Hyperspectral Anomaly Detection by the Use of Background Joint Sparse Representation

Jiayi Li, *Student Member, IEEE*, Hongyan Zhang, *Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, and Li Ma, *Member, IEEE*

***Abstract*—In this paper, we propose a hyperspectral image anomaly detection model by the use of background joint sparse representation (BJSR). With a practical binary hypothesis test model, the proposed approach consists of the following steps. The adaptive orthogonal background complementary subspace is first estimated by the BJSR, which adaptively selects the most representative background bases for the local region. An unsupervised adaptive subspace detection method is then proposed to suppress the background and simultaneously highlight the anomaly component. The experimental results confirm that the proposed algorithm obtains a desirable detection performance and outperforms the classical RX-based anomaly detectors and the orthogonal subspace projection-based detectors.**

***Index Terms*—Anomaly detection (AD), hyperspectral imagery, joint sparse representation (JSR), robust background estimation.**

I. Introduction

HYPERSPECTRAL images (HSIs) [1], [2] span the visible, near-infrared, and mid-infrared portions of the spectrum (0.4–2.5 μm) in many contiguous and very narrow spectral bands (normally about 0.010 μm wide per band). The significant information about the spectral characteristics of the materials in the hyperspectral scene can be potentially used to exploit and discriminate different objects on the basis of their unique spectral signatures. This discriminative capability has led to two major applications: 1) classification [3], [4]; and 2) target detection [5]–[8]. Anomaly detection (AD) [9], [10] is a special case of the latter application, in which no *a priori* information about the spectra of the targets of interest is available [9]. In the literature, the most well-known benchmark AD method is the Reed–Xiaoli (RX) [11]–[13] algorithm, which assumes a multivariate Gaussian distribution for modeling an optical image, and is a constant false alarm rate (CFAR) [10] adaptive method [11]. As a promising technique, AD has been widely used in many applications, such as detecting crop stress locations in precision agriculture, rare minerals in geology, oil pollution in environmental research, landmines in the public safety and defense domain, and man-made objects in reconnaissance and surveillance applications [14].

Considering the complicated remote sensing background scene, the large number of possible objects of interest, and the uncertainty, AD still faces some obstacles. First, the way of defining a target of interest is ambiguous, as some pixels detected as a target in a local region will belong to the background in a global view. Second, the burdensome construction of an accurate model of the complicated background exacerbates the difficulty of the extraction of anomalies, as the background model cannot be precisely obtained by single- or multiple-Gaussian models. Moreover, with the limitation of the spatial resolution, subpixel objects are a fundamental challenge for AD in HSIs. In recent years, various approaches have been developed to deal with these problems. Carlotto [15] utilized a Gaussian mixture model (GMM) [10] and clustered all the pixels into several subspaces to detect man-made objects and changes in the scene. Developed from the linear mixing model (LMM) [16]–[18], an orthogonal subspace projection (OSP) method has also been introduced in an unsupervised manner [17], such that the background components can be suppressed via OSP to improve the AD performance. Moreover, some nonlinear detectors, such as kernel RX [18], support vector data description (SVDD) [19], and so on, have also been proposed to handle the nonlinear separability of hyperspectral data.

Sparsity of signals, an emerging and extremely powerful tool in many classical signal processing applications [20]–[22], denotes that most natural signals can be compactly represented by only a few coefficients in a certain basis or dictionary, with almost no performance loss [23]–[25]. With the development of compressed sensing (CS) [26] and sparse coding, this approach has been applied in many different computer vision and pattern recognition applications [7], [27]–[30]. With respect to some bases with meaningful semantic information, the similar high-dimensional natural signals can be assigned to a low-dimensional subspace, which can nonlinearly extract the desirable sematic information for the subsequent decision and analysis. The use of sparsity as a constraint often leads to state-of-the-art performance, since the CS theory ensures excellent recovery of the sparse signal. For hyperspectral target detection, Chen *et al.* [7] constructed two subdictionaries from the background and the target prior, and set the target detection problem as a special binary sparse representation classification (SRC) [27] task, under the assumption that the background and target should be distributed in two different subspaces.

Manuscript received February 17, 2014; revised May 06, 2015; accepted May 20, 2015. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707105, in part by the National Natural Science Foundation of China under Grant 61201342, Grant 41431175, and Grant 61102104, and in part by Open Research Fund of Key Laboratory of Digital Earth, Chinese Academy of Sciences.

J. Li, H. Zhang, and L. Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, and the Collaborative Innovation Center for Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zhanghongyan@whu.edu.cn).

L. Ma is with the Faculty of Mechanical and Electronic Information Engineering, China University of Geosciences, Wuhan 430079, China.

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/JSTARS.2015.2437073

Our major concern in this paper is the hyperspectral AD problem, and we propose a relaxed binary hypothesis model with the natural highly redundant prior of the background and the low-rank characteristic of the HSI. Considering the complicated land-cover distribution of HSIs, the statistical assumption is often ideal and the estimation of the background is hard to determine [16]. In the proposed method, the pixels in a local neighborhood with similar spectra are assumed to belong to limited subspaces, which belong to the dictionary constructed under the highly redundant background prior. If the center test pixel is anomalous, the active dictionary bases used to simultaneously represent the surrounding background pixels should be dissimilar to the center anomaly. On the other hand, when the center pixel is a normal one, it can be predicted by the aforementioned active bases. By comparing the reconstruction results, an unsupervised adaptive subspace detector is proposed to check whether there is any salient difference and judge whether the center pixel is an anomaly or not. The robust background estimation is effectively implemented by the use of the simultaneous sparsity constraint, and ensured by the low-rank prior of the local neighborhood, as theoretically analyzed in [31]. In view of this, the center test pixel can be linearly represented by the selected background subset, if and only if a nonignorable component of the center test pixel also belongs to the subspace spanned by the selected background subset. Based on the robust background estimation property, the proposed local AD method is named "background joint sparse representation detection" (BJSRD).

We next introduce the relationship between the proposed method and the other hyperspectral anomaly detectors. The overcomplete background dictionary in the proposed algorithm contains multiple background subspaces, which flexibly select the underlying bases for the local neighboring pixel set, and can effectively and simultaneously detect multiclass anomalies in a complicated background. The cluster-based anomaly detector (CBAD) [15], which utilizes a GMM to characterize the spectral pixels from nonhomogeneous and multicomponent scenes, is strictly dependent on the previous clustering procedure, and is limited in real time. The OSP-based algorithms [20], [34] often roughly divide the data into one background subspace and one target subspace, and cannot adequately analyze the patterns of each kind of land cover. Moreover, unlike the other detectors based on statistical hypothesis testing, the sparsity model in the proposed approach has the flexibility of modeling the complicated characteristics of the background of the HSI.

Compared with the sparsity model-based hyperspectral target detection method in [7], significant differences can be found as follows. First, the algorithm in [7] employs supervised target detection, a binary classification case based on the collaborative representation (CR) mechanism [32] of the well-known sparse representation classifier in [27]. However, the proposed method refers to unsupervised AD, which is based on the high-background redundancy of the HSI scene. The second difference is the dictionary learning stage. In fact, the BJSR step of the proposed AD method is equivalent to a dictionary learning approach which aims at pruning the original dictionary and adaptively obtaining the distinct background one, with an adaptive noise baseline. The target detection approach in [7] also uses the classical K-SVD [33] method to learn the dictionary, due to the computational efficiency.

This paper is organized as follows. Section II briefly reviews sparse representation (SR) and joint sparse representation (JSR) for HSIs. Section III proposes the BJSRD method for hyperspectral imagery. The experimental results of the proposed algorithm are given in Section 4. Finally, Section 5 concludes the paper.

II. Brief Review of Related Works

A. SR

In a sparsity model, it is assumed that a signal can be approximated by a sparse linear combination of elements from a basis set or dictionary. A hyperspectral signal can be denoted as $s \in \mathbb{R}^B$, where B is the number of bands. Given a matrix $\mathbf{A} \in \mathbb{R}^{B\times N}$ with $B \ll N$ as an overcomplete dictionary, it is believed that the signal s can be approximately represented by multiplying the dictionary $\mathbf{A}$ with a sparse vector $\boldsymbol{\alpha}$, in which only a few entries are nonzero. The sparse vector $\boldsymbol{\alpha}$ can be obtained by solving the following optimization problem:

$$\boldsymbol{\alpha} = \arg\min \|\mathbf{A}\boldsymbol{\alpha} - s\|_2 \quad \text{s.t. } \|\boldsymbol{\alpha}\|_0 \le K_0 \tag{1}$$

where K_0 is the given upper bound of the sparsity level. It is clear that these aforementioned problems are NP-hard. In general, there are two effective ways of solving these problems: the greedy pursuit-based algorithms [34], [35] and the ℓ_1-norm convex relaxation algorithms [36]–[38].

B. JSR

For a hyperspectral scene, the majority of the neighboring pixels in a local patch usually consist of similar materials. All the similar pixels can be jointly represented in the same low-dimensional feature subspace with different compact coefficients [39]. Tropp *et al.* [35] noted that a simultaneous sparse approximation problem requires a good approximation of several input pixels at once, using different linear combinations of the same elementary signals. This kind of approximation problem can be solved by the joint sparsity model (JSM) [40].

For a hyperspectral patch $\mathbf{S} = [s_1\ s_2 \dots s_n] \in \mathbb{R}^{B\times n}$, it can be reconstructed under the JSM with the $B \times N$ structured dictionary $\mathbf{A}$

$$\begin{aligned}\mathbf{S} &= [s_1\ s_2 \dots s_n] = [\mathbf{A}\boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_1\ \mathbf{A}\boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}_2 \dots \mathbf{A}\boldsymbol{\alpha}_n + \boldsymbol{\varepsilon}_n] \\ &= \mathbf{A}\underbrace{[\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2 \dots \boldsymbol{\alpha}_n]}_{\boldsymbol{\Psi}} + \boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Psi} + \boldsymbol{\Sigma}\end{aligned} \tag{2}$$

where $\boldsymbol{\Psi}$ is the set of all the sparse coefficient vectors $\{\boldsymbol{\alpha}_i\}_{i=1,\dots,n}$, n is the number of pixels in the patch, and $\boldsymbol{\Sigma}$ is the model error matrix. In the JSM, it is assumed that all the columns $\{s_i\}_{i=1,\dots,n}$ share a common sparsity pattern, which means that these signals are located in the same low-dimensional subspace of the dictionary. It is believed that the low-dimensional subspace spanned by the active atoms of $\mathbf{A}$ is the most informative and distinct base for $\mathbf{S}$, and $\boldsymbol{\Sigma}$ physically

represents the noise level of the patch. The optimization model corresponding to (2) can be expressed as

$$\mathbf{\Psi} = \arg\min \|\mathbf{A}\mathbf{\Psi} - \mathbf{S}\|_F, \quad \text{s.t. } \|\mathbf{\Psi}\|_{row,0} \leq K_0 \tag{3}$$

where $\|\bullet\|_{row,0}$ denotes the number of nonzero rows of $\mathbf{\Psi}$. The optimal solution $\mathbf{\Psi}$ is a row-sparse matrix with L nonzero rows. For the different solutions to the optimization model (3), please refer to [35], [41].

C. RX AD

Based on the assumption that HSI data can be represented by a multivariate normal (Gaussian) distribution, Reed and Yu [11] proposed the well-known RX anomaly detector. Denoting the pixel after subtracting the mean value from each one as s_c, the two competing hypotheses of the RX algorithm are

$$\begin{aligned} H_0 &: s_c = e, && \text{Anomaly absent} \\ H_1 &: s_c = at + e, && \text{Anomaly present} \end{aligned} \tag{4}$$

where $a = 0$ under H_0, an d $a > 0$ under H_1. e represents the background clutter noise process, and $\mathbf{t}$ is the spectral signature of the anomaly target. Both the target signature $\mathbf{t}$ and background covariance $\mathbf{C}_b$ are unknown, and the model of (4) assumes that the hyperspectral data arise from two normal probability density functions with the same covariance matrix but different means. In general, a more appropriate model should have different covariance structures, considering the multiple kinds of background and the complicated anomaly spectra. For this classical anomaly detector, the RX algorithm converges to a simple version

$$\mathrm{RX}(s_c) = (s_c - \widehat{\boldsymbol{\mu}})' \widehat{\mathbf{R}}^{-1} (s_c - \widehat{\boldsymbol{\mu}}) \underset{<H_0}{>H_1} \eta \tag{5}$$

where η is the threshold of the test, $\widehat{\mathbf{R}}$ is the background covariance matrix estimated from the reference background clutter data, and $\widehat{\boldsymbol{\mu}}$ is the estimated background sample mean.

III. Proposed Framework

In this section, we first review the spatial window setting for the local AD method [10], and we then describe the pixel detection processing in detail.

A. Dual Spatial Window Setting

For the dual windows shown in Fig. 1, the test pixel is colored in red, and the guard window sized g^2 is colored in light orange. The dual windows are effective in preventing some potential target pixels becoming mixed with the outer neighboring pixel set. On the outside of the guard window, we set a local neighboring window (colored in blue) with $n = b^2 - g^2$ pixels. The possible cases for this pixel set can be categorized as: 1) all the pixels in this local neighboring pixel set are background, and they consist of a few types of land cover; 2) the majority of the pixels, except for one or two anomalous pixels, are background; and 3) the majority or all the pixels are anomalies, which is believed to never happen in reality, due to the low probability of anomaly pixels in the hyperspectral scene. Therefore, the proposed AD method, which utilizes the local dual window set, focuses on the first two cases.

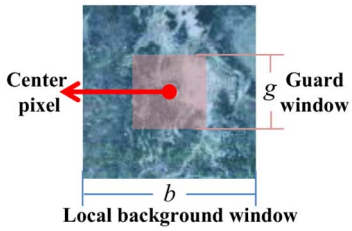


Fig. 1. Dual windows used in the proposed method.

B. Proposed AD Method for the Center Test Pixel

For the test pixel, we first construct the spatial window, as described in Section III-A, and we then set the center pixel as $s_c \in \mathbb{R}^B$ and the local neighboring pixel set as $\mathbf{S} = [s_1\ s_2 \ldots s_n] \in \mathbb{R}^{B \times n}$.

For the proposed AD method, the hypothesis test is

$$\begin{aligned} H_0 &: s_c = \mathbf{B}\boldsymbol{\alpha}_b + e, && \text{Anomaly absent} \\ H_1 &: s_c = \mathbf{B}\boldsymbol{\alpha}_b + t + e, && \text{Anomaly present} \end{aligned} \tag{6}$$

where $\mathbf{B}$ is the background basis matrix to s_c, $\boldsymbol{\alpha}_b$ is the associated background coefficient vector, $\boldsymbol{t}$ is the unknown anomaly component of s_c, and $\boldsymbol{e}$ is the random noise with low energy, which is complicated in real HSIs. When the H_1 case holds, it is believed that there is some unignorable anomaly component $\boldsymbol{t}$ in s_c. In this way, the H_1 hypothesis can be rewritten as $s_c = \mathbf{B}\boldsymbol{\alpha}_b + \boldsymbol{m}$, where $\boldsymbol{m}$ is mixed with the noise and anomaly component, which occupies an unignorable energy part of the test pixel. In view of this, the main issues under the hypothesis test are two aspects: 1) background basis matrix $\mathbf{B}$ construction with the unsupervised information; and 2) the evaluation criteria for the judgment of the aforementioned "unignorable" contribution. In the proposed method, the first term is handled by the BJSR process, and the latter problem is addressed by a combination of the following procedures (Fig. 2).

1) Robust BJSR: Background estimation and suppression techniques [3], [42]–[44] have been studied for a long time, but they are often sensitive to the following judgment procedures. To judge whether the center pixel is abnormal or not, the most straightforward idea is to predict it with its surrounding pixels in the outer window, and a successful prediction can be expected when only the background information contributes to the prediction of the center pixel. However, as the local neighboring pixel set $\mathbf{S}$ may be mixed with several potential abnormal signatures, such decision-making hardly suggests a desirable AD performance. In view of this, we propose an adaptive background basis matrix construction method by the use of the BJSR, which adaptively estimates the local background information of the center pixel.

We first construct the overcomplete dictionary from the hyperspectral scene for the BJSR to linearly represent the local neighboring pixel set $\mathbf{S} = [s_1\ s_2 \ldots s_n] \in \mathbb{R}^{B \times n}$. We stack all

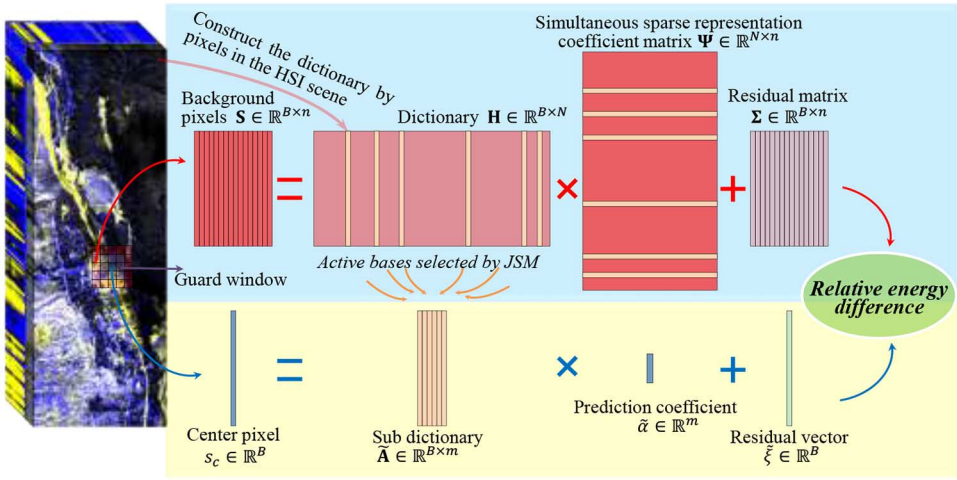


Fig. 2. Schematic illustration of BJSR for hyperspectral AD. For a pixel to be tested, we first construct the joint signal matrix from pixels surrounding the guard window, which is used to prevent the potential target pixels from mixing in this local background characterization. The local neighboring pixel set $\mathbf{S}$ contains n columns, where n is the size of $\mathbf{S}$. All the atoms of the overcomplete dictionary $\mathbf{H}$ for the JSR are the raw hyperspectral pixels in the scene. First, the matrix $\mathbf{S}$ is sparsely reconstructed by the dictionary $\mathbf{H}$ with an acceptable tolerance error matrix $\mathbf{\Sigma}$. We then select the atoms acting as active bases in the aforementioned SR and consider them as the most informative spectral components of this local neighboring pixel set. Utilizing these active components to linearly represent the center test pixel, we obtain the prediction result with a residual vector ξ. We compare these two residual energies and determine whether the center test area contains any anomaly component or not.

the pixels in the scene as a matrix $\mathbf{H} \in \mathbb{R}^{B\times H}$, in which H is the number of pixels, and B refers to the number of spectral bands. Assuming that there are C_{bn} categories of background materials and C_{tn} categories of anomaly targets, the columns of the matrix $\mathbf{H}$ are reordered by category, without loss of generality. In view of this, the matrix can be shown as $\mathbf{H} = [\mathbf{B}_{C_{b1}} \dots \mathbf{B}_{C_{bn}} \mathbf{T}_{C_{t1}} \dots \mathbf{T}_{C_{tn}}]$, where $\mathbf{B}_{C_{bi}}$ contains N_{bi} background pixels belonging to category C_{bi} $(bi = b1, \dots, bn)$, $\mathbf{T}_{C_{ti}}$ contains N_{ti} target pixels belonging to category C_{ti} $(ti = t1, \dots, tn)$, and $H = \sum_{bi=b1}^{bn} N_{bi} + \sum_{ti=t1}^{tn} N_{ti}$. That is to say, there are $C_{bn} + C_{tn}$ low-dimensional subspaces associated with the respective land-cover types in the hyper-dimensional space spanned by $\mathbf{H}$.

Some representative cases of the background information from the local scenario $\mathbf{S}$ are analyzed and shown in Fig. 3. For simplicity, the noise components of $\mathbf{S}$ [as in the instance of Fig. 3(d)] are assumed to be subtracted in advance, and then the detailed distribution of each class cluster in each case is shown in the associated subfigure.

- Case 1: Only background pixels in $\mathbf{S}$, as in Fig. 3(b). Selection of the representative spectral components of $\mathbf{S}$ can reach the desired goal for background estimation.
- Case 2: A nonnegligible anomalous cluster in the local scenario $\mathbf{S}$, as shown in Fig. 3(c). Considering the very small size of the anomaly panel, this case can never happen in reality.
- Case 3: Several anomaly pixels implanted in the surrounding background cluster, as shown in Fig. 3(e). In this case, reducing the influence of these unexpected anomalies is required for an acceptable background estimation step.

To sum up the potential cases of $\mathbf{S}$, it is suggested that a robust background estimation calls for dealing with the

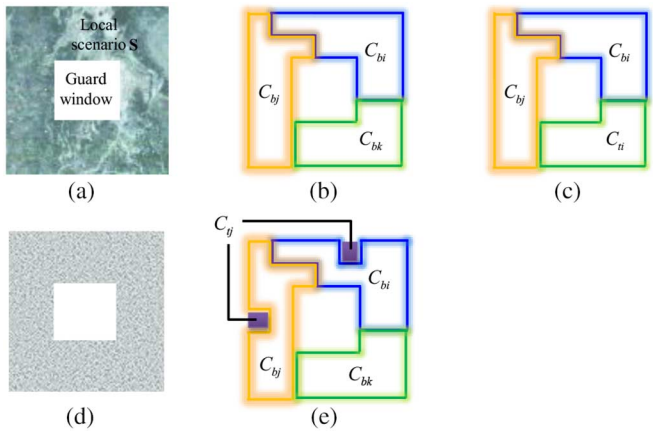


Fig. 3. Illustration of different cases of the local neighboring pixel set $\mathbf{S}$. (a) False-color map of $\mathbf{S}$. (b) Only background land cover constitutes $\mathbf{S}$. (c) Background land cover and a nonnegligible anomalous cluster in $\mathbf{S}$. (d) Random noise components of the local scenario. (e) Similar composition to (b), except for a few anomaly pixels implanted in certain background spots.

cases in which there are some existing anomaly pixels in the surrounding neighborhood. We therefore introduce a robust background estimation method based on a sparse constraint. Suppose a column of $\mathbf{S}$ named s_j $(j = 1, \dots, n)$ belongs to class C_{bi}, then it can be compactly represented by a linear combination of the pixels of $\mathbf{B}_{C_{bi}}$

$$s_j = \mathbf{B}_{C_{bi}} \boldsymbol{\beta}_{C_{bi},j} + \boldsymbol{\xi}_j \tag{7}$$

where $\boldsymbol{\beta}_{C_{bi},j}$ is an unknown N_{bi}-dimensional coefficient vector, and $\boldsymbol{\xi}_j$ is the random noise. Since there is no prior for judging the label of the pixel s_j as C_{bi} to remove the anomaly target subdictionaries from the matrix $\mathbf{H}$, we utilize $\mathbf{H}$ as the overcomplete dictionary to solve the sparse coefficient vector

TABLE I
PROPOSED BJSRD ALGORITHM FOR HSI

Input: A hyperspectral image and parameters: 1) the two sizes for the dual window (b and c) ; 2) the size of the search window; and 3) the parameters of SOMP
Initialization: Linearly normalize each band of the HSI
For each test pixel $\boldsymbol{s}_c$ in the hyperspectral scene:
1) Construct the neighboring pixel set $\mathbf{S}$ and the overcomplete dictionary $\mathbf{H}$
2) Select the $\tilde{\mathbf{A}}$ from $\mathbf{H}$ via SOMP
3) Calculate $\mathbf{P}_B^{\perp} = \mathbf{I} - \tilde{\mathbf{A}}\left(\left(\tilde{\mathbf{A}}\right)^T \tilde{\mathbf{A}}\right)^{-1}\left(\tilde{\mathbf{A}}\right)^T$
4) Calculate the detected value of $\boldsymbol{s}_c$ via (12)
End For
Output: A 2-D matrix which records the detection values of the hyperspectral scene

$\boldsymbol{\beta}_j$, with respect to the dictionary, and the optimization problem can be given by

$$\boldsymbol{\beta}_j = \arg\min \|\mathbf{H}\boldsymbol{\beta}_j - s_j\|_2, \quad \text{s.t. } \|\boldsymbol{\beta}_j\|_0 \le l \tag{8}$$

where l is the predefined sparsity level.

To avoid a trivial solution to (8), all the pixels in the local window sized $b \times b$ are removed from $\mathbf{H}$. In addition, utilizing pixels in a local search window instead of the whole scene is more suitable in practice, as the local similarity suggests more stability in the joint representation step, as analyzed in detail in [32]. In the following part, we still take "$\mathbf{H}$" as the original dictionary.

Similar to the pixel s_j, all the other pixels in the local neighboring window can also be sparsely represented by the overcomplete dictionary $\mathbf{H}$, shown as $\mathbf{S} = [s_1 \dots s_n] = \mathbf{H}[\boldsymbol{\alpha}_1 \dots \boldsymbol{\alpha}_n] + [\xi_1 \dots \xi_n]$. Since the local neighboring window contains only one or a few kinds of background categories, it is assumed that all the pixels in the local neighboring window can be linearly represented in the same low-dimensional feature subspace with different compact coefficients, where this subspace can be associated with a background subdictionary $\tilde{\mathbf{B}}_{C_{bi}}$ or a combination of a few background subdictionaries. In this way, the local neighboring pixel set $\mathbf{S}$ can be better represented under the JSM

$$\boldsymbol{\Psi} = \arg\min \|\mathbf{H}\boldsymbol{\Psi} - \mathbf{S}\|_F, \quad \text{s.t. } \|\boldsymbol{\Psi}\|_{row,0} \le L \tag{9}$$

where $\boldsymbol{\Psi}$ is the set of all the sparse coefficient vectors $\{\tilde{\boldsymbol{\alpha}}_i\}_{i=1,\dots,n}$, and L denotes the upper bound of the number of nonzero rows of $\boldsymbol{\Psi}$. In this paper, we utilize simultaneous orthogonal matching pursuit (SOMP) [35] to obtain the row-sparse coefficient matrix $\boldsymbol{\Psi}$, the residual matrix $\boldsymbol{\Sigma}$, as well as the active atom subset $\tilde{\mathbf{A}}$ of $\mathbf{H}$, as shown in Table I. Considering the low probability of target pixels, it is mentioned again that the principal component of $\mathbf{S}$ must surely belong to the background. Based on the JSM, it is natural that the atoms of $\tilde{\mathbf{A}}$ effectively reflect the characteristics of the feature space associated with the principal background components of $\mathbf{S}$, which are robust in dealing with the case where $\mathbf{S}$ includes only one or two anomaly pixels, and can elegantly handle the cases in Fig. 3(b), (e), and (f). That is to say, the subset $\tilde{\mathbf{A}}$, which filters out the bases associated with the underlying anomaly subspaces and the irrelevant background subspaces from the original overcomplete dictionary $\mathbf{H}$, is the adaptive background basis set to the central test pixel s_c, and is equivalent to $\mathbf{B}$ in (6).

By the use of SOMP, the characteristics of $\boldsymbol{\Sigma}$ can be denoted as

$$\begin{aligned}\boldsymbol{\Sigma} &= \mathbf{S} - \tilde{\mathbf{A}}\boldsymbol{\Psi} = \mathbf{S} - \tilde{\mathbf{A}}\left(\left(\tilde{\mathbf{A}}\right)^T \tilde{\mathbf{A}}\right)^{-1}\left(\tilde{\mathbf{A}}\right)^T \mathbf{S} \\ &= \left(\mathbf{I} - \tilde{\mathbf{A}}\left(\left(\tilde{\mathbf{A}}\right)^T \tilde{\mathbf{A}}\right)^{-1}\left(\tilde{\mathbf{A}}\right)^T\right)\mathbf{S} = \mathbf{P}_B^{\perp}\mathbf{S}\end{aligned} \tag{10}$$

where $\mathbf{P}_B^{\perp}$ represents the background orthogonal complementary subspace. As the majority of $\mathbf{S}$ is background land cover, the average energy of $\boldsymbol{\Sigma}$ reflects the noise level of this neighborhood, and the contribution of the potential anomaly component in the neighborhood can be smoothed.

To select the active set from $\mathbf{H}$, the stopping criterion of SOMP contains three parameters. First, the upper bound of the joint sparsity level L is related to the intrinsic dimension of the hyperspectral scene, which is low rank (as discussed in [45]). For the background estimation, it is believed that L plays the dominant role, and the final residual threshold should be robust over a wide reasonable range. In addition, it should be mentioned that the residual is proportional to the noise of the scene, and the reference scopes of these parameters may require some adjustment, according to the specific quality and the content of the hyperspectral image. The detailed characteristics of the parameters, as illustrated by the experiments, are described in Section III-B2.

2) Anomaly Detector Construction: We next analyze the characteristics of the center pixel s_c, based on the active dictionary atoms selected by the BJSR in the last step. Induced from the orthogonality utilized in each iteration of SOMP, each atom of $\tilde{\mathbf{A}}$ can be considered as the most representative basis of the local neighboring pixel set. In this way, the linear projection (LP) by the use of $\tilde{\mathbf{A}}$ can be shown as

$$s_c = \tilde{\mathbf{A}}\tilde{\boldsymbol{\alpha}} + \boldsymbol{\xi}_c \tag{11}$$

where $\tilde{\boldsymbol{\alpha}}$ denotes the fractional abundance of the most representative spectral component, and $\boldsymbol{\xi}_c$ denotes the tolerance error. For simplicity, the unconstrained LP is utilized as (10). Since $\tilde{\mathbf{A}}$ is an adaptive subdictionary selected from $\mathbf{H}$, such an unconstrained LP approach should induce a sparse and superior solution, as suggested in [32], [46]. To minimize the unconstrained LP error, it can be determined using a least-squares solution $\xi_c = s_c - \tilde{\mathbf{A}}\tilde{\boldsymbol{\alpha}} = \mathbf{P}_B^{\perp} s_c$.

For the proposed AD algorithm, the different cases can be analyzed as follows:

Case 1: s_c contains no anomalies, and the H_0 hypothesis holds. Considering the spatial correlation of s_c with its surrounding neighboring background, s_c can be successfully reconstructed by (11), and $\boldsymbol{\xi}_c$ denotes the noise, which is similar to the atoms of $\boldsymbol{\Sigma}$. That is to say, the energy of $\boldsymbol{\xi}_c$, represented as $\|\xi_c\|_2$, should be comparable to the average energy of $\boldsymbol{\Sigma}$.

Case 2: s_c is a pure anomaly pixel to its surrounding background. In this case, s_c can be modeled as

$s_c = \mathbf{t} + \mathbf{n} = \xi_c$, where $\boldsymbol{\xi}_c$ should maintain the principal component with respect to the orthogonal subspace of $\tilde{\mathbf{A}}$, and the LP (10) mostly fails. Therefore, $\|\boldsymbol{\xi}_c\|_2$ should be significantly larger than the average energy of $\boldsymbol{\Sigma}$.

Case 3: s_c is a mixed pixel containing some anomalous components, and the H_1 hypothesis holds. In this case, some unignorable anomalous signatures $\boldsymbol{t}$ are mixed in s_c, as shown in (6). For an HSI with a reasonable signal-to-noise ratio (SNR), it is believed that $\boldsymbol{\xi}_c = t + n$ should be dominated by t, and $\|\xi_c\|_2$ should still be in multiples of the base average energy of $\boldsymbol{\Sigma}$.

In view of this, it is reasonable that the detector be modeled as

$$d = \frac{\left\|\mathbf{P}_B^{\perp} s_c\right\|_2^2}{\sum \left\|\mathbf{P}_B^{\perp} s_i\right\|_2^2 / n} = \frac{{s_c}^T \mathbf{P}_B^{\perp} s_c}{\text{mean}\left({s_i}^T \mathbf{P}_B^{\perp} s_i\right)}, \quad i = 1, \ldots, n \tag{12}$$

where s_i is the ith pixel in the dual window. It can be seen that the formulation of the proposed method is a matched filter, and it can be considered as an unsupervised extended version of the adaptive subspace detector [47].

As described above, the proposed AD algorithm is robust in two aspects. First, the only assumption for the AD method refers to the overcompleteness of the dictionary $\mathbf{H}$ and the limitation of the anomalies, no matter what the distribution or size of the anomalous objects and the background land cover are. In view of this, it is worth mentioning that the dictionary $\mathbf{H}$ can also be constructed by a local search window centered at the test pixel, instead of the whole scene, in consideration of the computational cost in practical applications. Although this kind of approach may weaken the prior of the high redundancy of the dictionary $\mathbf{H}$, some potential interfering bases will also be removed in advance, which improves the representation performance. The effectiveness of this technique is validated in the experimental section. Second, the proposed AD method can tolerate some potentially confusing anomaly signatures in the local background $\mathbf{S}$. In the BJSR step, the anomaly part is modeled into the noise, and only the principal component of $\mathbf{S}$, that is the background information, is utilized to automatically prune the rest of the subspace of $\mathbf{H}$ and to maintain the adaptive and pure local background basis matrix $\tilde{\mathbf{A}}$.

C. Final Scheme for BJSRD

The implementation details of the proposed BJSRD algorithm for HSI are shown in Table I.

When utilizing the full scene as the original overcomplete dictionary, the computational burden for each pixel to be tested by the proposed AD algorithm is $O(nLBH + LB^2)$, where the local background representation costs $O(nLBH)$ by SOMP, and the $\mathbf{P}_B^{\perp}$ construction and the final detection cost $O(LB^2)$, as $((\tilde{\mathbf{A}})^T \tilde{\mathbf{A}})^{-1}(\tilde{\mathbf{A}})^T$ can be obtained simultaneously from SOMP. In this way, the computational burden for the whole scene is $O(nLBH^2 + LB^2H)$. In practice, the computational complexity can be reduced to $O((nH + B)LBd^2)$ when utilizing a smaller search window instead of the full scene to construct $\mathbf{H}$, where $d^2 < H$ denotes the size of the search window.

The proposed approach is superior to the well-known local RX (LRX) detector for the following reasons. First, $\mathbf{P}_B^{\perp}$ alleviates the interference from the potential anomaly component in the neighborhood, as the joint sparsity constraint ensures the purity of the selected representative background basis. Second, the computational complexity of the local version of the proposed approach is less than that of LRX, as the covariance matrix inverse costs $O(n^3 + B^3)$ for each pixel.

IV. Experiments

We investigated the effectiveness of the proposed algorithm with two widely used real HSI datasets. The detectors of global RX (GRX), LRX [11], low possibility detection (LPD) [3] based on OSP [18], and a recent low-rank subspace decomposition method for global RX detection (referred to as LRD-GRX) [48] were used as the benchmarks in this study. To obtain a fair comparison, each detection map was linearly normalized by its maximum value in the performance evaluation step. All the methods and experiments were carried out using MATLAB on a PC with a single 3.50-GHz processer and 16.0 GB of RAM.

A. Dataset Description

The first real dataset, obtained from an aircraft platform with a HYperspectral Digital Imagery Collection Experiment (HYDICE) sensor, can be freely downloaded from the website of the Army GeoSpatial Center (www.agc.army.mil/hypercube/). With a spectral resolution of 10 nm and a spatial resolution of 1 m, this dataset contains a vegetation area, a construction area, and several roads, on which there are some vehicles. The full image size is 307×307 pixels, with 210 spectral channels in the VNIR–SWIR range. In this experiment, the low-SNR and water absorption bands were eliminated so that 160 bands remained. A subscene sized 80×100 in the upper right of the whole scene was used in this experiment, as shown in Fig. 4(a). It can be seen that the anomaly panels with different sizes are mainly vehicles located in different backgrounds, the positions of which are shown in Fig. 4(b). With the help of the minimum noise fraction (MNF) and the pixel purity index (PPI) modules in ENVI 4.7, it was revealed that the rest of the pixels in the scene were all composed of the four main background objects: 1) asphalt; 2) grass; 3) trees; and 4) soil, the detailed spectra of which are shown in Fig. 4(c). In addition, the details of the anomaly target spectra are shown in Fig. 4(d).

The second real dataset covers an agricultural area of the State of Indiana, USA, in 2008, and was obtained by the Hyperion sensor onboard NASA's Earth Observing-1 (EO-1) satellite. This dataset has 242 bands with a spectral resolution of 10 nm over 357–2576 nm. After removal of the low-SNR bands and the uncalibrated bands, a 150×150 subimage containing 149 spectral bands was used. The anomalies come from the storage silo and the roof, and are shown in Fig. 5(d). The background object endmembers were extracted, including paddy field, crops, and water, and constituted the majority of the whole

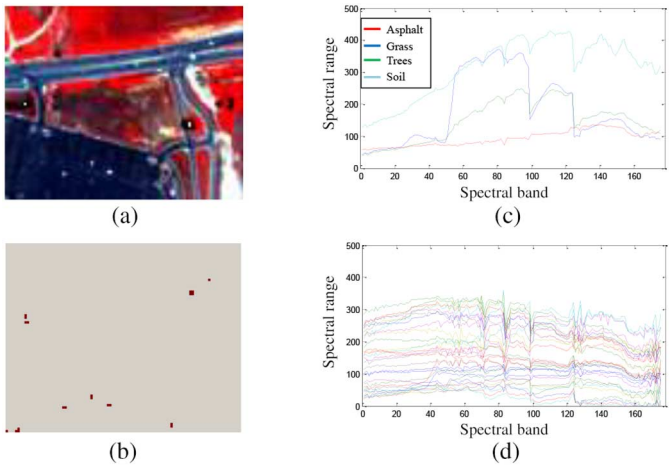


Fig. 4. HYDICE hyperspectral dataset. (a) False-color map. (b) Subscene detection area. (b) Anomaly target positions. (c) Spectra of the background objects. (d) Spectra of all of the anomalies.

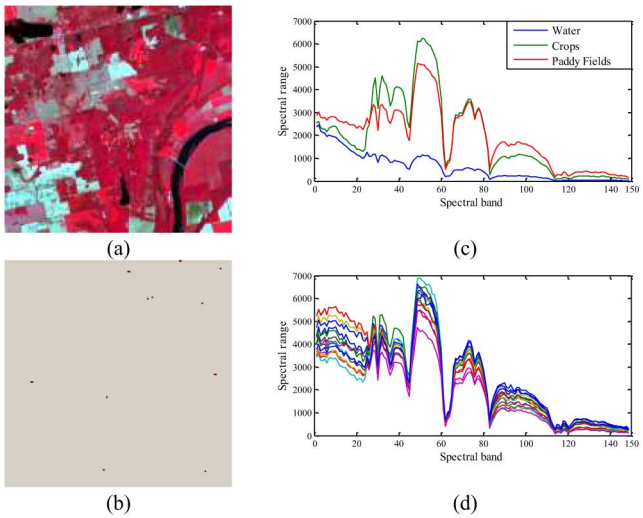


Fig. 5. Hyperion hyperspectral dataset. (a) False-color map. (b) Anomaly target positions. (c) Spectra of the background objects. (d) Spectra of all of the anomalies.

image pixel dataset, as shown in Fig. 5(c). We ensured that all of the pixels spectrally different from these endmembers were selected, and the rest of the pixels were all from the three background endmembers.

B. Detection Performance

For the proposed algorithm, the upper bound of the sparsity level L of the overcomplete background dictionary $\mathbf{H}$ was varied from 5 to 50. For the dual spatial window, the size of the local neighboring window was varied from 3×3 to 17×17, while the search window $d \times d$ was varied from 5×5 to 33×33 for both data sets. The size of the guard window was set from 1×1 to 5×5. In all the experiments, the parameters for each algorithm in each case were set as the optimal, as described in this section, and a detailed analysis of the roles of these parameters is presented in Section IV-C.

TABLE II
AUC VALUES/RUNNING TIMES OF THE ALGORITHMS OBTAINED WITH THE HYPERSPECTRAL DATASETS

AD		GRX	LRX	LPD	LRD-GRX	BJSR D	BJSR D(f)
HYDICE	AUC	0.9856	0.9949	0.9949	0.9910	0.9989	0.9956
	Times	0.89	280.90	0.75	3.25	15.99	1.09E4
Hyperion	AUC	0.9977	0.9996	0.9998	0.9978	0.9999	–
	Times	2.04	115.08	1.72	5.69	16.94	

The quantitative results and the running times of the two hyperspectral image datasets are shown in Table II. The area under the receiver operating characteristic (ROC) curve (AUC) was utilized to evaluate the detection performance for each detector. The AUC metric varies between [0, 1], and a larger value is associated with a better detection result. In Table II, the best results for each quality index are labeled in bold, and the second-best results for each quality index are underlined. The detection results and the binary detection map at 1e-3 FAR for each AD method are shown in Figs. 6 and 7, respectively. In addition, the representative ROC [9] curves under different fraction cases and the statistical separability analysis for these two datasets are shown in Fig. 8, respectively.

AD methods project the vectored pixels into an ideal decision plane, where anomaly pixels should be highlighted and background pixels should be suppressed. For the HYDICE dataset, whose background constitution is more complicated than the Hyperion dataset, the number of subspaces cannot be easily fixed, which causes the inferior result of LPD. The detection performance of LRX was better than that of GRX, since the spatial distribution of the anomalies in this dataset encourages the local window to avoid the interruption from the complex background in the whole scene. For the global-based detectors, it is noted that LRD—RXD is superior to the others, as the low-rank prior can effectively enhance the detector. In the meantime, the global-based detectors mistreat several pixels on the right side and the upper left Y-shaped pathway as anomalies, as shown in Fig. 6(a), (c), and (d). For the relatively simple Hyperion dataset, a similar performance can be seen in Fig. 7(a), (c), and (d), as a vertical strip in the scene is also wrongly highlighted. In this dataset, LPD provides an acceptable detection result, as the parameter of the subspace number can successfully estimate the background in the scene. As can be seen in the binary maps, the global-based detectors cannot effectively consider some of the local environmental variation, as shown in the false alarms located in the right edge of Fig. 6(f) and (i) for the HYDICE dataset and those located in the lower right of Fig. 7(f) and (i) for the Hyperion dataset. In short, when compared with the global-based detectors, the local detectors are better at suppressing the background components, and the proposed algorithm gives the best result. From the quantitative point of view, the ROC curves and AUC values suggest a similar story.

We continue to analyze the separability of the aforementioned AD algorithms in Fig. 8(c) and (d). These subfigures report the output target detection test result range of all the AD methods in the test datasets in five groups. Each group has

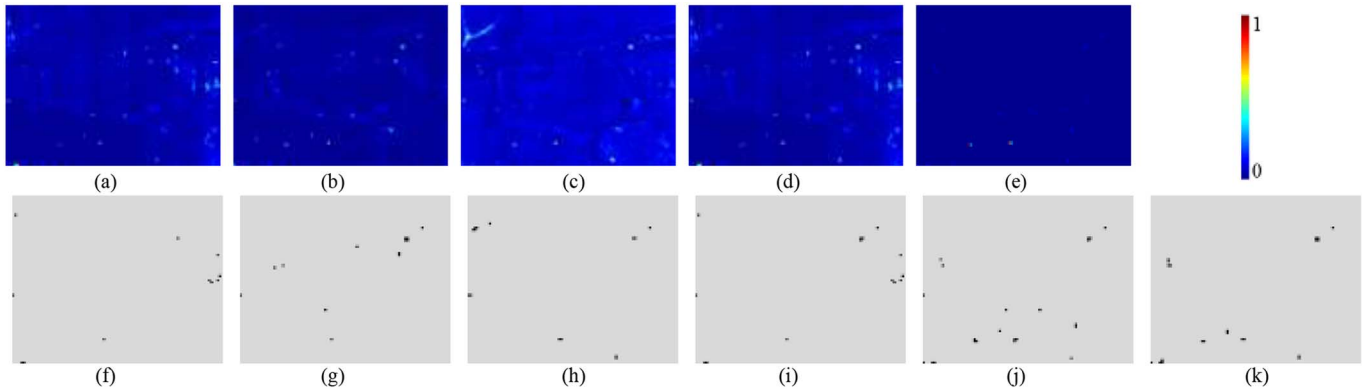


Fig. 6. Detection performances of the AD algorithms for the real HYDICE dataset: (a)–(e) refer to the detection probability result of (a) GRX, (b) LRX, (c) LPD, (d) LRD-GRX, and (e) BJSRD; (f)–(j) refer to the binary map @ 1e-3 FAR of (f) GRX, (g) LRX, (h) LPD, (i) LRD–GRX, and (j) BJSRD; and (k) is the reference location.

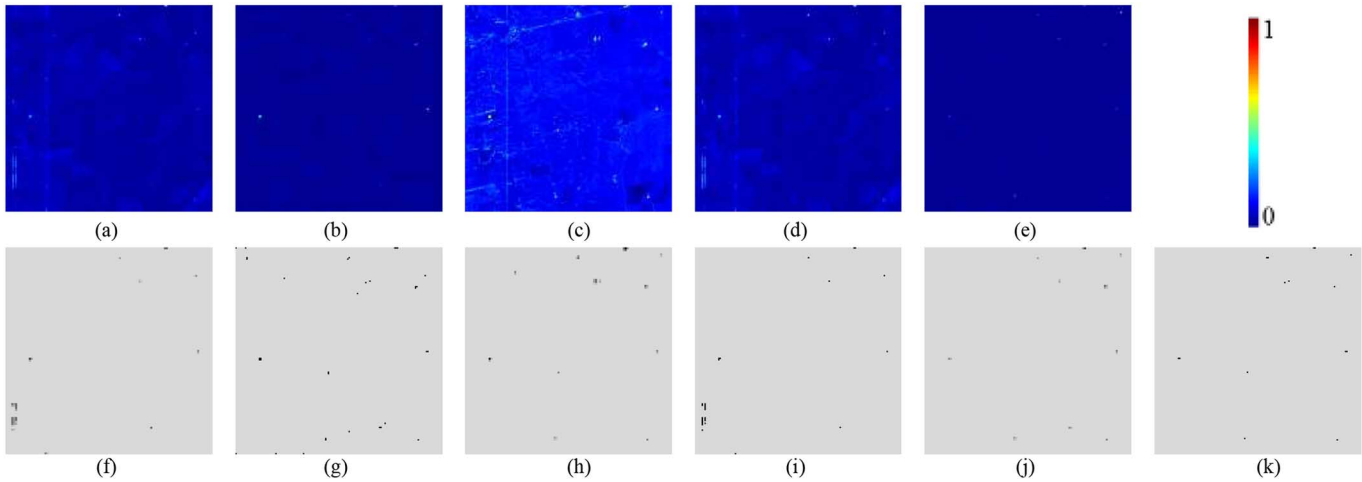


Fig. 7. Detection performances of the AD algorithms for the real Hyperion dataset: (a)–(e) refer to the detection probability result of (a) GRX, (b) LRX, (c) LPD, (d) LRD-GRX, and (e) BJSRD; (f)–(j) refer to the binary map @ 1e-3 FAR of (f) GRX, (g) LRX, (h) LPD, (i) LRD–GRX, and (j) BJSRD; and (k) is the reference location.

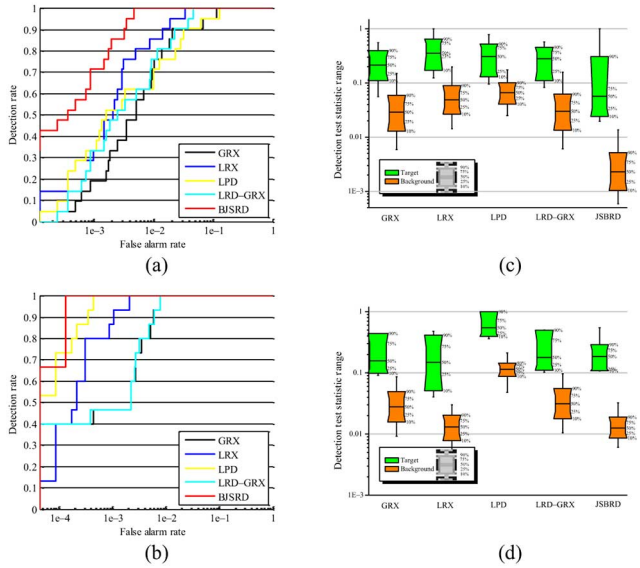


Fig. 8. ROC curves and statistical separability analysis of the AD algorithms: (a) and (b) ROC curves of the HYDICE and Hyperion datasets, respectively; (c) and (d) statistical separability analysis of the HYDICE and Hyperion datasets, respectively.

a green box representing the real target pixels, and an orange box representing the range of the background pixels in the scene. Each box provides the detailed value distribution of the detection test statistic results, as shown in the legend. The gap between the lower bound of the green box and the upper bound of its counterpart orange box refers to the separability between the target and the background pixels in the dataset. In both subfigures, BJSRD shows the best separability that confirms the effectiveness of the redundant prior of the background utilized in the proposed algorithm.

For the running time comparison, the running times of each of the AD algorithms implemented in MATLAB code are shown in the second item of Table II. Here, it can be seen that the global-based methods are faster than the local-based methods. Furthermore, LPD with one singular value decomposition (SVD) calculation and GRX with one covariance matrix inverse calculation are the fastest. The main computational burden for the LRD–GRX method comes from the SVD calculations in the iteration of the low-rank constraint-based preprocessing. The global version of the proposed work BJSRD(f), which utilizes the whole scene to construct the original dictionary $\mathbf{H}$, calls for enormous computational burden. The result of BJSRD(f)

TABLE III
OPTIMAL PARAMETER SETTINGS FOR THE LOCAL DETECTORS

	BJSRD				LRX	
	$b\times b$	$g\times g$	$d\times d$	L	$b\times b$	$g\times g$
HYDICE	17×17	5×5	19×19	3	63×63	1×1
Hyperion	9×9	1×1	11×11	6	29×29	3×3

with the Hyperion dataset is omitted, as it costs far more than 1e5 s. It can still be seen that the AUC value of BJSRD(f) is inferior to BJSRD, as the local approach ignoring the irrelevant pixels far away from $\mathbf{S}$ leads to a more stable estimation by the use of SOMP. For the local-based methods, the optimal parameter settings are shown in Table III. In these cases, it is shown that BJSRD is faster than LRX, which calls for multiple local inverse covariance estimation. Compared with the global methods, the computational load of the JSR is the main drawback of the proposed BJSRD. To sum up, although the proposed method takes the most time, the burden is still acceptable when considering its superior detection performance.

C. Parameter Analysis

This section examines the effect of the parameters on the detection performance of the proposed BJSRD algorithm. The optimal parameter settings for BJSRD are shown in Table III. In this experiment, we fixed the other parameters as the corresponding optimal and focused on specific parameter(s) at a time. For the two-dimensional (2-D) subfigures below, each horizontal axis is the parameter, and the vertical axis shows the AUC. For the three-dimensional (3-D) subfigures, the two horizontal axes are the related parameters, while the vertical axis records the AUC value.

1) Dual Spatial Window: We first varied the size of the local neighboring window $g \times g$ from 1×1 to 25×25 (where 1×1 means no guard window) and the outer window $b \times b$ was varied from 5×5 to 17×17 for the proposed BJSRD, and a larger range for LRX, to investigate the effect of the dual window size. In Fig. 9(b) and (d), the AUC increases with the growth of the dual window size, and then begins to slowly decrease after the maximum value. It should be noted that most dual window sets can still obtain a reasonable detection result, which verifies the robustness of this parameter set. Meanwhile, as seen in Fig. 9(a) and (c), LRX requires many more neighboring pixels to estimate the surrounding background information, and it costs more running time in practice.

2) Size of Dictionary: For the size of the local dictionary, the value was ranged from 9 to 37 for all the experiments, to investigate the sensitiveness of the proposed algorithm. As shown in Fig. 10, it is suggested that this term is relatively stable. In this way, it is believed that setting a local search window to construct the dictionary, instead of the whole scene, is meaningful, not only for the sake of the computational complexity, but also for a more robust detection performance. Considering the complicated background distribution of the real data experiments, the size of the local window and the size of the search window needed to be large enough to fully construct the major background information of the local patch.

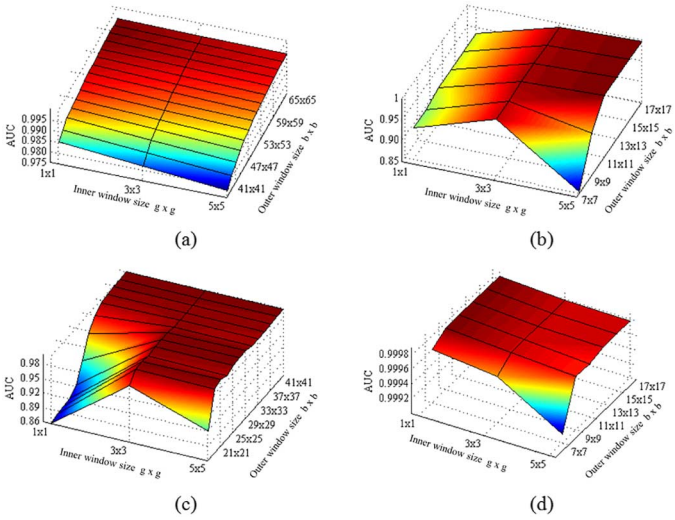


Fig. 9. AUC versus the dual spatial window size. (a) LRX with the HYDICE dataset. (b) BJSRD with the HYDICE dataset. (c) LRX with the Hyperion dataset. (d) BJSRD with the Hyperion dataset.

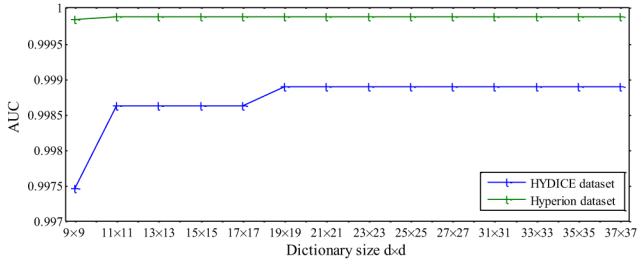


Fig. 10. AUC versus the local dictionary size for the proposed algorithm with the two hyperspectral datasets.

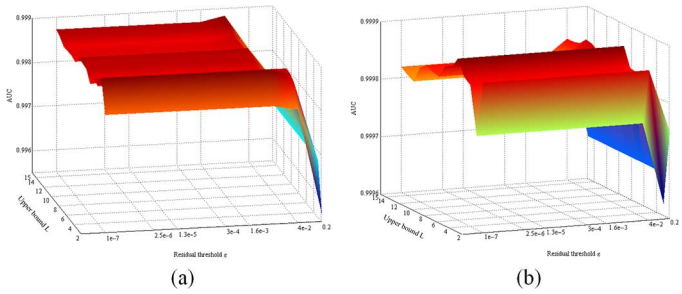


Fig. 11. AUC versus the SOMP-related parameters for the proposed algorithm with the two hyperspectral datasets. (a) HYDICE dataset. (b) Hyperion dataset.

3) Parameters of the SOMP Step: Finally, the residual threshold ε was varied from $L1e-7$ to $1e-1$, and the upper bound of the sparsity level L was varied from 2 to 15 for the two datasets. Since these two parameters are related to the SOMP step, a 3-D plot is utilized to simultaneously display the performances. It can also be seen in Fig. 11 that the residual threshold is insensitive to the detection result over a wide reasonable range, as the optimal residual threshold value is omitted in Table III. Once ε exceeds a certain threshold, the local neighboring pixel set will be linearly reconstructed with a large residual, which cannot be utilized to measure the separability of the test pixel and its surrounding neighborhood pixels. For the upper bound of the sparsity level, the performance is

quite stable, which validates the effectiveness of utilizing only a few active dictionary atoms to linearly represent the local neighboring pixel set. As fully analyzed in [45],it also worth mentioning that the optimal L value is related to the dominant background components. In view of this, the detailed optimal values mentioned before suggest that the proposed algorithm is not sensitive to the joint sparsity level, and the low-rank characteristic of the HSI plays the key role. All in all, the parameter setting is quite robust within the proposed algorithm, which shows an acceptable detection performance over a large parameter range.

V. Conclusion

This paper has proposed a JSR-based framework for HSI AD. The proposed algorithm utilizes the redundant background information in the hyperspectral scene, and automatically deals with the complicated multiple background classes, without estimating the statistical information of the background. Whether the test pixel belongs to an anomaly or not, it is judged by measuring the length of the matched projection on the orthogonal complementary background subspace that is estimated by the JSR. The proposed BJSRD method was tested on two widely used real HSI datasets, and the experimental results confirm the effectiveness of the proposed anomaly detector.

The proposed framework could still be further improved in some aspects. For instance, the proposed method works well under the assumption that the subspace for each background and target is linearly divisible; however, nonlinear cases should also be considered. A method of designing a unified overcomplete dictionary is also of interest. Our future work will therefore focus on how to extend the proposed detector to a nonlinear kernel and a global version.

Acknowledgment

The authors would like to thank the anonymous reviewers for their insightful and constructive comments. They would also like to thank the researchers for providing the real HYDICE and Hyperion datasets.

References

[1] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, 1998.

[2] C. O. Justice *et al.*, "The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1228–1249, Jul. 1998.

[3] J. C. Harsanyi and C. I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[4] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

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

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

[7] 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.

[8] L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

[9] 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.

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

[11] 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.

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

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

[14] S. Khazai *et al.*, "An approach for subpixel anomaly detection in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 769–778, Apr. 2013.

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

[16] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[17] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.

[18] 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.

[19] 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.

[20] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[21] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. New York, NY, USA: Academic, 2008.

[22] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[23] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[24] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[25] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C.R. Math.*, vol. 346, no. 9, pp. 589–592, 2008.

[26] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[28] J. Yang *et al.*, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[29] M. D. Iordache, J. 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.

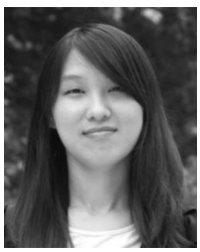
[30] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intra-class variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.

[31] D. Weisheng, S. Guangming, and L. Xin, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, Feb. 2013.

[32] J. Li *et al.*, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014.

[33] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[34] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[35] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.
[36] S. J. Kim *et al.*, "An interior-point method for large-scale l1-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.
[37] D. L. Donoho and Y. Tsaig, "Fast solution of l1-norm minimization problems when the solution may be sparse," Dept. Statistics, Stanford Univ., Stanford, CA, USA, Tech. Rep. 1, 2006 [Online]. Available: http://statweb.stanford.edu/~donoho/reports.html
[38] R. Tibshirani, "Regression shrinkage and selection via the lasso: A retrospective," *J. Roy. Stat. Soc. B*, vol. 73, no. 3, pp. 273–282, 2011.
[39] M. F. Duarte *et al.*, "Distributed compressed sensing of jointly sparse signals," in *Proc. Conf. Rec. 39th Asilomar Conf. Signals Syst. Comput.*, 2005, pp. 1537–1541.
[40] H. Zhang *et al.*, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.
[41] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Berlin, Germany: Springer-Verlag, 2010.
[42] V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Automatic target recognition based on simultaneous sparse representation," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, 2010, pp. 1377–1380.
[43] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, Jun. 2014.
[44] S. Matteoli *et al.*, "Models and methods for automated background density estimation in hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2837–2852, May 2013.
[45] H. Zhang *et al.*, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
[46] J. Li, H. Zhang, and L. Zhang, "Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification," *ISPRS J. Photogramm.*, vol. 94, no. 4, pp. 25–36, 2014.
[47] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.
[48] S. Y. Chen, S. Yang, K. Kalpakis, and C. I. Chang, "Low-rank decomposition-based anomaly detection," in *Proc. SPIE Def., Secur., and Sens.*, International Society for Optics and Photonics, 2013, p. 87430N.



Jiayi Li (S'13) received the B.S. degree in geomatics engineering from the Central South University, Changsha, China, in 2011. Currently, she is pursuing the Ph.D. degree at the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China.

Her research interests include hyperspectral imagery, sparse representation, computation vision and pattern recognition, and remote sensing images.

Miss. Li is a reviewer of more than five international journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE SIGNAL PROCESSING LETTER, and the *International Journal of Remote Sensing*.

Hongyan Zhang (M'13) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

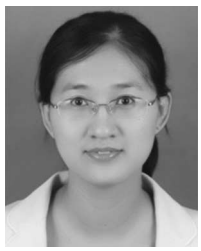
Currently, he is an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He has authored more than 30 research papers. His research interests include image reconstruction, hyperspectral image processing, sparse representation, and low-rank methods for sensing image imagery.

Dr. Zhang is a reviewer of about 10 international academic journals.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

Currently, he is the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. Currently, he is a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 410 research papers. He is the holder of 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Engineering and Technology, an Executive Member (Board of Governor) of the China National Committee of the International Geosphere-Biosphere Programme, and an Executive Member of the China Society of Image and Graphics. He regularly serves as a Co-Chair of the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-Spatial Information Science*, the *Journal of Remote Sensing*, and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was the recipient of the 2010 the Best Paper Boeing Award and 2013 Best Paper ERDAS Award of American Society of Photogrammetry and Remote Sensing (ASPRS), respectively.



Li Ma (M'13) received the B.S. degree in biomedical engineering and M.S. degree in pattern recognition and intelligent system from Shandong University, Jinan, China, in 2004 and 2006, respectively, and the Ph.D. degree in pattern recognition and intelligent system from Huazhong University of Science and Technology, Wuhan, China, in 2011.

During 2008–2010, she was a Visiting Scholar at Purdue University, West Lafayette, IN, USA. Currently, she is an Assistant Professor with the Faculty of Mechanical and Electronic Information, China University of Geosciences, Wuhan, China. Her research interests include hyperspectral data analysis, pattern recognition, and remote sensing applications.