Anomaly Detection in Hyperspectral Images Based on Low-Rank and Sparse Representation

Yang Xu, *Student Member, IEEE*, Zebin Wu, *Member, IEEE*, Jun Li, *Member, IEEE*, Antonio Plaza, *Fellow, IEEE*, and Zhihui Wei

***Abstract*—A novel method for anomaly detection in hyperspectral images (HSIs) is proposed based on low-rank and sparse representation. The proposed method is based on the separation of the background and the anomalies in the observed data. Since each pixel in the background can be approximately represented by a background dictionary and the representation coefficients of all pixels form a low-rank matrix, a low-rank representation is used to model the background part. To better characterize each pixel's local representation, a sparsity-inducing regularization term is added to the representation coefficients. Moreover, a dictionary construction strategy is adopted to make the dictionary more stable and discriminative. Then, the anomalies are determined by the response of the residual matrix. An important advantage of the proposed algorithm is that it combines the global and local structure in the HSI. Experimental results have been conducted using both simulated and real data sets. These experiments indicate that our algorithm achieves very promising anomaly detection performance.**

***Index Terms*—Anomaly detection, dictionary construction, hyperspectral image (HSI) analysis, low-rank representation (LRR), sparse representation.**

I. Introduction

HYPERSPECTRAL images (HSIs) convey abundant information about the spectral characteristics of materials, with hundreds or even thousands of bands covering specific wavelengths [1]. The spectrum of each hyperspectral pixel can be viewed as a vector with each entry representing the radiance of reflectance value at each spectral band [2]. Since different materials normally reflect electromagnetic energy at different and specific wavelengths, HSI data are suitable for target detection, which have been of great interest in many military and civilian applications for several years [2]–[5]. The goal of target detection is to separate the specific target spectra or anomalous properties. Anomaly detection can be seen as target detection without any prior information about the target signature.

In anomaly detection, pixels that have a significantly different spectral signature from their neighboring background clutter pixels are defined as spectral anomalies. The well-known Reed–Xiaoli (RX) algorithm [6] is based on the assumption that the background follows a multivariate normal distribution. Then, the RX detector uses the probability density functions of the multivariate normal distribution to measure the probability of the test pixel to be part of the background. The solution of the resulting generalized likelihood ratio test turns out to be the Mahalanobis distance between the spectral vectors of an input test pixel and its surrounding neighbors. However, in a real hyperspectral scene, a multivariate normal distribution is too simple to describe the complicated background. Moreover, due to the existence of noisy and other anomalous pixels, the estimated covariance matrix and the mean vector as a form of background representation may not be accurate. To overcome this limitation, some improved methods have been proposed. For example, the regularized-RX [7] regularizes the covariance matrix estimated by all HSI pixels. The segmented-RX [8] is a recently proposed method that uses a clustering of all image pixels. The weighted-RX and linear-filter-based-RX methods are introduced in [9], aiming at improving the background information estimation. Moreover, kernel-based methods such as kernel-RX [10], [11] and support vector data description [12], [13] were proposed based on the kernel theory for extending the original space to a higher dimensional feature space. These methods can deal with very high dimensional data. Moreover, some non-RX-based methods have been proposed recently. A random-selection-based anomaly detector was presented in [14]. By randomly selecting representative background pixels and employing a sufficient number of random selections, the contamination of the background statistics was reduced. The discriminative metric learning anomaly method is described in [15]. It exploits a robust anomaly degree metric for increasing the separability between anomaly pixels and other background pixels using discriminative information. Subspace-based methods are presented in [16] and [17]. To capture local spectral variations, multiple-window anomaly detection is developed in [18].

Manuscript received March 10, 2015; revised June 30, 2015; accepted September 23, 2015. Date of publication November 9, 2015; date of current version March 9, 2016. This work was supported in part by the National Natural Science Foundation of China under Grants 61471199 and 11431015 and in part by the Fundamental Research Funds for the Central Universities under Grant 30915012204. *(Corresponding author: Zebin Wu.)*

Y. Xu and Z. Wei are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: xuyangth90@gmail.com).

Z. Wu is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China, and also with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres E-10003, Spain (e-mail: zebin.wu@gmail.com).

J. Li is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China.

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres E-10003, Spain.

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

Digital Object Identifier 10.1109/TGRS.2015.2493201

In addition to the aforementioned methods, recently, representation-based methods have gained much attention. These methods assume that hyperspectral signatures can be represented by using a dictionary. With different constraints on the representation coefficients, we can obtain different detectors. The sparse-representation-based detector (SRD) is introduced in supervised hyperspectral target detection [19], [20] and assumes that each sample can be represented by a few atoms in the dictionary. The collaborative-representation-based detector is proposed in [21]. It is based on the concept that each pixel in the background can be approximately represented by its spatial neighborhoods, whereas anomalies cannot. The representation is the linear combination of neighboring pixels, and the collaboration among these pixels is reinforced by the ℓ_2-norm minimization of the representation weight vector. However, none of these methods takes the correlations of all the pixels in the HSI into consideration; thus, the global information is not accounted for in these methods.

In this paper, a novel anomaly detection method based on low-rank and sparse representation (LRASR) is proposed. As opposed to other representation-based methods, the proposed method is built on the separation of the anomaly part and background part, and the background information is contained in the lowest rank representation of the HSI pixels. Low-rank representation (LRR) [22], [23] can be used to find the lowest rank representation of all the pixels jointly. Then, the anomaly part can be obtained by the residual of the original image and the recovered background part, using the lowest rank representation. In this way, the relationship of all HSI pixels is characterized from a global viewpoint. Then, the local structure of each pixel's coefficient is of great importance for better representation. A sparsity criterion is designed to characterize the local structure of the data set in the proposed LRASR, which gives an accurate representation of the observed data. Moreover, the background dictionary has a great impact on the representation power. In anomaly detection, the dictionary should consist of the background pixels and cover all the background classes. Thus, a novel dictionary construction strategy is proposed in our method to make the representation more stable and discriminative. The main contributions of this paper can be therefore summarized as follows.

1) To the best of our knowledge, this is the first time that the LRR is adopted for anomaly detection purposes in HSI. The background information is characterized by the low rankness of the representation coefficients, and the anomaly information is contained in the residual.
2) To better describe the local structure of each pixel's representation, a sparsity-inducing regularizer is included in the proposed model, resulting in a more accurate representation.
3) The construction of the dictionary takes two factors into consideration: One is the fact that the dictionary is composed of the background pixels, and the other is that it contains all the background classes.

The remainder of this paper is organized as follows. Section II provides a detailed description of the proposed LRASR detector. In Section III, both the simulated experiment and real data experiments are described and analyzed, followed by the conclusions in Section IV.

II. LRASR for Anomaly Detection

A. LRR for Anomaly Detection

Consider that N pixels form a band HSI $\mathbf{X} = \{x_i\}_{i=1}^{N} \in \mathbb{R}^{B\times N}$. In HSI, an anomalous pixel should be different from the background pixels. Moreover, there usually exists strong correlation among the background pixels, i.e., the background pixels can be represented by some of the other background pixels. This means that the matrix $\mathbf{X}$ can be decomposed into a background part and an anomalous part as follows:

$$\mathbf{X} = \mathbf{DS} + \mathbf{E} \tag{1}$$

where $\mathbf{DS}$ denotes the background part, $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_m]$ is the background dictionary formed by the background pixels (m is the total number of atoms in the dictionary), $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N]$ denotes the representation coefficients, and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_N]$ denotes the remaining part corresponding to the anomalies. This means that the original data can be reconstructed by the background dictionary.

There are many feasible solutions to problem (1). To address this issue, we need some criteria for characterizing matrices $\mathbf{S}$ and $\mathbf{E}$. On the one hand, only a very small fraction of the pixels belong to anomalies which means that matrix $\mathbf{E}$ is sparse. On the other hand, the spectrum of each pixel corresponds to one kind of material (which is called pure pixel) or to the mixture of several materials (which is called mixed pixel). As the spectrum of every material can be represented in a subspace, all the spectra in the HSI can be drawn from multiple subspaces. Thus, the coefficients matrix $\mathbf{S}$ should give the lowest-rank representation of all spectra jointly. In summary, for matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ with each $\mathbf{x}_i$ representing the ith pixel, it is appropriate to infer the anomalies by solving the following LRR [22] problem:

$$\begin{aligned} \min_{\mathbf{S},\mathbf{E}} \quad & \text{rank}(\mathbf{S}) + \lambda\|\mathbf{E}\|_{2,1} \\ \text{s.t} \quad & \mathbf{X} = \mathbf{DS} + \mathbf{E} \end{aligned} \tag{2}$$

where rank$(\cdot)$ denotes the rank function, parameter $\lambda > 0$ is used to balance the effects of the two parts, and $\|\cdot\|_{2,1}$ is the $\ell_{2,1}$ norm defined as the sum of ℓ_2 norm of the column of a matrix

$$\|\mathbf{E}\|_{2,1} = \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{d} \left([\mathbf{E}]_{j,i}\right)^2} \tag{3}$$

where $[\mathbf{E}]_{j,i}$ is the entry of $\mathbf{E}$. The $\ell_{2,1}$ norm encourages the columns of $\mathbf{E}$ to be zero, which assumes that the corruptions are "sample specific," i.e., some pixels are corrupted and the others are clean. For a column vector corresponding to the ith pixel, a larger magnitude implies that the pixel is more anomalous. As a consequence, the matrix $\mathbf{E}$ naturally measures the anomalies. Different from the method proposed in [24]

which used robust principal component analysis (RPCA) [25] to separate the original data to a low-rank part and the sparse error part, we use LRR to separate the data. As pointed in [22], RPCA relies on the assumption that the data lie in a single low-rank subspace. However, due to the existence of mixed pixels, the pixels of an HSI are drawn from multiple subspaces. Thus, it may not be appropriate to use RPCA in this context. On the contrary, by choosing an appropriate dictionary, the LRR (which is regarded as a generalization of RPCA) can recover the underlying multiple subspaces.

Algorithm 1 LADMAP Algorithm for LRASR

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

Initialize: $\mathbf{S}_0 = \mathbf{J}_0 = \mathbf{E}_0 = \mathbf{Y}_{1,0} = \mathbf{Y}_{2,0}$, $\mu_0 = 0.01$, $\mu_{\max} = 10^{10}$, $\rho_0 = 1.1$, $\varepsilon_1 = 10^{-6}$, $\varepsilon_2 = 10^{-2}$, $\eta_1 = \|\mathbf{D}\|_2^2$, $k = 0$.

1: **while** $\|\mathbf{X} - \mathbf{D}\mathbf{S}_0 - \mathbf{E}_0\|_F / \|\mathbf{X}\|_F \geq \varepsilon_1$ or $\mu_k \max(\sqrt{\eta_1}\|\mathbf{S}_k - \mathbf{S}_{k-1}\|_F, \|\mathbf{J}_k - \mathbf{J}_{k-1}\|_F, \|\mathbf{E}_k - \mathbf{E}_{k-1}\|_F)/\|\mathbf{X}\|_F \geq \varepsilon_2$ **do**

2: Update variable $\mathbf{S}_{k+1}$:

$$\mathbf{S}_{k+1} = \Theta_{(\eta_1\mu_k)^{-1}}\left(\mathbf{S}_k + \left[\mathbf{D}^T(\mathbf{X} - \mathbf{D}\mathbf{S}_k - \mathbf{E}_k + \mathbf{Y}_{1,k}/\mu_k) - (\mathbf{S}_k - \mathbf{J}_k + \mathbf{Y}_{2,k}/\mu_k)\right]/\eta_1\right)$$

3: Update variable $\mathbf{J}_{k+1}$:

$$\mathbf{J}_{k+1} = \mathcal{S}_{\beta\mu_k^{-1}}(\mathbf{S}_{k+1} + \mathbf{Y}_{2,k}/\mu_k)$$

4: Update variable $\mathbf{E}_{k+1}$:

$$\mathbf{E}_{k+1} = \Omega_{\lambda\mu_k^{-1}}(\mathbf{X} - \mathbf{D}\mathbf{S}_{k+1} + \mathbf{Y}_{1,k}/\mu_k)$$

5: Update Lagrange multipliers as follows:

$$\mathbf{Y}_{1,k+1} = \mathbf{Y}_{1,k} + \mu_k(\mathbf{X} - \mathbf{D}\mathbf{S}_{k+1} - \mathbf{E}_{k+1}).$$
$$\mathbf{Y}_{2,k+1} = \mathbf{Y}_{2,k} + \mu_k(\mathbf{S}_{k+1} - \mathbf{J}_{k+1})$$

6: Update μ as follows:

$$\mu_{k+1} = \min(\mu_{\max}, \rho\mu_k) \text{ where}$$
$$\rho = \begin{cases} \rho_0, & \text{if } \mu_k \max\left(\sqrt{\eta_1}\|\mathbf{S}_{k+1} - \mathbf{S}_k\|_F, \|\mathbf{J}_{k+1} - \mathbf{J}_k\|_F, \|\mathbf{E}_{k+1} - \mathbf{E}_k\|_F\right)/\|\mathbf{X}\|_F \leq \varepsilon_2 \\ 1, & \text{otherwise} \end{cases}$$

7: Update $k : k \leftarrow k + 1$.

8: **end while**

Output: an optimal solution $(\mathbf{S}_k, \mathbf{J}_k, \mathbf{E}_k)$.

However, solving problem (2) is NP-hard. Fortunately, it was suggested by matrix completion methods [25] that the following convex optimization provides a good surrogate for problem (2):

$$\min_{\mathbf{S},\mathbf{E}} \quad \|\mathbf{S}\|_* + \lambda\|\mathbf{E}\|_{2,1}$$
$$\text{s.t} \quad \mathbf{X} = \mathbf{D}\mathbf{S} + \mathbf{E} \tag{4}$$

where $\|\cdot\|_*$ denotes the matrix nuclear norm (sum of the singular values of a matrix) [26]. Once the representation process is finished, the anomalies for the ith spectrum $T(\mathbf{x}_i)$ can be determined by the response of the residual matrix $\mathbf{E}$ as follows:

$$T(\mathbf{x}_i) = \|[\mathbf{E}^*]_{:,i}\|_2 = \sqrt{\sum_j ([\mathbf{E}^*]_{j,i})^2} \tag{5}$$

where $\|[\mathbf{E}^*]_{:,i}\|_2$ denotes the ℓ_2 norm of the ith column of $\mathbf{E}^*$. If it is larger than a threshold, then $\mathbf{x}_i$ is claimed to be an anomalous pixel.

B. Sparse Regularization for LRR

As seen in [22], the low rankness criterion is superior at capturing the global structure of observed data $\mathbf{X}$. However, each spectrum has its own local structure. The more accurate the description of the local structure, the more accurate the representation of the observed data. The sparse signal representation has proven to be a powerful tool in many areas [27], [28]. This success is mainly due to the fact that most natural signals can be sparsely represented by a few coefficients carrying the most important information with respect to a certain dictionary or basis set [28]. In HSI target detection, sparsity-based target detection algorithms are widely used. The basic sparsity-based detector uses a similar sparsity model proposed in [28] to sparsely represent a test image by a few training samples, including both target and background samples, and then, it directly employs the reconstruction residuals to perform the detection. In the anomaly detection stage, as most of the samples are background pixels, they have a sparse representation in terms of the background dictionary. Therefore, the sparse nature of the matrix allows us to describe the local struct model with the sparse regularization can be written as

$$\min_{\mathbf{S},\mathbf{E}} \quad \|\mathbf{S}\|_* + \beta\|\mathbf{S}\|_1 + \lambda\|\mathbf{E}\|_{2,1}$$
$$\text{s.t} \quad \mathbf{X} = \mathbf{D}\mathbf{S} + \mathbf{E} \tag{6}$$

where $\|\cdot\|_1$ is the ℓ_1 norm of a matrix, i.e., the sum of the absolute value of all entries in the matrix; $\beta > 0$ is a parameter to trade off low rankness and sparsity. The model (6) incorporates the global structure by the low-rank property and the local structure by the sparsity property, which results in a more accurate representation of the original data. Thus, the residual matrix $\mathbf{E}$ provides a better description of the anomalies.

The LRASR problem (6) could be solved by the popular alternating direction method [22], [29]. However, two auxiliary variables need to be introduced when solving (6), and expensive matrix inversions are required in each iteration. Thus, a newly developed method called the linearized alternating direction method with adaptive penalty (LADMAP) [30] is adopted to solve (6).

Algorithm 2 Anomaly detection algorithm for HSI based on LRASR

Input: data matrix $\mathbf{X}$, parameters $\beta > 0, \lambda > 0, K, P$
1: Divide the data matrix $\mathbf{X}$ into K parts using K-means algorithm, s.t.
$\mathbf{X} = \cup_{i=1,\ldots,K}\mathbf{X}^i$,$\mathbf{X}^i \cap \mathbf{X}^j = \emptyset, \quad i \neq j, \quad i, j = 1, \ldots K$,
Denote N_i the number of pixels in $\mathbf{X}^i$ and $\mathbf{D} = \emptyset$.
2: for $i = 1 : K$
2.1: If $N_i < P$
Skip and go to Step 2;
2.2: compute the mean vector $\boldsymbol{\mu}$ and covariance matrix $\sum$ of data $\{x_j | x_j \in \mathbf{X}^i, j = 1 \ldots . N_i\}$.
2.3: compute the pre-detection result:
$PD(x_j) = (\mathbf{x}_j - \boldsymbol{\mu})^T \sum^{-1} (\mathbf{x}_j - \boldsymbol{\mu})\ j = 1, 2, \ldots, N_i$
2.4: choose P pixels $\mathbf{D}^i = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_P]$ in $\{\mathbf{x}_j | \mathbf{x}_j \in \mathbf{X}^i, j = 1 \ldots . N_i\}$, s.t.
$PD(\mathbf{x}_i) < PD(\mathbf{x}_j), \mathbf{x}_i \in D^i, \mathbf{x}_j \in \{\mathbf{x}_j | \mathbf{x}_j \in \mathbf{X}^i, j = 1, \ldots, N_i\} \setminus D^i$
2.5: $\mathbf{D} = \mathbf{D} \cup \mathbf{D}_i$
end
3: Solve the following problem using Algorithm 1,

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

And obtain the optimal solution $(\mathbf{S}^*, \mathbf{E}^*)$.
4: compute $T(\mathbf{x}_i) = \|[\mathbf{E}^*]_{:,i}\|_2 = \sqrt{\sum_j ([\mathbf{E}^*]_{j,i})^2}$, $i = 1, 2, \ldots, N$
Output: Anomaly detection map.

To make the objective function separable, we introduce an auxiliary variable $\mathbf{J}$ which satisfies $\mathbf{S} = \mathbf{J}$; then, we can replace the second term $\|\mathbf{S}\|_1$ in the objective function with $\|\mathbf{J}\|_1$. Thus, the original problem (6) can be converted to the following problem:

$$\begin{aligned} \min_{\mathbf{S},\mathbf{E}} \quad & \|\mathbf{S}\|_* + \beta\|\mathbf{J}\|_1 + \lambda\|\mathbf{E}\|_{2,1} \\ \text{s.t} \quad & \mathbf{X} = \mathbf{DS} + \mathbf{E}, \mathbf{S} = \mathbf{J} \end{aligned} \tag{7}$$

The augmented Lagrangian function of problem (7) is

$$\begin{aligned} &L(\mathbf{S}, \mathbf{J}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2, \mu) \\ &\quad = \|\mathbf{S}\|_* + \beta\|\mathbf{J}\|_1 + \lambda\|\mathbf{E}\|_{2,1} + \langle \mathbf{Y}_1, \mathbf{X} - \mathbf{DS} - \mathbf{E}\rangle \\ &\qquad + \langle \mathbf{Y}_2, \mathbf{S} - \mathbf{J}\rangle + \frac{\mu}{2}\left(\|\mathbf{X} - \mathbf{DS} - \mathbf{E}\|_F^2 + \|\mathbf{S} - \mathbf{J}\|_F^2\right) \\ &\quad = \|\mathbf{S}\|_* + \beta\|\mathbf{J}\|_1 + \lambda\|\mathbf{E}\|_{2,1} \\ &\qquad + f(\mathbf{S}, \mathbf{J}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2, \mu) - \frac{1}{2\mu}\left(\|\mathbf{Y}_1\|_F^2 + \|\mathbf{Y}_2\|_F^2\right) \end{aligned} \tag{8}$$

where $(\mathbf{Y}_1, \mathbf{Y}_2)$ are Lagrange multipliers, $\mu > 0$ is the penalty parameter, and

$$\begin{aligned} &f(\mathbf{S}, \mathbf{J}, \mathbf{E}, \mathbf{Y}_1, \mathbf{Y}_2, \mu) \\ &\quad = \frac{\mu}{2}\left(\|\mathbf{X} - \mathbf{DS} - \mathbf{E} + \mathbf{Y}_1/\mu\|_F^2 + \|\mathbf{S} - \mathbf{J} + \mathbf{Y}_2/\mu\|_F^2\right) \end{aligned} \tag{9}$$

The LADMAP is a multiple-variable optimization problem, which can be solved by updating one variable alternately by minimizing L with the other variables fixed. Suppose that we are in the kth iteration; the problem can be divided into the following subproblems.

1) Fix $\mathbf{J}$ and $\mathbf{E}$, and update $\mathbf{S}$; the objective function can be written as follows:

$$\begin{aligned} \mathbf{S}_{k+1} &= \arg\min_{\mathbf{S}} \|\mathbf{S}\|_* \\ &\quad + \langle \nabla_{\mathbf{S}} f(\mathbf{S}_k, \mathbf{J}_k, \mathbf{E}_k, \mathbf{Y}_{1,k}, \mathbf{Y}_{2,k}, \mu_k), \mathbf{S} - \mathbf{S}_k\rangle \\ &\quad + \frac{\eta_1\mu_k}{2}\|\mathbf{S} - \mathbf{S}_k\|_F^2 \\ &= \arg\min_{\mathbf{S}} \|\mathbf{S}\|_* + \frac{\eta_1\mu_k}{2} \\ &\quad \times \left\|\mathbf{S} - \mathbf{S}_k + \left[-\mathbf{D}^T(\mathbf{X} - \mathbf{DS}_k - \mathbf{E}_k + \mathbf{Y}_{1,k}/\mu k)\right.\right. \\ &\qquad \left.\left. + (\mathbf{S}_k - \mathbf{J}_k + \mathbf{Y}_{2,k}/\mu_k)\right]/\eta_1\right\|_F^2 \end{aligned} \tag{10}$$

where the quadratic term f is replaced by its first-order approximation at previous iterations, and then adding a proximal term [30], $\nabla_{\mathbf{S}} f$ is the partial differential of f with respect to $\mathbf{S}$, and $\eta_1 = \|\mathbf{D}\|_2^2$.

2) Fix $\mathbf{S}$ and $\mathbf{E}$, and update $\mathbf{J}$; the objective function can be written as follows:

$$\mathbf{J}_{k+1} = \arg\min_{J} \beta\|\mathbf{J}\|_1 + \frac{\mu_k}{2}\|\mathbf{S}_{k+1} - \mathbf{J} + \mathbf{Y}_{2,k}/\mu_k\|_F^2 \tag{11}$$

3) Fix $\mathbf{S}$ and $\mathbf{J}$, and update $\mathbf{E}$; the objective function can be written as follows:

$$\begin{aligned} \mathbf{E}_{k+1} &= \arg\min_{\mathbf{E}} \lambda\|\mathbf{E}\|_{2,1} \\ &\quad + \frac{\mu_k}{2}\|\mathbf{X} - \mathbf{DS}_{k+1} - \mathbf{E} + \mathbf{Y}_{1,k}/\mu_k\|_F^2 \end{aligned} \tag{12}$$

The solution is outlined in **Algorithm 1**.

In Algorithm 1, Θ, $\mathcal{S}$, and Ω are the singular value thresholding [31], shrinkage [29], and $\ell_{2,1}$ minimization operator [22], respectively. Here, the orders of updating $\mathbf{S}$, $\mathbf{J}$, and $\mathbf{E}$ can be changed.

C. Construction of Dictionary $\mathbf{D}$

In the aforementioned algorithm, the dictionary $\mathbf{D}$ plays an important role in detecting anomalies. In sparse representation for target detection problems, the dictionary is composed of the background dictionary and the target dictionary, which are supposed to be known. However, in the case of anomaly detection, the dictionary is not known in advance and should represent the background information as much as possible. One way to construct the dictionary is to use the original data $\mathbf{X}$ directly. Although the original data $\mathbf{X}$ contains anomalies, the number of anomalous pixels is very small, which can be ignored in the representation process. However, the computation burden

(a) (b) (c)

Fig. 1. Simulated data set. (a) False color image of the whole scene, (b) false color image of the detection area, and (c) ground-truth map.

is very large since the algorithm involves a singular value decomposition of a matrix with size $M \times M$, where M is the number of atoms. Another way is to choose some of the pixels from the HSI randomly to form the dictionary. In this way, the atom number can be decreased. However, the dictionary should cover all the ground material classes except for the anomalies. Fortunately, in an HSI, most of the scene is covered by a few kinds of major materials, so pixels are much more likely to belong to these major materials than to others. If we construct the dictionary by choosing pixels randomly, the probability of choosing pixels belonging to major materials is very high, so the constructed dictionary will be composed of the major materials' pixels, and some materials with other less relevant samples will be ignored. As a result, the pixels which correspond to the material with less representative samples will be detected as anomalies.

In this paper, a new strategy for dictionary construction is adopted. First, K-means is used to divide all the pixels into K clusters $\mathbf{X} = \{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^K\}$. The Euclidean distance is used in the K-means algorithm. K should be larger than the true number of ground material classes in order to make sure that the K clusters cover all the ground materials. As the atom should be a background pixel, a prepredictive strategy is adopted in order to choose the background pixels in each cluster. Similar to the RX algorithm, the square of the Mahalanobis distance between the test pixel and the local background mean in each cluster is calculated. The RX detector assumes that the higher the value of the detection result, the more likely the pixel will be anomalous. However, noisy pixels or pixels corresponding to rare materials can be detected as outliers because they are more uncommon in the HSI.

On the other hand, pixels with small prepredictive values are background pixels with certainty. Therefore, the P pixels which give the smallest Mahalanobis distance are chosen to generate atoms in the dictionary. If the total number of pixels in the cluster is smaller than P, this cluster can be skipped as we have set K larger than the real number of ground material classes.

(a) (b)

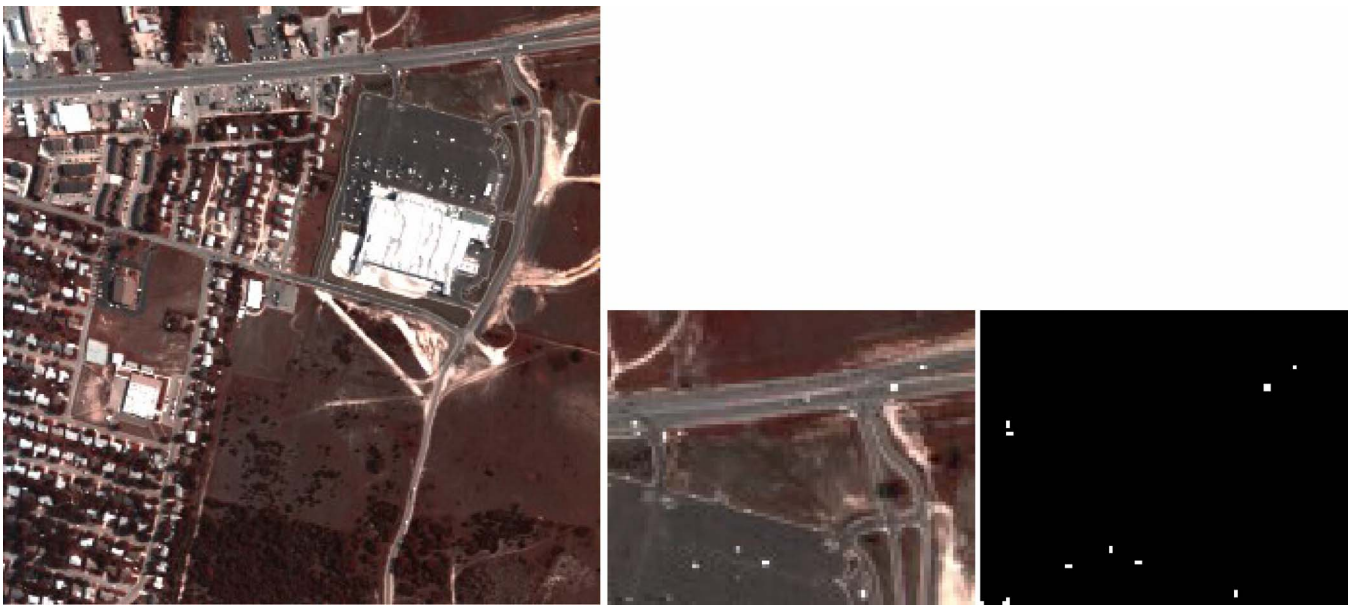
Fig. 2. First real-world data set. (a) False color image of the detection area and (b) ground-truth map.

The anomaly detection algorithm for HSIs based on LR using the clustered dictionary is summarized in Algorithm

III. Experimental Results

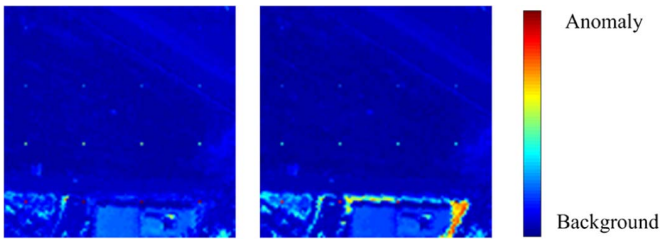
A. Data Set Description

In this paper, both simulated and real hyperspectral data sets are used to evaluate our method. The simulated data were generated based on a real HSI data set. It was collected by the Airborne Visible/Infrared Imaging Spectrometer over San Diego, CA, USA. The spatial resolution is 3.5 m per pixel. The image has 224 spectral channels in wavelengths ranging from 370 to 2510 nm. After removing the bands that correspond to the water absorption regions, low SNR, and bad bands (1–6, 33–35, 94–97, 107–113, 153–166, and 221–224), 186 available bands of the data are retained in the experiments. The whole data set has a size of 400×400, as shown in Fig. 1(a). From this



(a) (b) (c)

Fig. 3. Second real-world data set. (a) False color image of the whole image, (b) false color image of the detection area, and (c) ground-truth map.



(a) (b)

Fig. 4. Two-dimensional plots of the detection results achieved by LRASR using different dictionaries. A color nearer to red suggests that the pixels are anomalous, while a color nearer to blue suggests that the pixels belong to the background. The results are shown using (a) our dictionary and (b) a random dictionary.

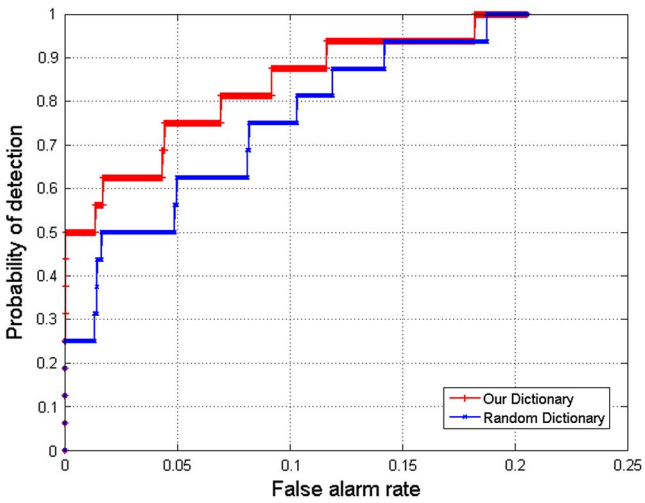


Fig. 5. ROC curves for the simulated data set using different dictionaries.

hyperspectral data set, a region with a size of 100×100 pixels is selected to form the simulated data. The anomalous pixels are simulated by the target implantation method [32]. Based on the linear mixing model, a synthetic subpixel anomaly target with

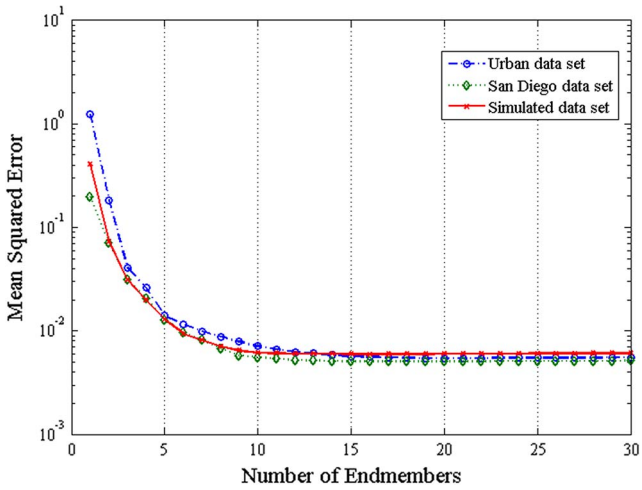


Fig. 6. Plot of MSE versus the number of endmembers.

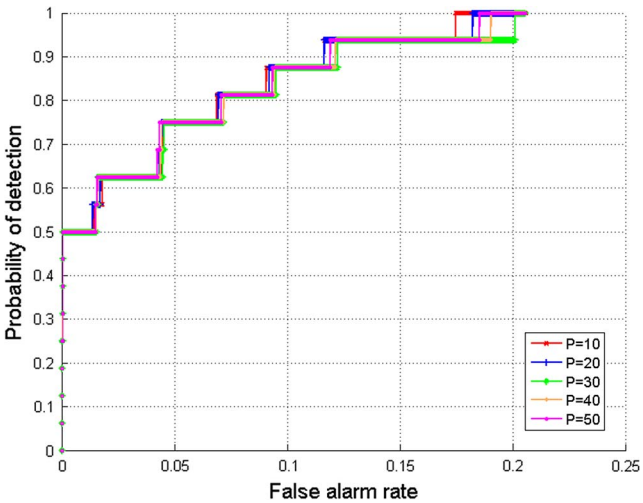


Fig. 7. ROC curves obtained for the simulated data set with different values of P.

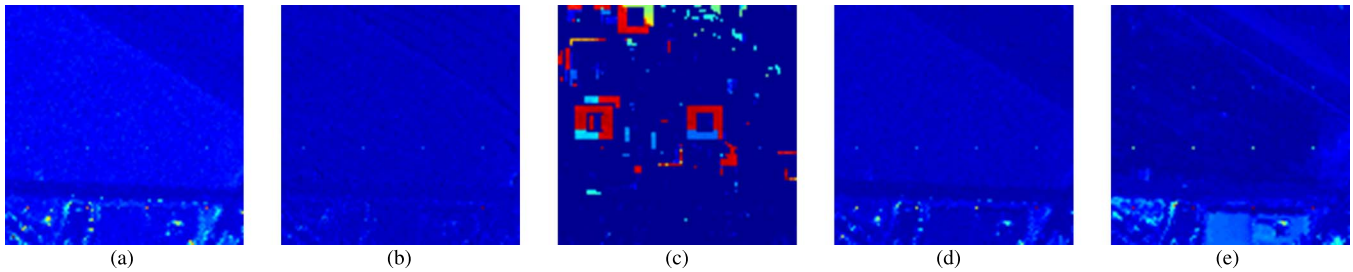


Fig. 8. Two-dimensional plots of the detection results obtained by different methods for the simulated data set: (a) Global-RX, (b) SegRX, (c) SRD, (d) RPCA-RX, and (e) LRASR.

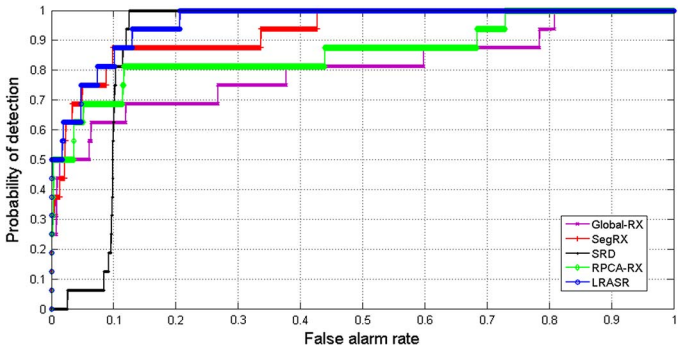


Fig. 9. ROC curves obtained by different methods for the simulated data set.

TABLE I
AUC FOR THE DETECTORS REPORTED IN FIG. 7

Algorithm	Global-RX	SegRX	SRD	RPCA-RX	LRASR
AUC	0.8050	0.9283	0.9029	0.8612	**0.9597**

spectrum $\mathbf{z}$ and a specified abundance fraction f is generated by fractionally implanting a desired anomaly with spectrum $\mathbf{t}$ in a given pixel of the background with spectrum $\mathbf{b}$ as follows [33]:

$$\mathbf{z} = f \cdot \mathbf{t} + (1 - f) \cdot \mathbf{b}. \tag{13}$$

In the image, 16 anomalous targets have been implanted. These anomalies are distributed in four rows and four columns. The abundance fractions f are 0.05, 0.1, 0.2, and 0.4 for different rows, respectively, and remain unchanged for the anomalies in the same row. The anomalous spectrum $\mathbf{t}$ is chosen outside the selected scene in the whole image, and the same anomalous spectrum $\mathbf{t}$ is applied to the 16 target pixels. It corresponds to the plane in the middle left of the whole scene. The image scene used in the simulated experiments is shown in Fig. 1(b), and the ground-truth map of anomalies is illustrated in Fig. 1(c).

The first data used for real-world detection are also part of the San Diego image. The up-left 100×100 of the scene is chosen as the test image. The scene is mainly composed of buildings with different roofs, parking aprons with different materials, an airport runway, and a small quantity of vegetation. The airplanes are the anomalies to be detected. The false color image and the ground-truth map are shown in Fig. 2(a) and (b), respectively. Fifty-seven pixels were selected as anomalies, composed of full-pixel anomalies in the main body of the airplanes and subpixel targets on the edges of the airplanes. In this scene, the anomaly is big compared to that of the simulated data set.

The second data set used in real-world experiments is a HYDICE hyperspectral data set obtained from an aircraft platform. It covers an urban area that comprises a vegetation area, a construction area, and several roads including some vehicles. The image has a spectral resolution of 10 nm and a spatial resolution of 1 m. The low-SNR and water vapor absorption bands (1–4, 76, 87, 101–111, 136–153, and 198–210) are eliminated so that 162 bands remain. The whole data set has a size of 307×307 pixels, as shown in Fig. 3(a). However, the ground truth defines that the anomalous targets are the cars and roofs embedded in the different backgrounds in the upper rightmost area of the scene. Therefore, the considered subscene consists of pixels covering this area. A color representation and the ground-truth map are shown in Fig. 3(b) and (c), respectively. The 21 anomalous target pixels are the vehicles with different sizes in the urban scene.

B. Detection Performance

First, we illustrate the effects of the dictionary constructed by our method. In our dictionary construction step, K and P are set to 15 and 20, respectively. We compare our method with another method using a different dictionary. In that method, all the atoms in the dictionary are randomly chosen from the whole image. For a fair comparison, the number of atoms is set to 300. The 2-D plots illustrating the detection results obtained by the two methods are shown in Fig. 4(a) and (b). The anomalies are obvious in both maps. However, in the results of the method using a random dictionary, many pixels on the lower rightmost corner of the scene exhibit a high detection value despite not being anomalous. This indicates that the random dictionary cannot avoid the presence of noisy pixels or pixels corresponding to rare materials being detected as anomalies. For numerical comparison, the receiver operating characteristic (ROC) curve is constructed, which acts as the classic comparison measurement for different detection methods [2], [34]. The target detection rate and false alarm rate are computed by a certain segmentation threshold. For each target detector's results, the segmentation threshold can be changed to obtain a group of target detection rate and false alarm rate for each target detector, which can be used to plot the ROC curve. A better detector would lie nearer the upper leftmost corner and result in a larger area under the curve [2]. Fig. 5 shows the ROC

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

Fig. 10. Two-dimensional plots of the detection results obtained by different methods for the real San Diego data set: (a) Global-RX, (b) SegRX, (c) SRD, (d) RPCA-RX, and (e) LRASR.

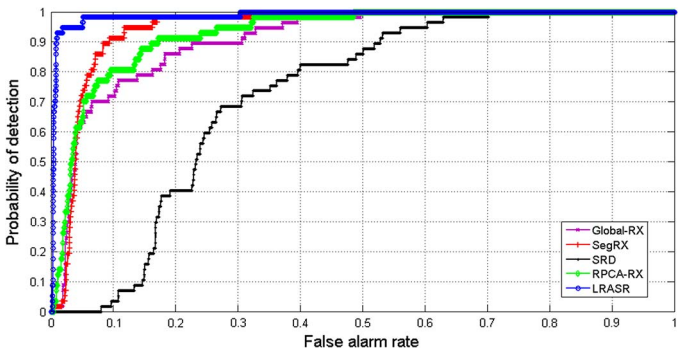


Fig. 11. ROC curves obtained by different methods for the real San Diego data set.

TABLE II
AUC FOR DETECTORS REPORTED IN FIG. 9

Algorithm	Global-RX	SegRX	SRD	RPCA-RX	LRASR
AUC	0.9090	0.9446	0.7234	0.9288	**0.9882**

curves obtained after applying the two considered methods. It can be seen that the method using our dictionary can achieve better results, which is expected because our dictionary can cover most of the background materials, thus providing more robust and reliable results.

The number of clusters should be larger than the real number of the ground materials, but this number is difficult to know *a priori*. Thus, the HySime [35] algorithm is applied to estimate the number of endmembers that can represent the real ground materials. HySime estimates the signal and the noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the minimum mean squared error (MSE) sense. Fig. 6 shows the MSE versus the number of endmembers of the three considered data sets. From Fig. 6, we can see that, when the number of endmembers is larger than 10, the MSE values are small for the three data sets. Thus, we set $K = 15$ empirically in our experiments. Then, we compare the impact of P as shown in Fig. 7. From Fig. 7, we can see that the results obtained using different values of P are almost the same. Thus, our method is robust to P. For simplicity, P is set to 20 in our experiments.

Next, we evaluate the detection performance of our proposed LRASR detector, comparing it with the conventional global RX, segmentation-based-RX (SegRX) detector, SRD, and RPCA-RX. For SegRX, the K-means clustering is first applied on the whole data to divide the data into several clusters. Then, the RX is used in each cluster to detect the anomalies. The number of clusters in SegRX is set empirically. The regularization parameters for RPCA-RX and SRD are optimized in our experiments. The parameters of window size (w_{out}, w_{in}) are set to (15, 7) for SRD after extensive searching. The 2-D plots of the detection results of all the compared algorithms are shown in Fig. 8(a)–(e). From these figures, we can see that all the methods (except SRD) can distinguish between the background and anomalies in high abundance fraction pixels. The proposed LRASR can also provide a clear discrimination in the low abundance fraction pixels. Fig. 9 illustrates the ROC curves obtained for the simulated data set. An important observation from Fig. 8 is that the SegRX outperforms the global RX and RPCA-RX. Compared to the SegRX, the proposed LRASR exhibits a slightly lower probability of detection for a low false alarm rate (when the false alarm is 0.033 to 0.047). Compared to SRD, it exhibits a lower probability of detection when the false alarm is from 0.125 to 0.2; however, the proposed LRASR is the best method in terms of the overall detection performance. Furthermore, we have also computed the area under the ROC curve (AUC) to evaluate the performance of these methods. The results are shown in Table I. The proposed LRASR achieves the highest score as expected. It is 0.03 higher than the second highest score, which is achieved by SegRX. The improvement over the other tested methods is due to the fact that our proposed LRASR performs better at both noise pixel suppression and weak anomaly detection.

For the real San Diego data set experiment, the 2-D plots for the obtained detection results are illustrated in Fig. 10(a)–(e). From this figure, it can be seen that the proposed LRASR gives a map where the anomalies are obvious. The ROC curves of all the methods are shown in Fig. 11 for illustrative purposes. The proposed LRASR achieves the highest probability of detection for all false alarm rate values. The AUC scores are provided in Table II. The proposed LRASR achieves a score that is 0.04 higher than SegRX, which is an obvious improvement. This confirms that the proposed method can outperform the traditional detectors.

For the real urban data set, the 2-D plots of the obtained detection results are shown in Fig. 12(a)–(e). The anomalies in Fig. 12(e) are more obvious, and few nonanomalous pixels have high detection values. The ROC curves and AUC scores are also shown in Fig. 13 and Table III, respectively. Although the representation-based SRD gains a higher probability of detection when the false alarm rate ranges from 0.012 to 0.051, the proposed LRASR achieves the highest AUC score among all the detectors. Therefore, it can be concluded that LRASR

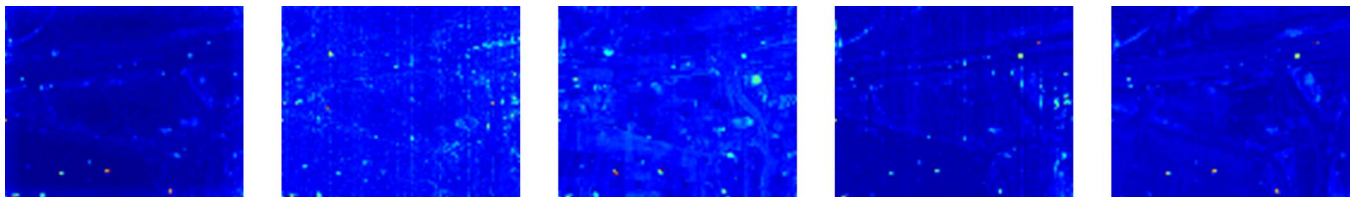


Fig. 12. Two-dimensional plots of the detection results obtained by different methods for the real urban data set: (a) Global-RX, (b) SegRX, (c) SRD, (d) RPCA-RX, and (e) LRASR.

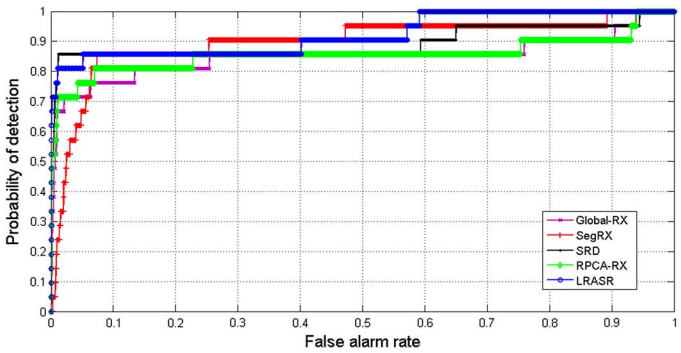


Fig. 13. ROC curves obtained by different methods for the real urban data set.

TABLE III
AUC FOR DETECTORS REPORTED IN FIG. 11

Algorithm	Global-RX	SegRX	SRD	RPCA-RX	LRASR
AUC	0.8504	0.8978	0.8928	0.8564	**0.9220**

TABLE IV
EXECUTION TIMES (IN SECONDS) FOR ALL EXPERIMENTS

	Global-RX	SegRX	SRD	RPCA-RX	LRASR
Simulated data	0.28	8.45	7.13	19.27	169.06
Real San Diego data	0.27	2.67	5.81	22.41	162.49
Real urban data	0.2	2.41	4.56	9.37	132.17

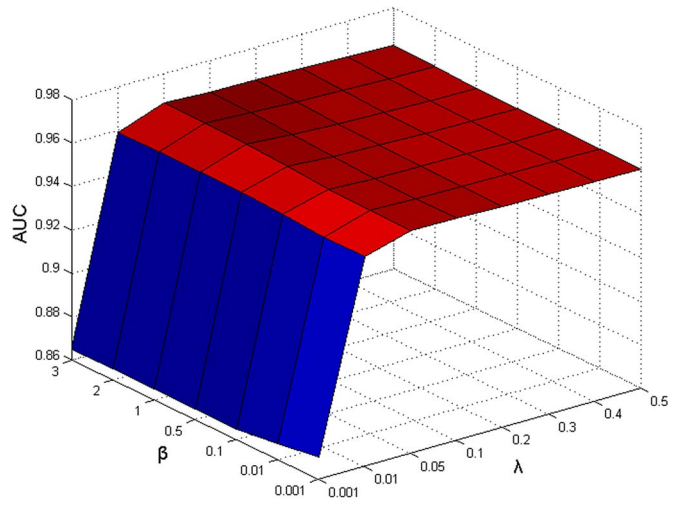


Fig. 14. Joint consideration of β and λ for the simulated data set.

is a promising method for the detection of anomalous pixels in HSIs.

The computation costs of all the aforementioned methods have also been compared. Detailed results are presented in Table IV. The algorithms are tested on a computer with a 64-b quad-core Intel Xeon CPU 3.33-GHz processor under Windows 7. It reveals that our method does take more computation time than other methods. Due to the fact that our algorithm requires hundreds of iterations to converge, the computational cost of the proposed method is an important issue to be settled in our future developments.

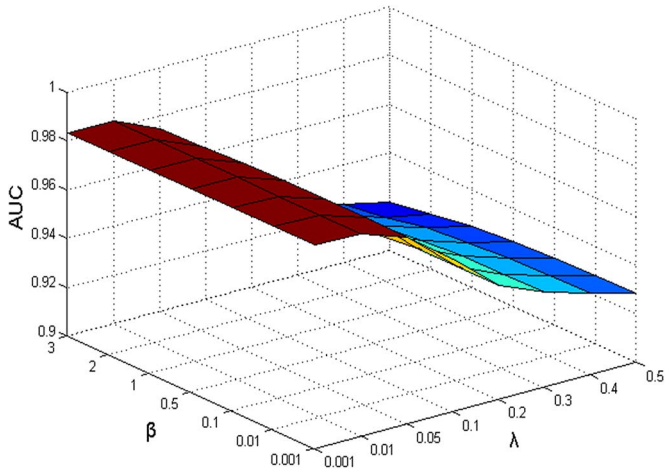


Fig. 15. Joint consideration of β and λ for the real San Diego data set.

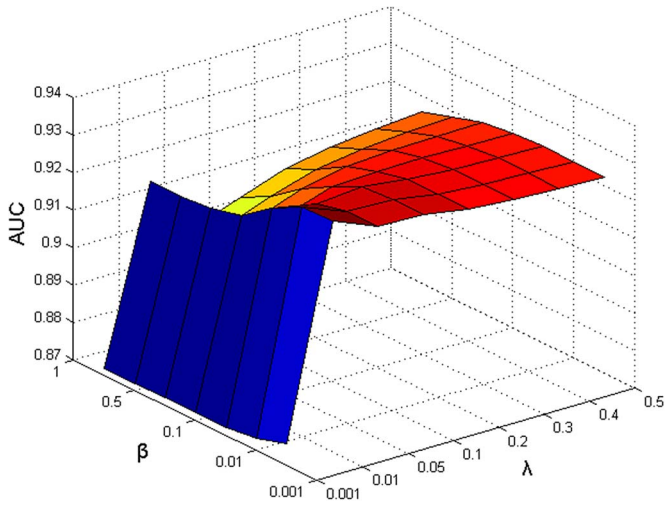


Fig. 16. Joint consideration of β and λ for the real urban data set.

C. Sensitivity to the Regularization Parameters

The proposed method involves two regularization parameters: β and λ. Normally, we can use a more robust method to

compute them by a *maximum a posteriori* (MAP) estimation problem. To improve the computational efficiency, we do not use a robust estimation method in this paper. Figs. 14–16 illustrate the obtained performance when jointly taking the two regularization parameters into consideration. β is chosen from $\{0.001, 0.01, 0.1, 0.5, 1, 2, 3\}$, and λ is chosen from $\{0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$. It reveals that, for the San Diego data set, LRASR is sensitive to β. For the simulated data set and the second real-world data set, it achieves high AUC when λ is larger than 0.001. From a general viewpoint, the changes of AUC are not obvious when $\beta, \lambda \in [0.01\ 0.1]$. For simplicity, in our experiments, we have empirically set $\beta = 0.1$ and $\lambda = 0.1$ for all the considered the data sets, achieving satisfactory results in all cases.

IV. Conclusions and Future Research Lines

This paper has proposed a new anomaly detection method based on LRASR. To estimate the background, each pixel is represented via a linear combination of the background dictionary's atoms. The representation coefficient matrix, which contains the background information, has a low-rank property. A sparse constraint is added to achieve a more accurate description of the local structure of each sample. As the dictionary represents the background information, a novel way to construct the dictionary is proposed. By this way, the atoms of the dictionary are more likely to belong to the background, and the dictionary covers all the ground material classes in the scene. The anomalies are calculated from the residual of the LRASR. It is demonstrated that the proposed LRASR provides better detection performance than other methods. An important aspect deserving future research is the computational complexity of the proposed method. In this regard, we are currently developing efficient implementations using high performance computing architectures such as commodity graphics processing units.

References

[1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing—Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2008.

[2] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[3] D. W. J. Stein *et al.*, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

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

[5] L. Zhang, B. Du, and Y. Zhong, "Hybrid detectors based on selective endmembers," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2633–2646, Jun. 2010.

[6] 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.

[7] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," in *Proc. SPIE*, vol. 6966, 2008, pp. 1–12.

[8] 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. 1–16, 2010.

[9] Q. Guo *et al.*, "Weighted-RXD and Linear Filter-Based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.

[10] 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.

[11] H. Goldberg, H. Kwon, and N. M. Nasrabadi, "Kernel eigenspace separation transform for subspace anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 581–585, Oct. 2007.

[12] A. Bannerjee, 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.

[13] W. Sakla, A. Chan, J. Ji, and A. Sakla, "An SVDD-based algorithm for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 384–388, Mar. 2011.

[14] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.

[15] 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.

[16] M. D. Farrell and R. M. Mersereau, "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 192–195, Apr. 2005.

[17] K. I. Ranney and M. Soumekh, "Hyperspectral anomaly detection within the signal subspace," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 312–316, Jul. 2006.

[18] W. Liu and C. I. Chang, "Multiple-window anomaly detection for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 6, no. 2, pp. 644–658, Apr. 2013.

[19] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 676–680, Jul. 2011.

[20] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," in *Proc. IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 1233–1236.

[21] 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.

[22] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.

[23] G. Liu *et al.*, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[24] W. Sun, C. Liu, and J. Li, "Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 8, no. 1, May 2014, Art. ID 083641.

[25] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–39, May 2011.

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

[27] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[28] J. Wright *et al.*, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[29] 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, UIUC Tech. Rep. UILU-ENG-09-2215, 2009.

[30] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low rank representation," *Adv. Neural Inf. Process. Syst.*, pp. 612–620, 2011.

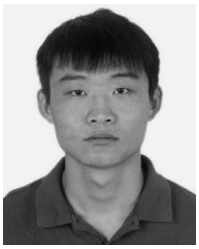
[31] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[32] M. S. Stefanou and J. P. Kerekes, "A method for assessing spectral image utility," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1698–1706, Jun. 2009.

[33] 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.

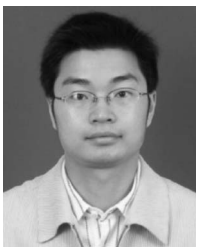
[34] B. Du and L. Zhang, "Unsupervised transfer learning for target detection from hyperspectral images," *Neurocomputing*, vol. 120, pp. 72–82, Nov. 2013.

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



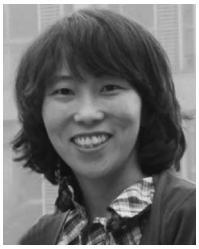
Yang Xu (S'14) was born in Jiangsu, China, in 1990. He received the B.Sc. degree in applied mathematics from the Nanjing University of Science and Technology, Nanjing, China, in 2011, where he is currently working toward the Ph.D. degree in computer science.

His research interests are in the area of hyperspectral image classification, image processing, and machine learning.



Zebin Wu (M'13) was born in Zhejiang, China, in 1981. He received the B.Sc. and Ph.D. degrees in computer science from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2003 and 2007, respectively.

He is currently a Visiting Scholar at the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Caceres, Spain, and also an Associate Professor at the School of Computer Science and Engineering of NUST. His research interests include hyperspectral image processing, high performance computing, and computer simulation.



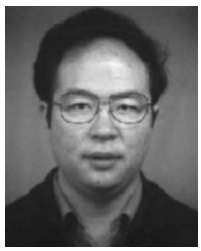
Jun Li (M'13) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotécnica e de Computadores and the Instituto de Telecomunicações, IST, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing (SIGNAL). She has also been actively involved in the Hyperspectral Imaging Network, a Marie Curie Research Training Network, involving 15 partners in 12 countries and intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Since 2011, she has been a Postdoctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Caceres, Spain. Currently, she is a Professor with Sun Yat-Sen University, Guangzhou, China. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li has been a Reviewer for several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Pattern Recognition*, *Optical Engineering*, *Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*. She received the 2012 Best Reviewer Award of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Antonio Plaza (M'05–SM'07–F'15) received the bachelor, M.Sc., and Ph.D. degrees from the University of Extremadura, Extremadura, Spain, in 1997, 1999, and 2002, respectively, all in computer engineering. He was born in Caceres, Spain, in 1975. He is an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, Caceres, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp). His main research interests comprise remotely sensed hyperspectral image analysis and efficient implementations of large-scale scientific problems on high performance computing architectures. He has been the advisor of 12 Ph.D. dissertations and more than 30 M.Sc. dissertations. He was the Coordinator of the Hyperspectral Imaging Network, a European project with a total funding of 2.8 million euro. He authored more than 400 publications, including 130 Journal Citation Reports journal papers (82 in IEEE journals), 20 book chapters, and over 240 peer-reviewed conference proceeding papers (94 in IEEE conferences). He has edited a book on high-performance computing in remote sensing for CRC Press/Taylor and Francis (the first book on this topic in the published literature) and guest edited eight special issues on hyperspectral remote sensing for different journals.

Dr. Plaza is a Fellow of IEEE "for contributions to hyperspectral data processing and parallel computing of Earth observation data." He is a recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009) and a recipient of the recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), a journal for which he served as Associate Editor in 2007–2012. He is also an Associate Editor for IEEE Access and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (2013). He was also a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is currently an Associate Editor for the *Journal of Real-Time Image Processing*. He is a recipient of the 2013 Best Paper Award of the JSTARS journal and a recipient of the most highly cited paper (2005–2010) in the Journal of Parallel and Distributed Computing. He is a coauthor of the 2011 Best Student Paper at the IEEE International Conference on Space Technology and a recipient of the 2008 Best Paper Award at the IEEE Symposium on Signal Processing and Information Technology. He is a recipient of the Best Ph.D. Dissertation Award at the University of Extremadura in 2002. Four of his students have received the Best Ph.D. Dissertation Award at the University of Extremadura, and one of his students has received the Best Ph.D. Dissertation Award at the Complutense University of Madrid, Madrid, Spain. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012 and is currently serving as the President of the Spanish Chapter of IEEE GRSS (since November 2012). He has served as a Proposal Evaluator for the European Commission (Marie Curie Actions, Engineering Panel), the European Space Agency, the Belgium Science Policy, the Israel Science Foundation, and the Spanish Ministry of Science and Innovation. He has participated in the Tenure Track Selection Committee of different universities in Italy, Spain, and Australia. He has reviewed more than 500 manuscripts for over 50 different journals. He is also currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal.



Zhihui Wei was born in Jiangsu, China, in 1963. He received the B.Sc. and M.Sc. degrees in applied mathematics and the Ph.D. degree in communication and information system from South East University, Nanjing, China, in 1983, 1986, and 2003, respectively.

He is currently a Professor and Doctoral Supervisor with the Nanjing University of Science and Technology, Nanjing. His research interests include partial differential equations, mathematical image processing, multiscale analysis, sparse representation, and compressive sensing.