf{b}$ as follows:

$$\mathbf{z} = f \cdot \mathbf{t} + (1 - f) \cdot \mathbf{b}. \tag{30}$$

In the simulated data set, 16 anomalous panels are embedded with different sizes of 1×1, 1×2, 2×1, 2×2, and every four panels have the same size with different mixing fractions f $(= 0.05, 0.1, 0.2, 0.4)$. The mixing fraction f remains unchanged in the same panel. The spectrum of the plane in the middle left of the whole image scene is chosen as the anomalous spectral signature $\mathbf{t}$, which is shown in Fig. 4(b). These panels are randomly distributed in the scene.

Moreover, we add the zero-mean white Gaussian noise to the simulated images (30 dB). Here, the SNR is defined as

$$\text{SNR} = 10 \log_{10} \frac{\mathbb{E}[\mathbf{y}^T \mathbf{y}]}{\mathbb{E}[\mathbf{e}^T \mathbf{e}]} \tag{31}$$

where $\mathbb{E}[\cdot]$ is the expectation operator, $\mathbf{y}$ and $\mathbf{e}$ denote the observation and the additive noise of a pixel, respectively.

It should be pointed out that the simulated data set produced in this article is more consistent with the practical scenarios in three aspects: 1) anomalous panels with different sizes are generated and embedded; 2) the spatial locations of these panels are randomly distributed; and 3) the mixing fraction f is relatively low, and it results in a weak anomaly detection problem, which is much more difficult and meaningful compared to that of pure anomaly $(f = 1)$. The false-color image and the ground-truth map of the simulated data set are shown in Fig. 4(c) and (d), respectively.

1) Dictionary Selection: In the proposed GTVLRR, the dictionary $\mathbf{A}$ plays an important role in the process of representing background pixels. In supervised target detection problems, the sparsity-based algorithms require a structured dictionary consists of both target and background training samples [22], which are supposed to be known in advance. However, as for the hyperspectral anomaly detection, the dictionary is usually unknown beforehand. Additionally, in order to achieve good separation of anomalies from the natural background, the dictionary should cover all the background materials and represent the background as well as possible. One may choose the original data matrix itself $\mathbf{Y}$ as the dictionary [29]. Although the original data comprise anomalies, the effect is negligible due to the low occurrence probability of anomalies. Nevertheless, this simple choice suffers an unbearable computational complexity (the size of $\mathbf{X}$ is $N \times N$). To reduce the number of atoms and the time consumption as well, a straightforward way is to randomly choose some of the pixels to constitute the dictionary. However, this method is unable to ensure that the constructed dictionary covers all the background materials, and thus, unstable detection result may be produced.

In this article, we adopt the dictionary construction strategy proposed in [30]. In this strategy, K-means is first used to divide all pixels in the hyperspectral images into M clusters $\mathbf{Y} = \{\mathbf{Y}^1, \mathbf{Y}^2, \ldots, \mathbf{Y}^M\}$. Then the RX detector is applied
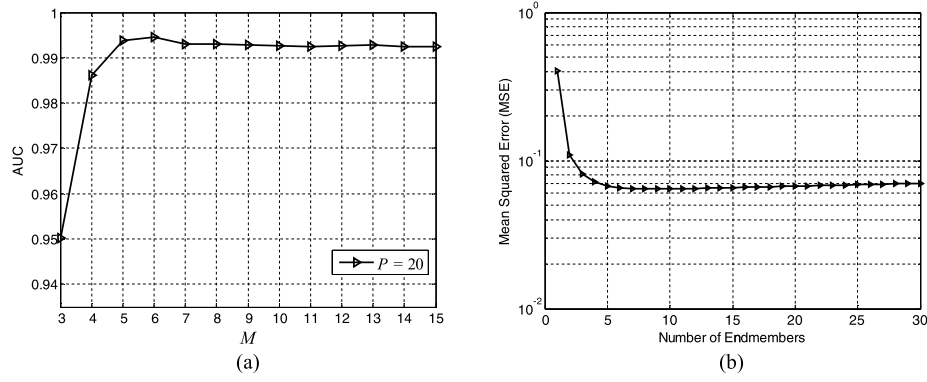
Fig. 5. Analysis of endmember number on the simulated data set. (a) Detection accuracy of GTVLRR with different values of the number of clusters M. (b) MSE varies with respect to the number of endmembers.
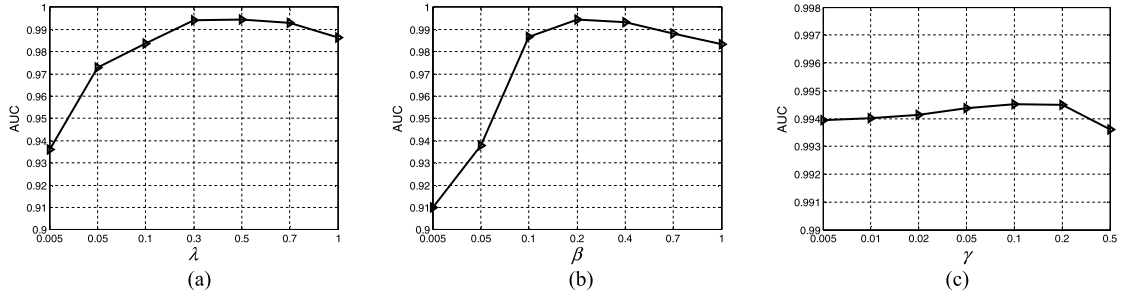

Fig. 6. Detection accuracy of GTVLRR on the simulated data set, with different parameter settings. (a) λ varies. (b) β varies. (c) γ varies.

to each cluster, the Mahalanobis distance between each test pixel and the corresponding local background mean vector is calculated. The P pixels which give the smallest outcomes in each cluster can be regarded as background pixels with high probability, which are chosen to constitute the dictionary $\mathbf{A}$.

In this experiment, we investigate the effect of the number of clusters M on the detection performance, P is set to 20 which is recommended in [30] which can ensure sufficient diversity of spectral signature in each cluster. Fig. 5(a) gives the variation of detection accuracy with respect to the number of clusters M when the tradeoff parameters in the GTVLRR model are set to $\lambda = 0.5$, $\beta = 0.2$, $\gamma = 0.05$, and $k = 10$, $\sigma = 1$ in the weight matrix. This default parameter setting will be analyzed in detail in Section IV-A2. The detection accuracy rises rapidly when M changes from 3 to 5 and achieves relatively high values when M exceeds 5. The main rationale in choosing an appropriate M is that it should be higher than the actual number of background classes so as to cover all the background materials. Moreover, M need not be too large to avoid a cluster made up of only anomalies being produced, which may weaken the discriminative capability of the constructed dictionary. The Hysime algorithm [55] can be utilized to estimate the actual number of background classes, and the result is given in Fig. 5(b). Consequently, the parameter M is set as $M = 6$ for the simulated data set in our experiment.

2) Selection of Parameters: We first perform extensive searching to find the best parameter pair (λ, β) for the TVLRR model. In addition, there are another three parameters in GTVLRR model, i.e., the tradeoff parameter γ, the number
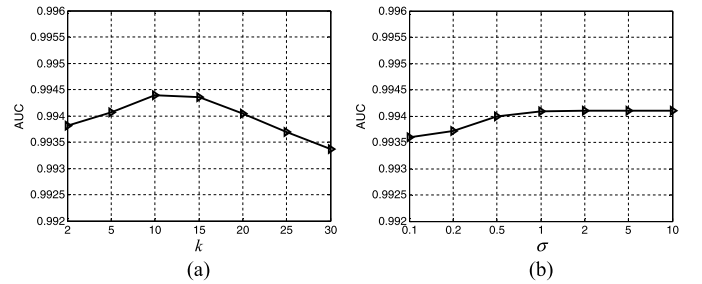

Fig. 7. Detection accuracy of GTVLRR on the simulated data set with respect to (a) number of nearest neighbors k and (b) scalar parameter σ.

of nearest neighbors k and the scalar parameter σ. While fixing the best parameter pair (λ, β), extensive searching is performed once again to find the best setting for (α, k, σ). In this way, an empirical parameter setting for GTVLRR can be $\lambda = 0.5$, $\beta = 0.2$, $\gamma = 0.05$ and $k = 10$, $\sigma = 1$.

In order to investigate the influence of the parameters in GTVLRR on the detection accuracy, we vary a parameter while keeping the others fixed as the reference setting. Fig. 6 shows how the detection accuracy varies with respect to each tradeoff parameter. The results on λ and β indicate that both low rankness and spatial information are of great importance for data representation. We can also note that the performance of the proposed detector is stable over a large range of γ. Fig. 7 shows the variation of detection accuracy with regard to k and σ in regard to the weight matrix. From Fig. 7(a), a relatively good choice can be $k = 10$. Fig. 7(b)

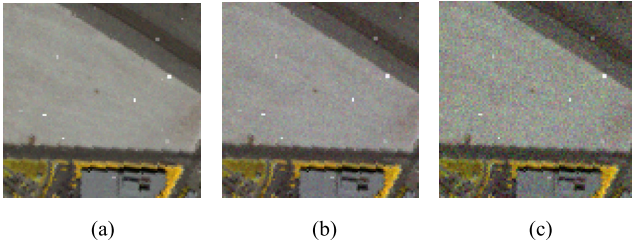(a)   (b)   (c)

Fig. 8.   False-color images of the simulated data sets with different levels of zero-mean Gaussian noise. (a) 30 dB. (b) 25 dB. (c) 20 dB.

TABLE I
AUC SCORES OF LRR, TVLRR, AND GTVLRR FOR THE
SIMULATED DATA SETS WITH DIFFERENT LEVELS
OF ZERO-MEAN GAUSSIAN NOISE

| Noise level	LRR	TVLRR	GTVLRR
30 dB	0.9520	0.9915	**0.9944**
25 dB	0.9492	0.9889	**0.9938**
20 dB	0.8771	0.9728	**0.9806**

Fig. 9.   2-D plots of detection results obtained by LRR, TVLRR, and GTVLRR for the simulated data sets with different levels of white Gaussian noise. (a) 30 dB. (b) 25 dB. (c) 20 dB.

shows the change of detection accuracy with respect to σ, and the AUC values seem not to change anymore when σ goes to 1.

Moreover, we conduct a global perturbation to investigate the robustness of the proposed detector to parameters. Here, the values for k and σ are fixed since the performance of GTVLRR is insensitive to the choice of these two parameters, which can be observed from Fig. 7. When GTVLRR is performed with all possible combinations of the parameter values listed in Fig. 6, the AUC values of the worst and best cases are 0.8929 and 0.9944, respectively. However, when parameters λ and β take value in [0.1, 0.7] from Fig. 6(a) and (b), respectively, the AUC value of the worst-case becomes 0.9820. This indicates that GTVLRR is quite stable in a broad range of parameters. Generally, parameter tuning is nontrivial, and often a computationally expensive task. Therefore, we simply employ the above-mentioned reference setting as a default choice throughout the experiments, with the purpose of saving computations.

3) Effects of Regularization Terms: In this section, we investigate the contributions of the included regularization terms, i.e., TV regularization and graph regularization on the detection performance using the simulated data set. The LRR-based detectors investigated in this experiment are listed as follows:

1) LRR without any regularization terms;
2) LRR with TV regularization, i.e., TVLRR;
3) LRR with graph and TV regularizations, i.e., GTVLRR.

The tradeoff parameters involved in the first two models have been tuned to achieve their optimal performance, i.e., $\lambda = 0.004$ for LRR, and $\lambda = 0.5$ and $\beta = 0.2$ for TVLRR. In addition, to evaluate the robustness of LRR-based detectors to noise, different levels of zero-mean Gaussian noise (30, 25, and 20 dB, respectively) are added into the simulated data set, and the false-color images are displayed in Fig. 8. The 2-D plots of detection results are shown in Fig. 9. Meanwhile, Fig. 10 and Table I show the detection results in terms of
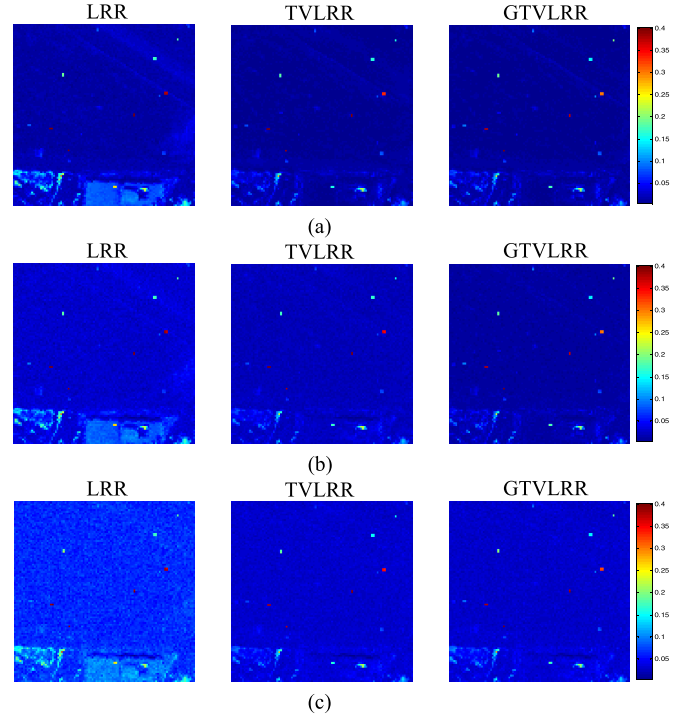
ROC curves and AUC scores, respectively. Note that in this analysis scenario, the ROC curves without confidence interval are employed in order to observe the difference better. From these results, it can be seen that both TVLRR and GTVLRR are good at suppressing the background and noise, and perform clearly better than the original LRR detector. It can be concluded that the inclusion of the TV regularizer offers the potential to improve detection performance significantly. Moreover, as the noise increases, the spatial information becomes particularly important, which can be observed from the results under SNR = 20 dB. This is due to the fact that the TV regularizer is powerful in suppressing the Gaussian noise.

Compared to TVLRR, the ROC curves of the proposed GTVLRR are closer to the upper left corner. It is noteworthy that TVLRR with the best parameter setting is actually the special case of GTVLRR with the default setting when γ is set to zero. Thus, the improvements indicate the positive effect of the local manifold information encoded in the graph regularizer, which is also important for discrimination between the anomalies and background pixels. The improvement becomes more obvious when the scene is contaminated by strong noise, and hence, the graph regularizer becomes more significant when the noise level is high, as reported in Table I.

4) Comparison With Different Anomaly Detectors: This experiment aims at evaluating the detection performance of the proposed GTVLRR in comparison with GRX, LRX, CRD, RPCA, and LRASR using the simulated data sets with different levels of zero-mean Gaussian noise (30, 25, and 20 dB, respectively). The dual windows (win$_{out}$, win$_{in}$) in LRX and CRD are set to (7, 5) and (11, 9), respectively,
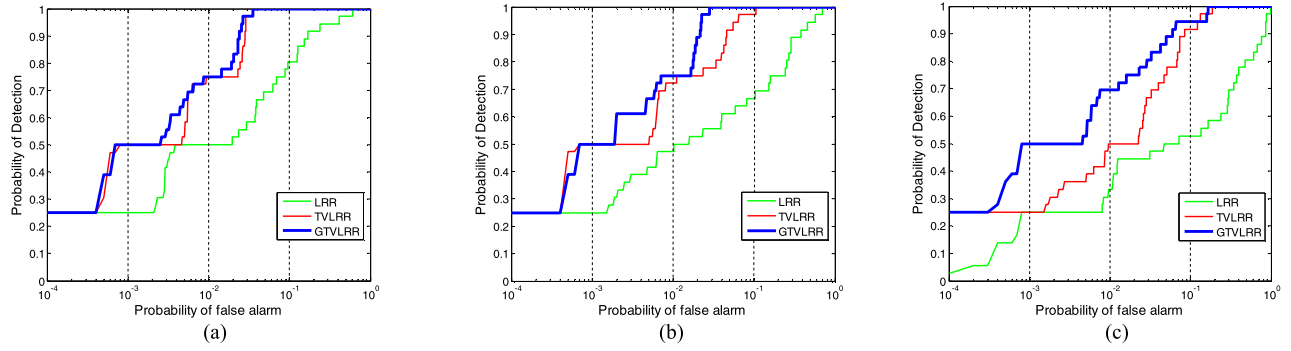
Fig. 10.   ROC curves of LRR, TVLRR, and GTVLRR for the simulated data sets with different levels of zero-mean Gaussian noise. (a) 30 dB. (b) 25 dB. (c) 20 dB.
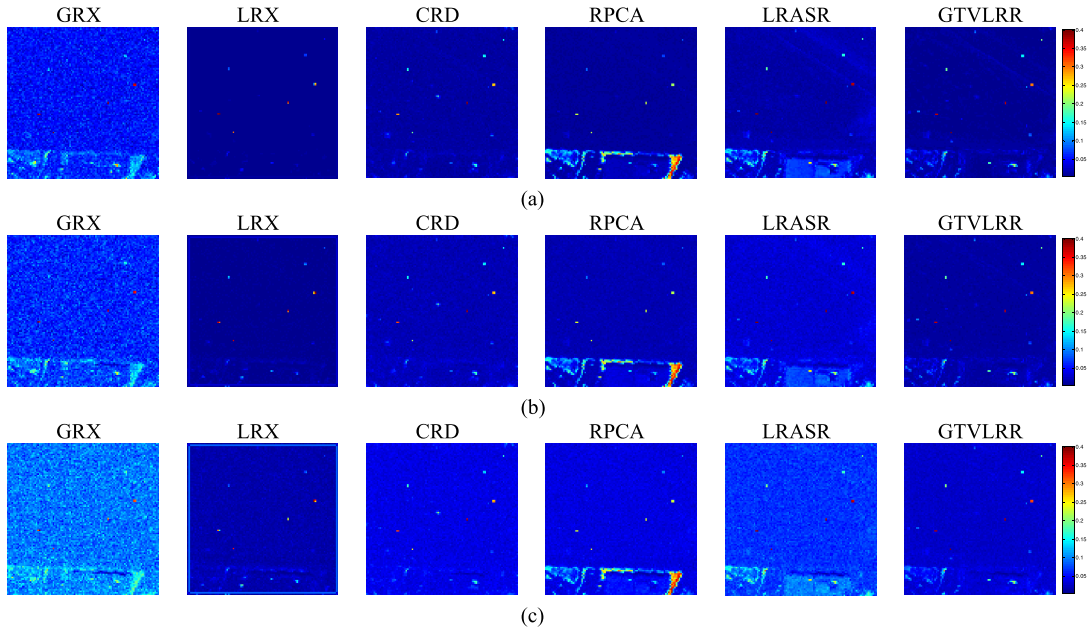

Fig. 11.   2-D plots of detection results obtained by different detectors for the simulated data sets with different levels of zero-mean Gaussian noise. (a) 30 dB. (b) 25 dB. (c) 20 dB.

after extensive searching. The regularization parameter in RPCA is set to $\lambda = 0.005$. The background dictionary in LRASR is identical with that constructed in GTVLRR for fair comparison, the tradeoff parameters β and λ are both set to 0.1, as recommended in [30].

The 2-D plots of the detection results are shown in Fig. 11. From these figures, it can be observed that all the detectors can distinguish between background pixels and the anomalous pixels with relatively high mixing fraction (i.e., $f = 0.4$). The proposed GTVLRR can detect all the implanted anomalies, and provides an obvious map for anomalies in all cases. Compared to LRASR and RPCA, the proposed detector can suppress the false alarms caused by pixels on the bottom of the scene effectively. The ROC curves with 95% confidence interval on the probability of false alarm and the AUC scores are provided in Fig. 12 and Table II, respectively. It can be observed that CRD can achieve a comparable probability of detection to that of the proposed GTVLRR when the probability of false alarm is less than 10^{-2}, and their confidence intervals overlap with each other. This is because the CRD

TABLE II

AUC SCORES OBTAINED BY DIFFERENT DETECTORS FOR THE SIMULATED DATA SETS WITH DIFFERENT LEVELS OF ZERO-MEAN GAUSSIAN NOISE

Noise level	GRX	LRX	CRD	RPCA	LRASR	GTVLRR
30 dB	0.8266	0.9818	0.9756	0.9180	0.9432	**0.9944**
25 dB	0.8593	0.9315	0.9736	0.8960	0.9399	**0.9938**
20 dB	0.7699	0.9178	0.9144	0.8182	0.8664	**0.9806**

is powerful in detecting small targets (anomalies) with the adoption of dual windows. However, GTVLRR is still the best in terms of the overall detection performance.

Another important observation is that the AUC scores of LRASR are even worse than those of LRR in Table I, which may result from the function of sparsity-inducing regularizer is a repeated expression and redundant with low rankness to some extent. It is apparent that the detection performance of GTVLRR is the most stable among all the detectors as
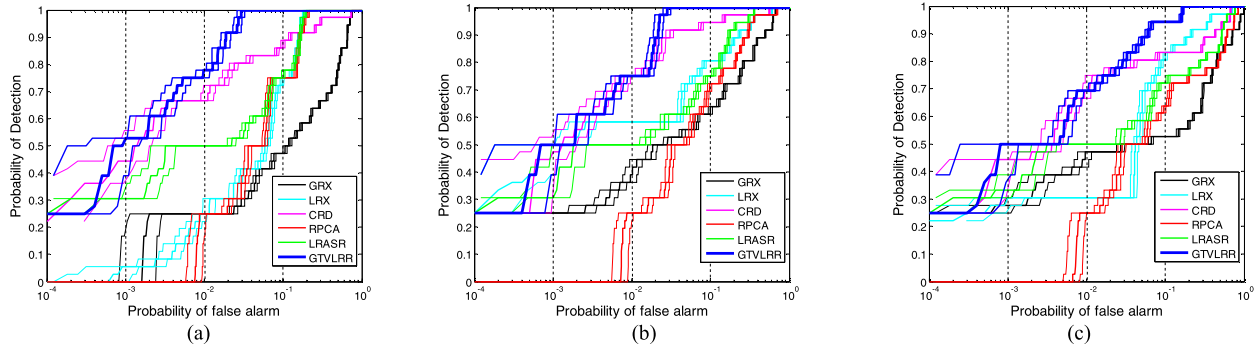
Fig. 12. ROC curves obtained by different detectors with 95% confidence intervals on the probability of false alarm for the simulated data sets with different levels of zero-mean Gaussian noise. (a) 30 dB. (b) 25 dB. (c) 20 dB.

TABLE III
AUC Scores and Computation Times of Different Detectors Based on 20 Runs (Average ± Std)

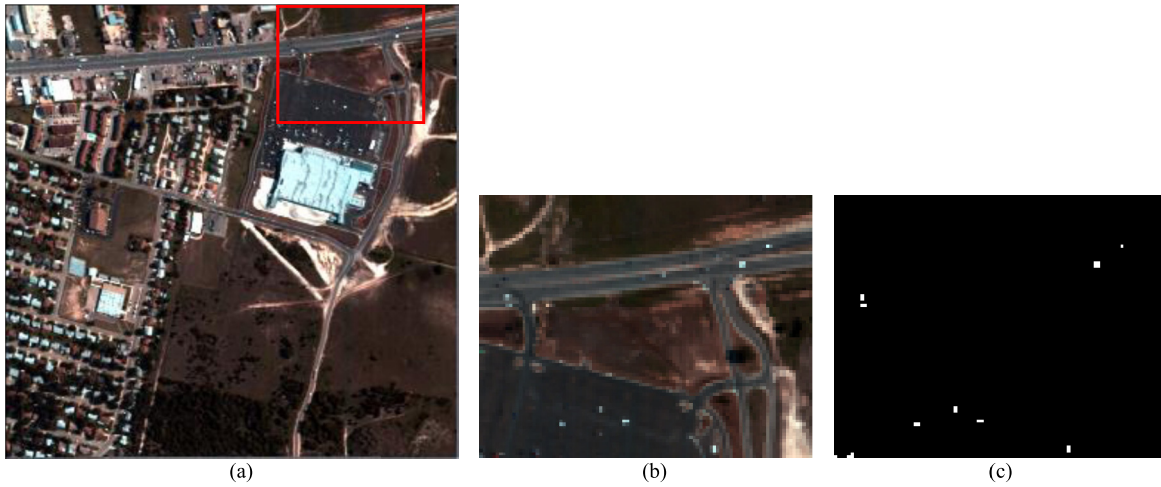| Algorithm	GRX	LRX	CRD	RPCA	LRASR	GTVLRR
AUC	0.8100±0.0280	0.9696±0.0074	0.9674±0.0107	0.9001±0.0137	0.9434±0.0164	**0.9935±0.0008**
Time(s)	0.22±0.03	83.21±9.19	100.82±11.51	6.03±0.43	32.47±3.45	158.28±19.51

Fig. 13. HYDICE data set. (a) Whole image scene. (b) False-color image of the selected region. (c) Ground-truth map.

the noise level increases. Particularly when the noise level is 20 dB, the AUC score of GTVLRR is 0.06 higher than the second best, i.e., LRX, which is a significant gap. The results clearly demonstrate that the proposed GTVLRR outperforms the compared anomaly detectors.

Furthermore, we repeat the above experiment for 20 times to reduce the impact brought by the random positions of implanted anomalies, and the noise level is fixed as 30 dB. Table III gives the average AUC values and the corresponding standard deviations (std). It can be observed that the proposed GTVLRR outperforms the compared detectors and can provide a much more stable detection result. In addition, we also provide the computational costs in Table III, and the proposed detector takes more computation time than other detectors.

The results obtained with the simulated data set are quite encouraging and reveal the positive effects of incorporating

spatial relationship and local manifold structure in conventional LRR formulation. In Section IV-B, we will further conduct experiments with real hyperspectral data sets to validate the effectiveness of the proposed detector in real scenarios.

B. Real Data Experiments

In this section, four commonly used real-world data sets are used to illustrate the practical performance of the proposed method. These four data sets are different in data collection instrument, anomaly size, background type, and image size.

The first real data were collected by the HYDICE in 1995.[1] It covers an urban region and has a spatial resolution of 1 m. The original image consists of 307×307 pixels and 210 spectral bands. This image scene is shown in Fig. 13(a); it comprises a vegetation area, a construction area, and several

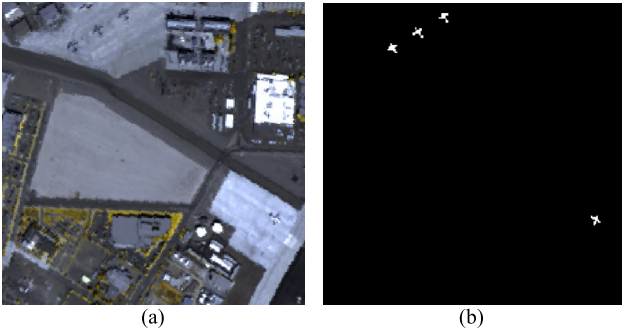[1]Available at http://www.escience.cn/people/feiyunZHU/index.html.

Fig. 14. San Diego data set. (a) False-color image of the detection region. (b) Ground-truth map.

Fig. 15. Hyperion data set. (a) False-color image. (b) Ground-truth map.

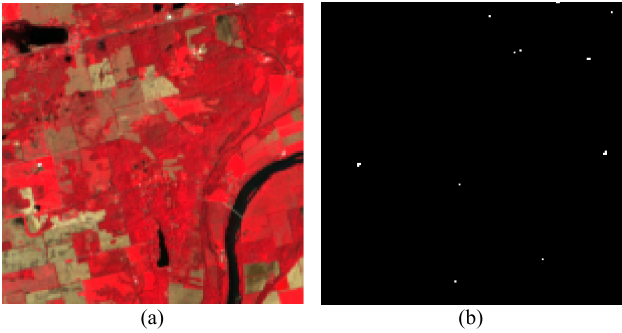Fig. 16. MUUFL Gulfport campus 1 data set. (a) False-color image. (b) Ground-truth map.

roads along with some vehicles. In our experiment, 160 bands remain after removing the low-SNR and water vapor absorption bands (1–4, 76, 87, 101–112, 136–153, 197–210). However, the anomalous targets are the cars and roofs embedded in the top right of the whole scene. Therefore, an 80 × 100 subimage is cropped for our experiment. The 21 pixels are selected as the anomalies, composed of man-made vehicles and roof with different sizes. The false-color image and the ground-truth map are depicted in Fig. 13(b) and (c), respectively.

The second real data set is also part of the aforementioned San Diego scene. The top-left subscene with a size of 220 × 220 is selected to test the performance of the proposed detector, as shown in Fig. 14(a). The background of this data set is mainly composed of roofs, shadows, and grass. The four airplanes are regarded as the anomalies to be detected. The ground-truth map is shown in Fig. 14(b). The sizes of anomalies in this scene are relatively large compared to those of the simulated data set.

The third real data were acquired by the Hyperion imaging sensor [14], covering an agricultural area of the State of Indiana, USA, in 2008. This data set has 242 bands with a spectral resolution of 10 nm, and 149 bands are left after removing the low-SNR bands and the uncalibrated bands. The 150 × 150 subimage with the ground-truth map for the anomalies is used, where the anomalies come from the storage silo and the "roof," these anomalies are not distinguishable visually from the background. The false-color image and the ground-truth map are shown in Fig. 15(a) and (b), respectively.
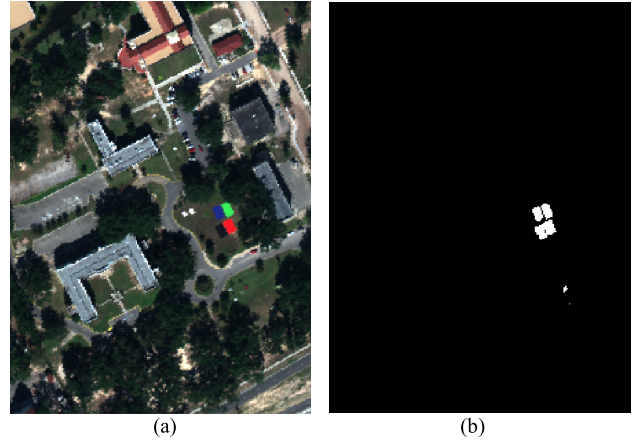
The fourth real data set was collected in November 2010 over the University of Southern Mississippi Gulf Park Campus [56]. The original MUUFL Gulfport data set campus 1 scene contains 325 × 337 pixels across 72 bands. After removing the invalid area and noisy bands, the cropped hyperspectral imagery with a size of 300 × 220 × 64 is used for anomaly detection. Cloth panels with various spectral signatures were placed in the scene, and those larger than pixel-level and visually identifiable are regarded as the anomalies with 269 pixels to be detected. Fig. 16(a) and (b) respectively show the false-color image of the scene and the ground-truth map.

For the HYDICE data set, the number of clusters M is set as $M = 15$. The dual windows (win$_{out}$, win$_{in}$) in LRX and CRD are both set to (5, 3), and the regularization parameter in RPCA is chosen as $\lambda = 0.007$ for achieving optimal performance. The 2-D plots of detection results are shown in Fig. 17(a). Among all the detectors, the proposed GTVLRR can not only provide an obvious map for the anomalies but also suppress the background effectively. For quantitatively analyzing the detection performance of all the detectors, the ROC curves with 95% confidence interval on the probability of false alarm and the AUC values are shown in Fig. 18(a) and Table IV, respectively. The proposed detector presents a lower probability of detection compared to LRASR and RPCA when the probability of false alarm is less than 6×10^{-4}, and the performance achieved by CRD is comparable to GTVLRR when the probability of false alarm is within the range of $(10^{-2}, 10^{-1})$. Nevertheless, the proposed GTVLRR is the best detector in terms of the overall performance. The AUC value of the proposed detector is 0.9976 and it is the best among all the detectors.

For the real San Diego data set, M is set as $M = 20$. For LRX and CRD, the dual windows (win$_{out}$, win$_{in}$) are both set to (13, 11). The regularization parameter in RPCA is set to $\lambda = 0.01$. The 2-D plots of detection results are shown in Fig. 17(b), and it can be seen that the proposed GTVLRR provides a clearer map for anomalies compared to the other detectors. The ROC curves with 95% confidence interval on the probability of false alarm are shown
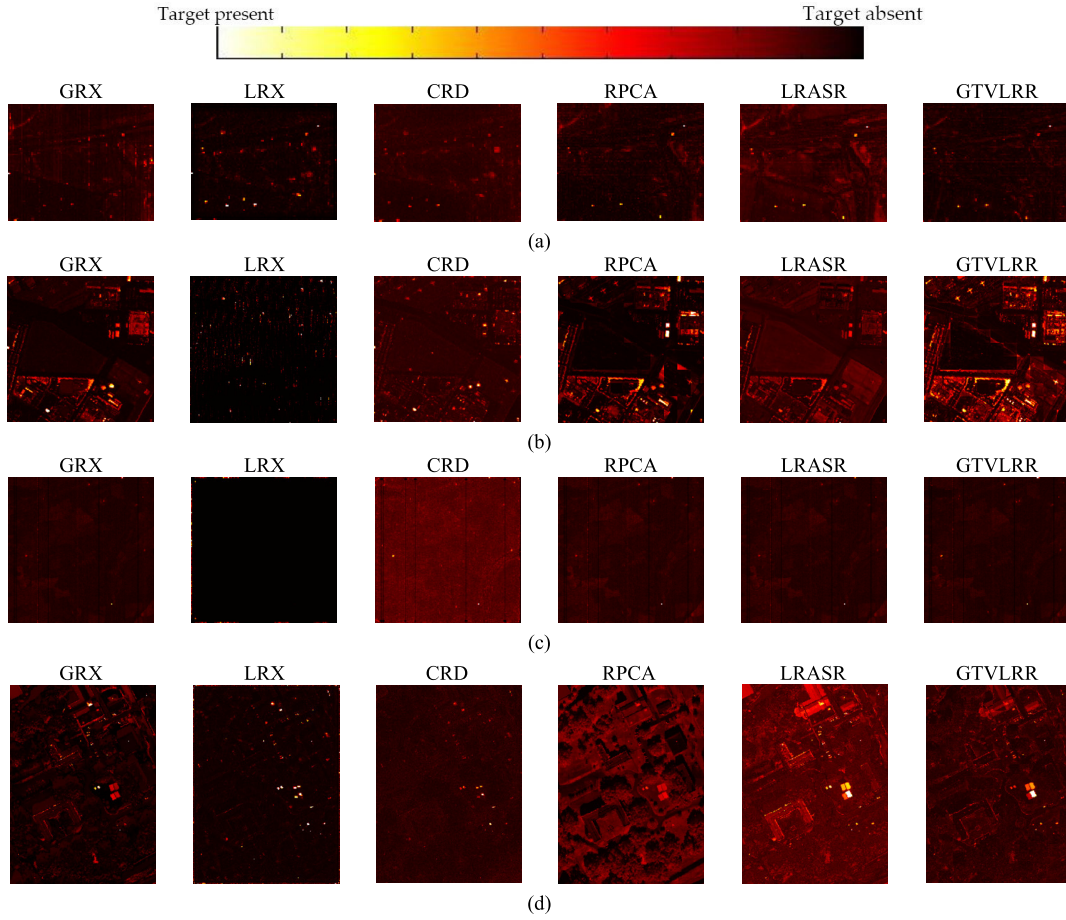
Fig. 17. 2-D plots of detection results obtained by different detectors for the four real data sets. (a) HYDICE. (b) San Diego. (c) Hyperion. (d) MUUFL.
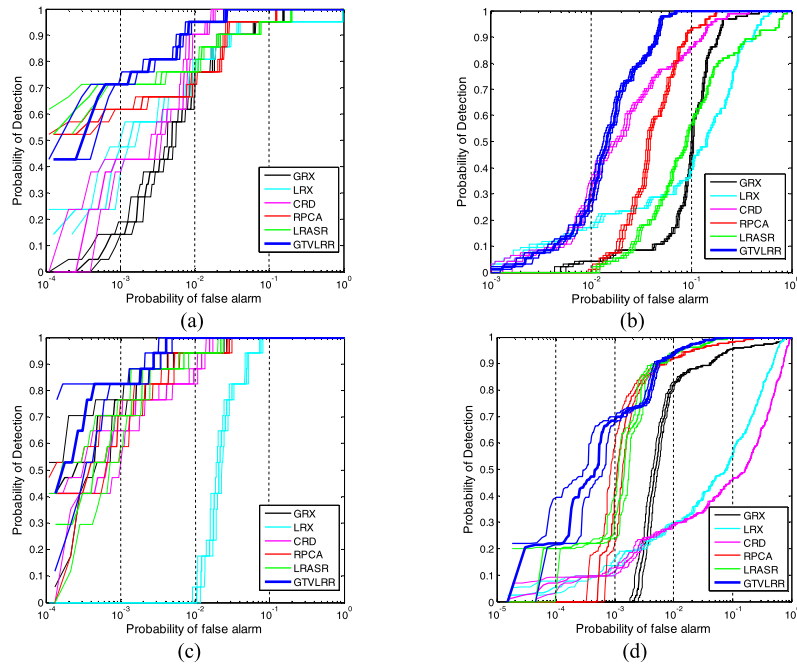

Fig. 18. ROC curves obtained by different detectors with 95% confidence interval on the probability of false alarm for the four real data sets. (a) HYDICE. (b) San Diego. (c) Hyperion. (d) MUUFL.

in Fig. 18(b). From these figures, it can be evidently observed that GTVLRR exhibits higher probabilities of detection compared to other detectors when the probability of false alarm is greater than 10^{-2}. The AUC values provided in Table IV also indicate that the proposed detector outperforms the other detectors.

TABLE IV
AUC VALUES OBTAINED BY DIFFERENT DETECTORS
FOR THE FOUR REAL DATA SETS

Data set	GRX	LRX	CRD	RPCA	LRASR	GTVLRR
HYDICE	0.9841	0.9475	0.9955	0.9884	0.9847	**0.9976**
San Diego	0.8908	0.8343	0.9064	0.9845	0.9850	**0.9928**
Hyperion	0.9978	0.9737	0.9972	0.9973	0.9978	**0.9994**
MUUFL	0.9699	0.8409	0.7422	0.9931	0.9961	**0.9972**

For the Hyperion data set, M is chosen as $M = 10$. The dual windows (win$_{out}$, win$_{in}$) for LRX and CRD are set to (17, 5) and (11, 3), respectively, and the regularization parameter in RPCA is set to $\lambda = 0.001$ after extensive searching. The 2-D plots of detection results are shown in Fig. 17(c). We can find that all the detectors produce similar detection maps for anomalies. The proposed GTVLRR performs better in background suppression. The ROC curves with 95% confidence interval on the probability of false alarm are shown in Fig. 18(c). All the detectors can achieve satisfactory results except LRX, and the ROC curves with confidence interval overlap with each other when the probability of false alarm is lower than 2×10^{-3}. However, the proposed GTVLRR achieves the best overall detection performance compared with other detectors, which can also be observed from the AUC values reported in Table IV.

For the MUUFL data set, M is chosen as $M = 10$. The dual windows (win$_{out}$, win$_{in}$) for LRX and CRD are both set to (17, 13) after extensive searching, and the best setting for the regularization parameter in RPCA is $\lambda = 0.008$. The 2-D plots of detection results are shown in Fig. 17(d). All the methods can provide a distinct map for the anomalies except those with dual windows, i.e., LRX and CRD. GTVLRR performs better than RPCA and LRASR. This demonstrates that the introduction of spatial information and local manifold structure is beneficial for discrimination. The ROC curves with 95% confidence interval on the probability of false alarm and the AUC values are shown in Fig. 18(d) and Table IV, respectively. It can be concluded that GTVLRR achieves the best overall result among all the detectors.

C. Summary

We conducted extensive experiments on both simulated and real hyperspectral data sets. To sum up, the advantages of the proposed GTVLRR can be listed as follows.

1) *Effectiveness:* The proposed GTVLRR can fully exploit the properties of the hyperspectral images, i.e., the high spectral correlations, spatial homogeneity, and local geometrical structure. The detection accuracies obtained by GTVLRR are satisfactory, illustrating that the proposed GTVLRR is very effective for hyperspectral anomaly detection.

2) *Robustness to Noise:* The incorporation of graph and TV regularizers also greatly improves the robustness of the proposed detector to additive noise; this is mainly owing to the fact that the introduced information can enhance the discrimination between the intrinsic structure of the data and the noise. The experimental results show that

satisfactory detection results can be achieved even when the noise level is high.

3) *Simplicity of Parameter Setting:* A fixed-parameter setting is used throughout the experiments, except the number of clusters M in the construction of a dictionary, which can be automatically determined by means of Hysime algorithm. This default setting provides satisfactory results for all the test images, which is mainly attributed to the considered characteristics of hyperspectral images from different points of view contributing to each other.

V. CONCLUSION

In this article, we have proposed a novel hyperspectral anomaly detection method based on the GTVLRR model. Three issues have been taken into consideration to fully exploit the characteristics of hyperspectral images from completely different aspects: First of all, LRR is adopted to capture the high spectral correlations of all pixels by exploiting the intrinsic low-rank property. Second, in order to preserve the spatial relationships in hyperspectral images, the TV regularization is included in the original LRR formulation. Third, the graph regularization term is incorporated as well with the purpose of maintaining the local geometrical structure in the hyperspectral data. In this way, a more discriminating representation can be learned. Experimental results have demonstrated that the proposed detector can provide finer detection maps for anomalies compared with other methods.

The computational cost of GTVLRR remains relatively high and should be cut down for practical use. In this regard, more efficient optimization techniques and implementations will be investigated. Another issue worthy of further study is the setting of the three tradeoff parameters in our proposed model. Although satisfactory results are obtained for all the test data sets with a fixed setting, the performance of the proposed detector on each real data set might not be optimal. In future work, we will adopt a recently proposed approach named multi-objective optimization [57] to solve this issue.

REFERENCES
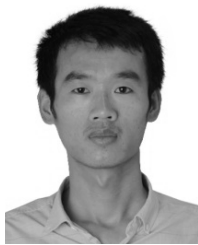
[1] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.
[2] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.
[3] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.
[4] D. Manolakis and G. S. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[5] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging—An emerging process analytical tool for food quality and safety control," *Trends Food Sci. Technol.*, vol. 18, no. 12, pp. 590–598, Dec. 2007.

[6] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, Jun. 2009.

[7] F. A. Kruse, J. W. Boardman, and J. F. Huntington, "Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1388–1400, Jun. 2003.

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

[9] C. Jiao and A. Zare, "Functions of multiple instances for learning target signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4670–4686, Aug. 2015.

[10] 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.

[11] J. Bolton and P. Gader, "Application of multiple-instance learning for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 889–893, Sep. 2011.

[12] A. Manandhar, P. A. Torrione, L. M. Collins, and K. D. Morton, "Multiple-instance hidden Markov model for GPR-based landmine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1737–1745, Apr. 2015.

[13] 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, Dec. 2018.

[14] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.

[15] 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.

[16] 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.

[17] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.

[18] S. Matteoli, M. Diani, and G. Corsini, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Opt. Eng.*, vol. 49, no. 4, pp. 046201-1–046201-16, 2010.

[19] 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.

[20] N. Billor, A. S. Hadi, and P. F. Velleman, "BACON: Blocked adaptive computationally efficient outlier nominators," *Comput. Stat. Data Anal.*, vol. 34, no. 3, pp. 279–298, Sep. 2000.

[21] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[22] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jun. 2011.

[23] 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.

[24] Q. Ling, Y. Guo, Z. Lin, and W. An, "A constrained sparse representation model for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2358–2371, Apr. 2019.

[25] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[26] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.

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

[28] 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.

[29] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[30] 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.

[31] Y. Niu and B. Wang, "Hyperspectral anomaly detection based on low-rank representation and learned dictionary," *Remote Sens.*, vol. 8, no. 4, p. 289, 2016.

[32] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2011.

[33] B. Yver and R. Marion, "A theoretical framework for hyperspectral anomaly detection using spectral and spatial *a priori* information," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 436–440, Jul. 2007.

[34] E. A. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 2, pp. 506–517, Mar. 1998.

[35] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: Clutter adaptation in anomaly detection," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1855–1871, Aug. 2000.

[36] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.

[37] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing System*, vol. 14. 2002, pp. 585–591.

[38] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[39] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[40] M. Zheng *et al.*, "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1336, May 2011.

[41] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[42] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 504–517, Mar. 2016.

[43] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, Jul. 2011.

[44] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[45] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 460–489, 2005.

[46] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: AMS, 1997.

[47] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[48] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 2002.

[49] X. Cui, Y. Tian, L. Weng, and Y. Yang, "Anomaly detection in hyperspectral imagery based on low-rank and sparse decomposition," *Proc. SPIE*, vol. 9069, Jan. 2014, Art. no. 90690R.

[50] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[51] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Dept. Elect. Comput. Eng., Univ. Illinois Urbana-Champaign, Champaign, IL, USA, Tech. Rep. UILU-ENG-09-2215, 2009.

[52] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Apr. 2008.

[53] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 646–650, Jul. 2011.

[54] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.

[55] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[56] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "MUUFL gulfport hyperspectral and LiDAR airborne data set," Dept. Comput. Inf. Sci. Eng., Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570, Oct. 2013.

[57] Z. Cai and Y. Wang, "A multiobjective optimization-based evolutionary algorithm for constrained optimization," *IEEE Trans. Evol. Comput.*, vol. 10, no. 6, pp. 658–675, Dec. 2006.

Bin Wang (M'03–SM'11) received the B.S. degree in electronic engineering and the M.S. degree in communication and electronic systems from Xidian University, Xi'an, China, in 1985 and 1988, respectively, and the Ph.D. degree in system science from Kobe University, Kobe, Japan, in 1999.

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

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

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

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