# Collaborative Representation for Hyperspectral Anomaly Detection

Wei Li, *Member, IEEE*, and Qian Du, *Senior Member, IEEE*

*Abstract*—In this paper, collaborative representation is proposed for anomaly detection in hyperspectral imagery. The algorithm is directly based on the concept that each pixel in background can be approximately represented by its spatial neighborhoods, while anomalies cannot. The representation is assumed to be the linear combination of neighboring pixels, and the collaboration of representation is reinforced by $\ell_2$-norm minimization of the representation weight vector. To adjust the contribution of each neighboring pixel, a distance-weighted regularization matrix is included in the optimization problem, which has a simple and closed-form solution. By imposing the sum-to-one constraint to the weight vector, the stability of the solution can be enhanced. The major advantage of the proposed algorithm is the capability of adaptively modeling the background even when anomalous pixels are involved. A kernel extension of the proposed approach is also studied. Experimental results indicate that our proposed detector may outperform the traditional detection methods such as the classic Reed–Xiaoli (RX) algorithm, the kernel RX algorithm, and the state-of-the-art robust principal component analysis based and sparse-representation-based anomaly detectors, with low computational cost.

*Index Terms*—Anomaly detection, collaborative representation, kernel collaborative representation, hyperspectral imagery (HSI), sparse representation.

## I. Introduction

**H**YPERSPECTRAL imagery (HSI), a 3-D "image cube," provides a wealth of spectral information to uniquely identify various materials by their reflective spectrum, which makes it possible to distinguish different objects of interest based on their spectral signatures. Target detection in remote sensing community seeks to discriminate uncommon observations, such as man-made objects, which usually have significantly different spectral signatures from natural background materials. Over the last two decades, target detection in HSI has drawn lots of attention due to its importance in many military and civilian applications [1]–[4].

Based on *a priori* target knowledge availability, target detection algorithms can be generally grouped into two cate-

W. Li is with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: liwei089@ieee.org).

Q. Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762 USA (e-mail: du@ece.msstate.edu).

gories: supervised and unsupervised. When the target spectral signature is known, the matched filtering strategy is usually considered. Spectral matched filter (SMF) [5] is a well-known detection approach using the known target spectral characteristics. The approach estimates the background covariance matrix and maximizes the signal-to-background ratio to distinguish targets of interest from the background. Its extension, called regularized SMF, has been introduced in [6]. Similar techniques include subspace-based methods, such as matched subspace detector [7] and adaptive subspace detector [8]. Some classifiers have also been modified for target detection, such as one-class support vector machine [9], sparse-representation-based detection [10]–[12], etc.

When the target spectral signature is unknown, anomaly detection has to be applied, which is to find anomalous pixels whose spectral signatures are different from their surroundings. The Reed–Xiaoli (RX) detector, which was introduced in [13], is built on the concept that a hypothesis testing can be formulated for a pixel vector and the conditional probability density functions (pdfs) under the two hypotheses (without and with anomaly) are assumed to be Gaussian. The solution of the resulting generalized likelihood ratio test turns out to be the Mahalanobis distance between the pixel under test and the background. Two typical versions of the RX have been studied: global RX, which estimates the background statistics (i.e., mean and covariance matrix) of the entire image, and local RX, which estimates the background using local statistics. Obviously, the key to success for global or local RX is an appropriate estimate of the global or local background covariance matrix for effective background suppression. In addition to the classical RX detector, a number of anomaly detection algorithms have also been proposed for hyperspectral data. A time-efficient method has been introduced for anomaly detection in [14] and [15], the subpixel anomaly detection was discussed in [16], multiple-window anomaly detection was developed in [17] to capture local spectral variations, a random-selection-based anomaly detector was described in [18], subspace-projection-based detectors were presented in [19]–[21], and kernel-based detectors, such as kernel RX and support vector data description (SVDD), were presented in [22]–[25].

In this paper, similar to SVDD, we propose a nonparametric anomaly detector that can adaptively estimate the background without assuming pdf or estimating its covariance matrix. The proposed detector is designed using the concept that each pixel in background can be approximately represented by its spatial neighborhoods, while anomalies cannot. The representation is assumed to be the linear combination of neighboring pixels, and the collaboration among these pixels is reinforced by $\ell_2$-norm

minimization of the representation weight vector. To adjust the contribution of each neighboring pixel, a distance-weighted regularization matrix is included in the optimization problem, which has a simple and closed-form solution. By imposing the sum-to-one constraint to the weight vector, the stability of the solution can be enhanced. Similar to the local RX, the proposed collaborative-representation-based detector (CRD) also adopts a dual-window strategy: surrounding spatial neighborhoods in the outer region are linearly combined to produce a prediction for the test (central) pixel within a sliding dual rectangular window. In this way, background estimation can be achieved. Once the background image is available, the anomalies can be determined in the residual image which is obtained by subtracting the predicted background from the original HSI. Note that the concept of collaborative representation has been presented for face recognition [26], [27], where it is a supervised method and the mechanism to estimate the weight vector is different from the proposed CRD.

Obviously, the counterpart of the proposed CRD is a sparse-representation-based detector (SRD). The supervised hyperspectral target detection using sparse representation has been discussed in [10] and [11]. In this paper, unsupervised target detection, i.e., anomaly detection, with sparse representation will be discussed and compared with the proposed CRD. Recently, robust principal component analysis (RPCA) [28] has been presented for low-rank and sparse matrix decomposition. The typical PCA provides the optimal low-rank representation in an $\ell_2$-sense when the data are corrupted as long as the magnitude of noise is small. However, it breaks down under arbitrary corruption, which affects only very few of the additional observations [29], [30]. Comparatively, RPCA seeks to exactly recover the underlying low-rank structure (background modeling) and a sparse structure (residual error) from the original data, even in the presence of large noise. Due to its characteristics, it is straightforward to apply RPCA for anomaly detection in HSI, which usually assumes that the background image is low-rank while anomalies are preserved in the residual (sparse) image. We will compare the proposed CRD with the traditional RX algorithm, SRD, and the state-of-the-art RPCA-based anomaly detectors. It is also extended to the kernel version and compared with the counterpart, i.e., kernel RX.

The uniqueness of the proposed CRD algorithm includes the following: 1) As far as we know, it is the first time that collaborative representation is applied for anomaly detection, which is a nonparametric approach without estimation of the background covariance matrix (a conference version of this paper with preliminary results can be found in [31]); 2) the collaborative representation is further regularized by a similarity matrix such that it can adaptively handle the situation when the background includes anomalous pixels, avoiding the contamination when using all of the pixels to estimate the background; 3) it has a closed-form solution, which is much simpler than a sparseness-constrained detector; and 4) by imposing the sum-to-one constraint to representation coefficients, the solution of the $\ell_2$-norm minimization becomes more stable, resulting in better discrimination power. Compared to the original collaborative representation in [26] and [27], our method includes distance-based regularization terms, which is critical

in anomaly detection since not all of the background pixels can be used for representation in practice. Note that our previous work on nearest regularized subspace (NRS) [32] uses the distance-weighted regularization, which is called Tikhonov regularization [33]; however, NRS in [32] was implemented in a supervised fashion because the representation is based on the labeled samples within each class.

This paper is organized as follows. Section II provides a detailed description of the proposed CRD detector as well as its kernel version. Section III describes the experimental setup and detection performance and shows how to optimize the proposed method. Finally, concluding remarks are provided in Section IV.

## II. COLLABORATIVE REPRESENTATION FOR ANOMALY DETECTION

### A. Collaborative-Representation-Based Detector

Consider a 3-D hyperspectral cube with resized samples $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ in $\mathbb{R}^d$ ($d$ is the number of spectral bands and $n$ is the total number of samples). For each pixel $\mathbf{y}$ (of size $d \times 1$), surrounding data are collected inside the outer window (of size $w_{\text{out}} \times w_{\text{out}}$) while outside the inner window (of size $w_{\text{in}} \times w_{\text{in}}$), centered at the pixel $\mathbf{y}$. The selected data are resized into a 2-D matrix $\mathbf{X}_s = \{\mathbf{x}_i\}_{i=1}^s$ ($s$ is the number of chosen samples; $s = w_{\text{out}} \times w_{\text{out}} - w_{\text{in}} \times w_{\text{in}}$). Therefore, the matrix $\mathbf{X}_s$ (of size $d \times s$) is obtained for every pixel $\mathbf{y}$ on its own local window. Note that, in the local case, $\mathbf{X}_s$ is adaptive for each test pixel. The objective is to find weight vector $\boldsymbol{\alpha}$ such that $\|\mathbf{y} - \mathbf{X}_s\boldsymbol{\alpha}\|_2^2$ is minimized under the constraint that $\|\boldsymbol{\alpha}\|_2^2$ is also minimized. Therefore, the objective function is

$$\arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{X}_s\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2 \qquad (1)$$

where $\lambda$ is a Lagrange multiplier. Equation (1) is equivalent to

$$\arg \min_{\boldsymbol{\alpha}} \left[ \boldsymbol{\alpha}^T \left( \mathbf{X}_s^T \mathbf{X}_s + \lambda\mathbf{I} \right) \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^T \mathbf{X}_s^T \mathbf{y} \right]. \qquad (2)$$

Taking derivative with regard to $\boldsymbol{\alpha}$ and setting the resultant equation to zero yield

$$\hat{\boldsymbol{\alpha}} = \left( \mathbf{X}_s^T \mathbf{X}_s + \lambda\mathbf{I} \right)^{-1} \mathbf{X}_s^T \mathbf{y}. \qquad (3)$$

The parameter $\lambda$ controls the penalty of the norm of weight vectors. However, some surrounding pixels may be quite similar to the center pixel and should allow them to have large coefficients in the representations; for those surrounding pixels that are quite different from the center pixel, the coefficients should be small, and the penalty for having a large coefficient must be heavy. In [31] and [32], a distance-weighted Tikhonov regularization has been considered to adjust the weight vector, and in this work, we also employ the technique by using the following diagonal regularization matrix:

$$\boldsymbol{\Gamma}_{\mathbf{y}} = \begin{bmatrix} \|\mathbf{y} - \mathbf{x}_1\|_2 & & 0 \\ & \ddots & \\ 0 & & \|\mathbf{y} - \mathbf{x}_s\|_2 \end{bmatrix} \qquad (4)$$
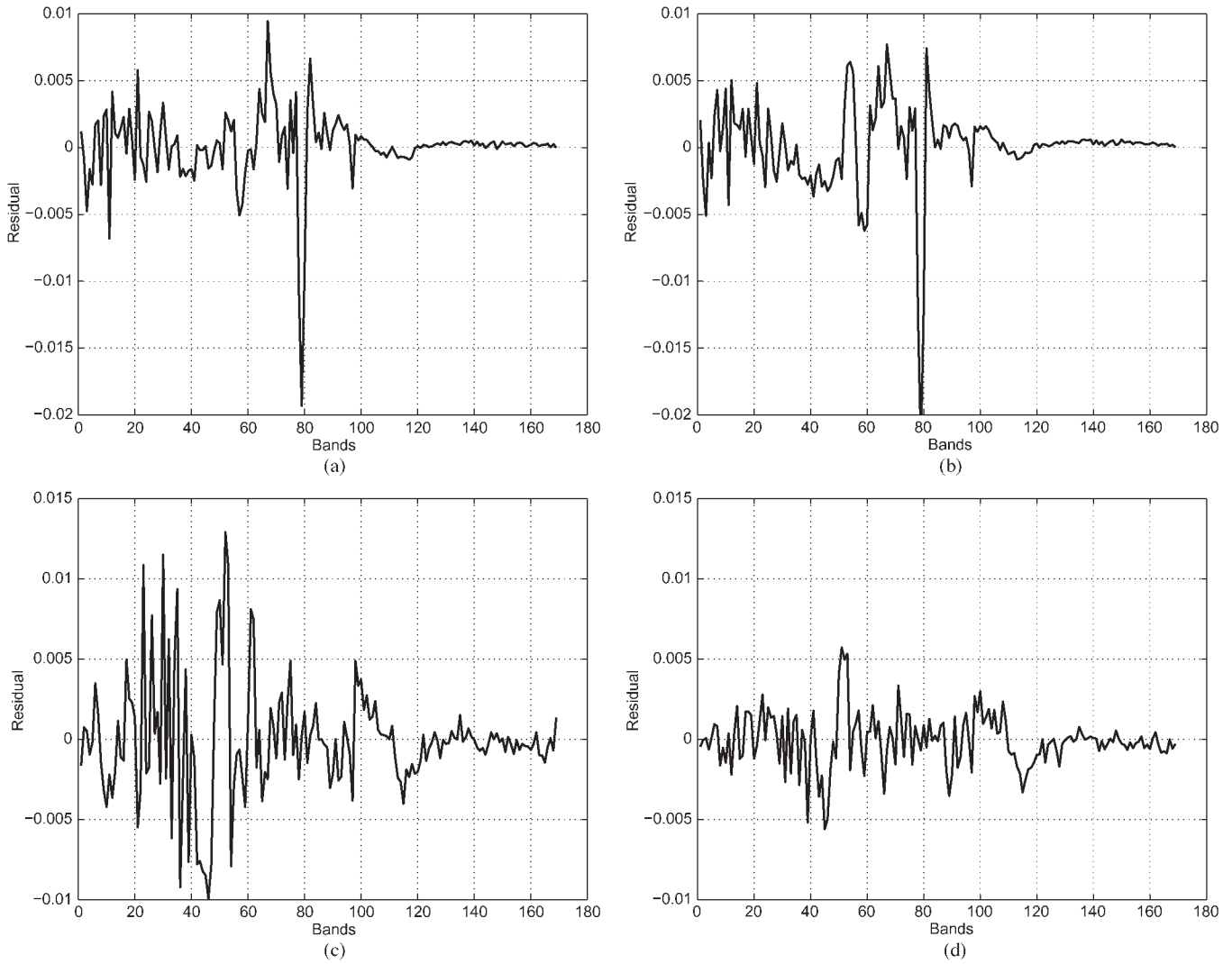
Fig. 1. Examples of how the sum-to-one constraint affects the output of the proposed CRD for HYDICE forest data. (a) and (b) represent residual vectors between a chosen normal pixel and its approximation, and (c) and (d) represent residual vectors between a chosen anomalous pixel and its approximation. (a) CRD with constraint for a normal pixel. (b) CRD without constraint for a normal pixel. (c) CRD with constraint for an anomalous pixel. (d) CRD without constraint for an anomalous pixel.

where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s$ are the columns of $\mathbf{X}_s$. It calculates the Euclidean distance between the center pixel to each of the pixels in $\mathbf{X}_s$. Then, the improved optimization problem becomes

$$\arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{X}_s\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\Gamma}_\mathbf{y}\boldsymbol{\alpha}\|_2^2 \tag{5}$$

which means that the penalty values on different elements in $\boldsymbol{\alpha}$ are different, depending upon the distance (or similarity) of the corresponding pixel to the center pixel. Intuitively, if the distance is small, which means that the corresponding pixel is similar to the center pixel, then its coefficient is allowed to be large; otherwise, the coefficient must be small. Following the same derivation, the solution to (5) is:

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{X}_s^T\mathbf{X}_s + \lambda\boldsymbol{\Gamma}_\mathbf{y}^T\boldsymbol{\Gamma}_\mathbf{y}\right)^{-1}\mathbf{X}_s^T\mathbf{y}. \tag{6}$$

To impose the sum-to-one constraint on $\boldsymbol{\alpha}$, the objective is modified such that $\|\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}_s\boldsymbol{\alpha}\|_2^2$ is minimized under the constraint that $\|\boldsymbol{\alpha}\|_2^2$ is also minimized, where $\widetilde{\mathbf{y}} = [\mathbf{y}; 1]$ and $\widetilde{\mathbf{X}}_s = [\mathbf{X}_s; \mathbf{1}]$, and $\mathbf{1}$ is a $1 \times s$ row vector of all ones; in other words,

the data dimensionality is increased by 1 after inserting 1's to the lately added dimension. After that, the new optimization problem can be represented as

$$\arg \min_{\hat{\boldsymbol{\alpha}}} \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}_s\hat{\boldsymbol{\alpha}}\|_2^2 + \lambda\|\boldsymbol{\Gamma}_\mathbf{y}\hat{\boldsymbol{\alpha}}\|_2^2 \tag{7}$$

and its solution is

$$\hat{\boldsymbol{\alpha}} = \left(\widetilde{\mathbf{X}}_s^T\widetilde{\mathbf{X}}_s + \lambda\boldsymbol{\Gamma}_\mathbf{y}^T\boldsymbol{\Gamma}_\mathbf{y}\right)^{-1}\widetilde{\mathbf{X}}_s^T\widetilde{\mathbf{y}}. \tag{8}$$

In CRD, once the representation process is finished, anomalies can be determined in the residual image which is obtained by directly subtracting the predicted background from the original hyperspectral data as

$$r_1 = \|\mathbf{y} - \hat{\mathbf{y}}\|_2 = \|\mathbf{y} - \mathbf{X}_s\hat{\boldsymbol{\alpha}}\|_2. \tag{9}$$

If it is larger than a threshold, then $\mathbf{y}$ is claimed to an anomalous pixel. The overall description of the CRD-based anomaly detection algorithm is given as Algorithm 1.
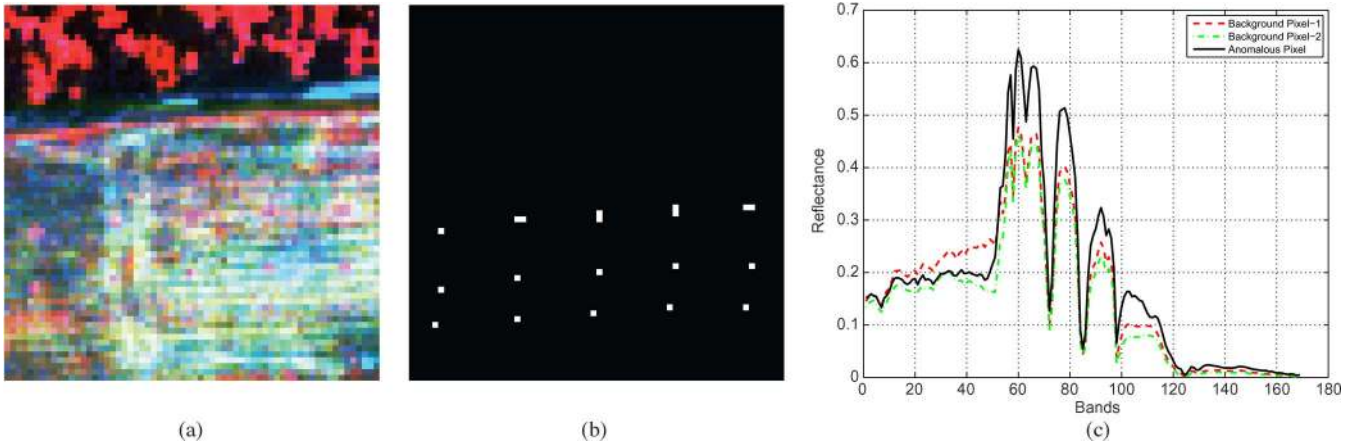
Fig. 2. (a) Pseudocolor image of the HYDICE forest scene. (b) Ground-truth map of 19 anomalous pixels. (c) Spectral signatures of two background (normal) pixels and one anomalous pixel.

---

**Algorithm 1** The CRD Algorithm for Anomaly Detection

---

**Input**: Three-dimensional hyperspectral cube, window size $(w_{\text{out}}, w_{\text{in}})$, and the regularization parameter $\lambda$
**for all** pixels **do**
   1) For each test pixel $\mathbf{y}$, a matrix $\mathbf{X}_s = \{\mathbf{x}_i\}_{i=1}^s$ is collected based on the dual window strategy;
   2) The weight vector $\boldsymbol{\alpha}$ is calculated via the $\ell_2$-regularized minimization with a closed-form of Eq. (6) or (8);
   3) Eq. (9) is used to obtain the distance measurement, which is further compared with the prescribed threshold;
**end for**
**Output**: Anomaly detection map.

---

It is worth mentioning that, for another important application of hyperspectral data, i.e., hyperspectral unmixing [34], sum-to-one is often imposed on abundances as a constraint, and the resulting constrained least squares problem has been presented in previous works (e.g., [34]). Nevertheless, the purpose of imposing the sum-to-one constraint in this research is mainly to help control the range of weights. Our proposed CRD builds on the concept that the central pixel is linearly reconstructed using surrounding neighbors. Empirically, we have found that CRD with and without constraint can produce similar weight vectors in homogeneous areas; however, in heterogeneous areas, the CRD without the sum-to-one constraint tends to use significant weights for the normal background pixels to represent a central anomalous pixel, causing the central pixel being misclaimed to be a normal one. Overall, the sum-to-one constraint makes the solution more stable, resulting in better discriminant power.

Fig. 1 illustrates how the sum-to-one constraint affects the output of the proposed CRD using the Hyperspectral Digital Imagery Collection Experiment (HYDICE) forest data to be introduced in Section III. Fig. 1(a) and (b) represents residual vectors between a chosen normal pixel and its estimation with and without the constraint, respectively, and Fig. 1(c) and (d) represents residual vectors between a chosen anomalous pixel and its estimation with and without the constraint, respectively.

For both cases, 112 surrounding normal pixels are used to calculate the weight vector. According to (9), the outputs of CRD (i.e., the norm of the residual vector) in Fig. 1(a) and (b) are 0.2535 and 0.3112, respectively, which means that the center pixel is easier to be claimed as a normal one with a smaller output when the constraint is imposed. On the other hand, in Fig. 1(c) and (d), the outputs are 0.3983 and 0.2017, respectively; the CRD with the sum-to-one constraint generates a larger output, thereby facilitating the discrimination of the anomaly from its background.

Since the anomalous pixels in hyperspectral data are rarely present, we assume that most of the samples $\mathbf{X}_s$ in an outer region are background (normal) pixels, and one of the following situations may appear.

1) If the test pixel $\mathbf{y}$ is a normal pixel and none of the samples from $\mathbf{X}_s$ is anomalous either, the regularization matrix $\boldsymbol{\Gamma}_{\mathbf{y}}$ will have less weighting power since most samples from surrounding data $\mathbf{X}_s$ are similar to $\mathbf{y}$, and the diagonal elements of $\boldsymbol{\Gamma}_{\mathbf{y}}$ are small, which causes the approximation $\widehat{\mathbf{y}}$ to be close to $\mathbf{y}$ (and the decision to $\mathbf{y}$ is "normal").

2) If the test pixel $\mathbf{y}$ is a normal pixel and one or two samples from $\mathbf{X}_s$ are anomalous, the corresponding values for the anomalous samples in the regularization matrix $\boldsymbol{\Gamma}_{\mathbf{y}}$ are large, resulting in less contributions. As a result, $\mathbf{y}$ is mainly represented by other normal pixels in $\mathbf{X}_s$, which causes the approximation of $\widehat{\mathbf{y}}$ to be close to $\mathbf{y}$ (and the decision to $\mathbf{y}$ is "normal").

3) If the test pixel $\mathbf{y}$ is an anomalous pixel and none of the samples from $\mathbf{X}_s$ is anomalous, the regularization matrix $\boldsymbol{\Gamma}_{\mathbf{y}}$ will have more weighting power since the diagonal elements of $\boldsymbol{\Gamma}_{\mathbf{y}}$ are relatively large, which causes the approximation $\widehat{\mathbf{y}}$ to be different from $\mathbf{y}$ (and the decision to $\mathbf{y}$ is most likely "abnormal").

4) If the test pixel $\mathbf{y}$ is an anomalous pixel and one or two samples from $\mathbf{X}_s$ are also anomalous but dissimilar to $\mathbf{y}$ (which means that they are different types of anomalies), the regularization matrix $\boldsymbol{\Gamma}_{\mathbf{y}}$ will have weighting power since all of the samples from the surrounding data are
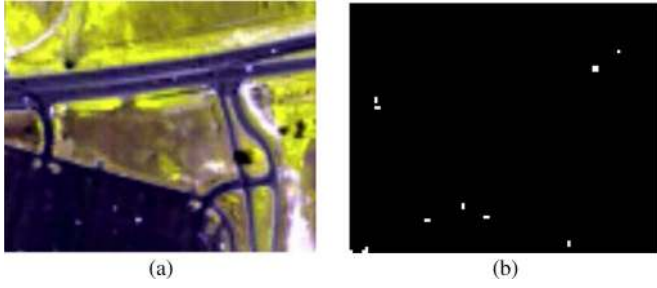
Fig. 3. (a) Pseudocolor image of the HYDICE urban scene. (b) Ground-truth map of 21 anomalous pixels.

dissimilar to $\mathbf{y}$, which causes the approximation $\widehat{\mathbf{y}}$ to be different from $\mathbf{y}$ (and the decision to $\mathbf{y}$ is "abnormal").

5) If the test pixel $\mathbf{y}$ is an anomalous pixel and one or two samples from $\mathbf{X}_s$ are also anomalous and similar to $\mathbf{y}$, $\mathbf{\Gamma_y}$ will have adaptive weighting power since the elements of $\mathbf{\Gamma_y}$ have such distribution—relevant positions have small values and others have large values, which causes the corresponding weight vector $\boldsymbol{\alpha}$ to be adaptively adjusted (i.e., if $\mathbf{x}_i$ is similarly anomalous as $\mathbf{y}$, then the $i$th element in $\mathbf{\Gamma_y}$ is small, which allows $\alpha_i$ to be large but others are small), and the approximation $\widehat{\mathbf{y}}$ is still close to $\mathbf{y}$ (and the decision is "locally normal" since there have been other similar anomalous pixels in the neighborhood).

## B. Kernel Collaborative-Representation-Based Detector

Kernel methods can project the linearly nonseparable data into a high-dimensional feature space in which those data become more separable [35], [36]. Kernel-based versions of feature extraction or anomaly detection algorithms have recently been investigated in hyperspectral image analysis, such as kernel principal component analysis [37], kernel discriminant analysis [38], kernel local Fisher discriminant analysis [39], and kernel RX [22]. Here, we further extend the proposed CRD into a kernel version, referred to as kernel CRD.

In a kernel-induced feature space, we can linearly represent the center pixel $\mathbf{y}$ in terms of its surrounding data $\mathbf{X}_s$. The new optimization problem becomes

$$\arg \min_{\boldsymbol{\alpha}^*} \|\Phi(\mathbf{y}) - \mathbf{\Phi}\boldsymbol{\alpha}^*\|_2^2 + \lambda \left\|\mathbf{\Gamma}_{\Phi(\mathbf{y})}\boldsymbol{\alpha}^*\right\|_2^2 \quad (10)$$

where the mapping function $\Phi$ maps the center pixel to the kernel-induced feature space: $\mathbf{y} \rightarrow \Phi(\mathbf{y}) \in \mathbb{R}^{D \times 1}$ ($D \gg d$ is the dimension of kernel feature space), and $\mathbf{\Phi} = [\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_s)] \in \mathbb{R}^{D \times s}$. The new biasing Tikhonov matrix $\mathbf{\Gamma}_{\Phi(\mathbf{y})}$ then has the form of

$$\mathbf{\Gamma}_{\Phi(\mathbf{y})} = \begin{bmatrix} \|\Phi(\mathbf{y}) - \Phi(\mathbf{x}_1)\|_2 & & 0 \\ & \ddots & \\ 0 & & \|\Phi(\mathbf{y}) - \Phi(\mathbf{x}_s)\|_2 \end{bmatrix} \quad (11)$$

where $\|\Phi(\mathbf{y}) - \Phi(\mathbf{x}_i)\|_2 = [k(\mathbf{y}, \mathbf{y}) + k(\mathbf{x}_i, \mathbf{x}_i) - 2k(\mathbf{y}, \mathbf{x}_i)]^{1/2}$, $i = 1, 2, \ldots, s$. After constituting $\mathbf{\Gamma}_{\Phi(\mathbf{y})}$, the weight vector $\boldsymbol{\alpha}^*$
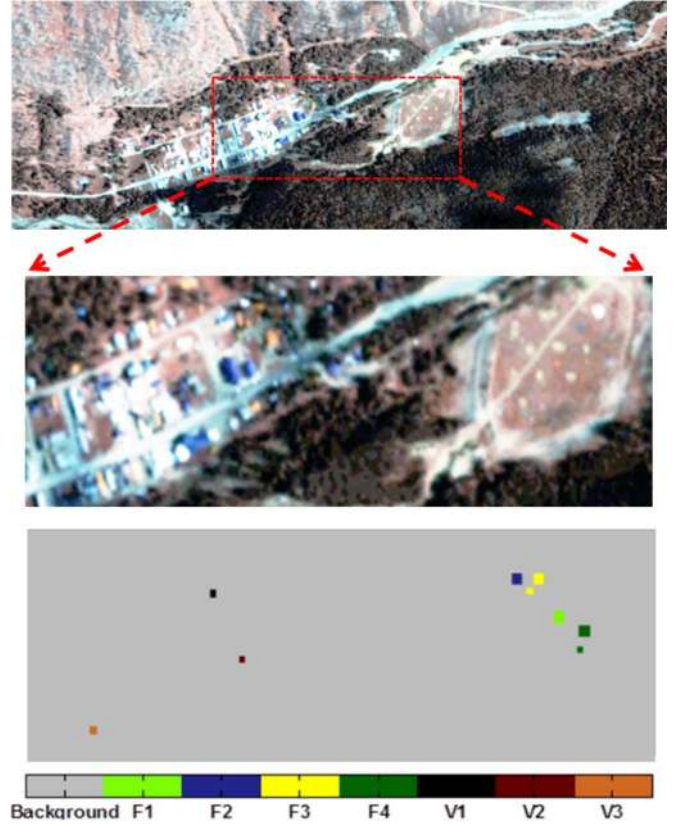


Fig. 4. Pseudocolor image of the HyMap image and ground-truth map of a total of seven types of targets (F1: 3-m red cotton target; F2: 3-m yellow nylon target; F3: 1- and 2-m blue cotton target; F4: 1- and 2-m red nylon target; V1: 1993 Chevy Blazer; V2: 1997 Toyota T100; and V3: 1985 Subaru GL Wagon).

with size of $s \times 1$ can be recovered in a closed-form solution

$$\boldsymbol{\alpha}^* = \left(\mathbf{\Phi}^T \mathbf{\Phi} + \lambda \mathbf{\Gamma}_{\Phi(\mathbf{y})}^T \mathbf{\Gamma}_{\Phi(\mathbf{y})}\right)^{-1} \mathbf{\Phi}^T \Phi(\mathbf{y})$$

$$= \left(\mathbf{K} + \lambda \mathbf{\Gamma}_{\Phi(\mathbf{y})}^T \mathbf{\Gamma}_{\Phi(\mathbf{y})}\right)^{-1} \mathbf{k}(\cdot, \mathbf{y}) \quad (12)$$

where $\mathbf{k}(\cdot, \mathbf{y}) = [k(\mathbf{x}_1, \mathbf{y}), k(\mathbf{x}_2, \mathbf{y}), \ldots, k(\mathbf{x}_s, \mathbf{y})]^T \in \mathbb{R}^{s \times 1}$ and kernel function $\mathbf{K} = \mathbf{\Phi}^T \mathbf{\Phi} \in \mathbb{R}^{s \times s}$ is the Gram matrix with $\mathbf{K}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$. The commonly used Gaussian radial basis function (RBF) kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|_2^2)$ ($\gamma > 0$ is the parameter of RBF kernel) is adopted here.

Then, the collaborative representation residual of the kernel CRD is represented as

$$r_2 = \|\Phi(\mathbf{y}) - \mathbf{\Phi}\boldsymbol{\alpha}^*\|_2$$

$$= \sqrt{\left(\Phi(\mathbf{y}) - \mathbf{\Phi}\boldsymbol{\alpha}^*\right)^T \left(\Phi(\mathbf{y}) - \mathbf{\Phi}\boldsymbol{\alpha}^*\right)}$$

$$= \sqrt{k(\mathbf{y}, \mathbf{y}) + \boldsymbol{\alpha}^{*T} \mathbf{K} \boldsymbol{\alpha}^* - 2\boldsymbol{\alpha}^{*T} \mathbf{k}(\cdot, \mathbf{y})}. \quad (13)$$

## C. Sparse-Representation-Based Detector

Note that in [10] and [12] the authors have proposed sparse-based $\ell_1$-minimization methods for hyperspectral target detection, which actually require *a priori* information, such as available training samples or the target spectral signature from a spectral library. In our work, the SRD with $\ell_1$-minimization
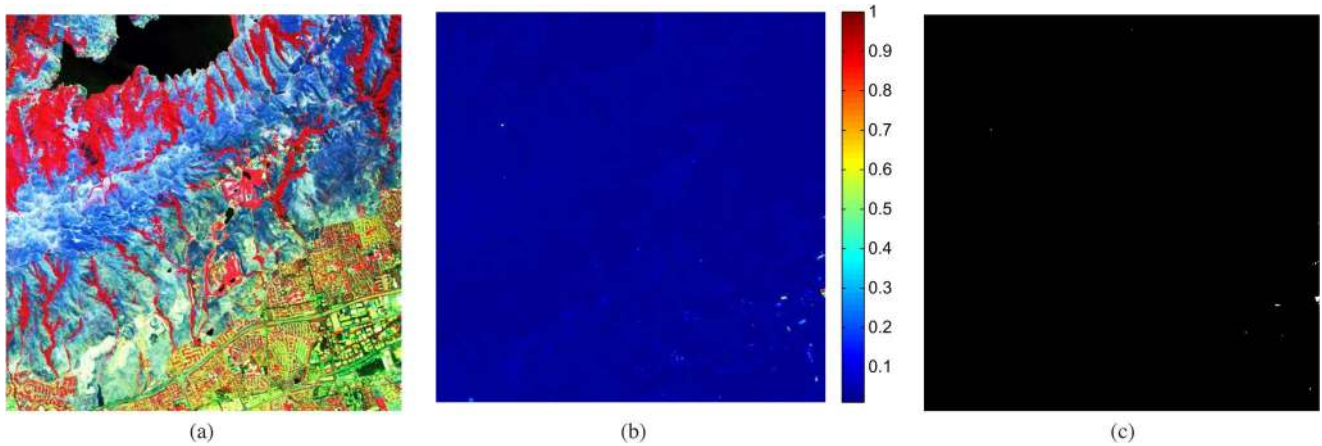
Fig. 5.　(a) Pseudocolor image of the Moffett Field scene. (b) Anomaly detection map of global RX. (c) Ground-truth map of 59 anomalous pixels.
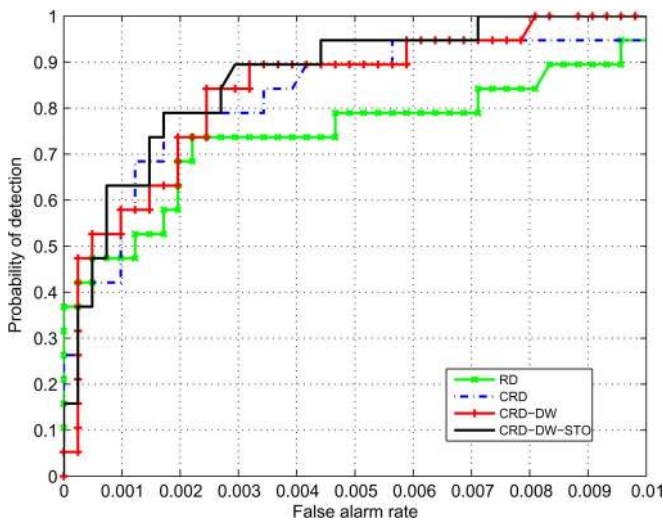


Fig. 6.　For the HYDICE forest data, detection performance of the RD, CRD, CRDDW, and CRD-DW-STO.
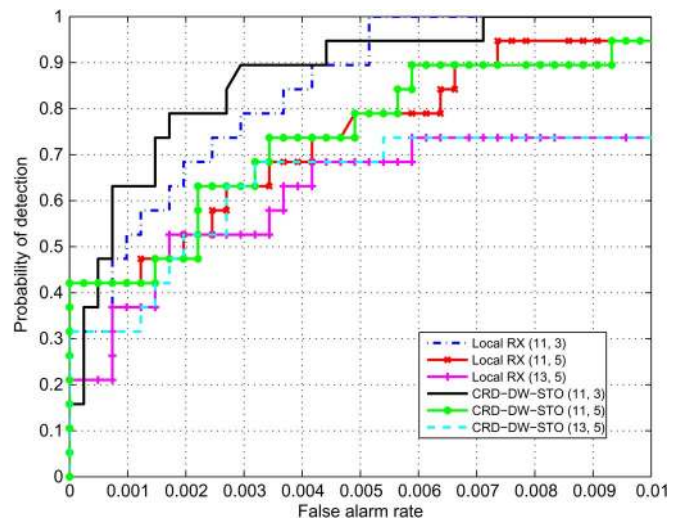


Fig. 7.　For the HYDICE forest data, sizes of window $(w_{\mathrm{out}}, w_{\mathrm{in}})$ adjustment for the proposed CRD-DW-STO and local RX.

is applied for background estimation in an unsupervised fashion. Different from (1), the weight vector $\boldsymbol{\alpha}$ using the $\ell_1$-minimization is defined as

$$\arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{X}_s \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1. \qquad (14)$$

The basic idea is to find a sparse solution $\boldsymbol{\alpha}$, which enables the approximation of $\mathbf{X}_s \boldsymbol{\alpha}$ to have desired matching compared to sample $\mathbf{y}$. It assumes that the matching pixels are sparsely distributed, and the number of matching pixels is limited.

### D. RPCA-Based Detector

RPCA [28]–[30] is a novel decomposition technique[1] tailored to the situation that the multidimensional data often comprise a low-rank term and a sparse term. In mathematical modeling, the observation data $\mathbf{X}$ can be expressed as

$$\mathbf{X} = \mathbf{L} + \mathbf{S} \qquad (15)$$

[1]http://perception.csl.illinois.edu/matrix-rank/sample_code.html.

TABLE I
AUC (IN PERCENT) PERFORMANCE OF THE PROPOSED CRD-DW-STO
AND LOCAL RX WITH VARYING WINDOW SIZE $(w_{\mathrm{out}}, w_{\mathrm{in}})$
AS WELL AS $\lambda$ FOR THE HYDICE FOREST DATA

| | | $w_{out}$ | | | | |
|---|---|---|---|---|---|---|
| | $w_{in}$ | 7 | 9 | 11 | 13 | 15 |
| $\lambda = 10^{-6}$ | 3 | 99.28 | 99.85 | **99.86** | 98.41 | 96.17 |
| | 5 | 99.65 | 99.71 | 99.70 | 97.59 | 95.02 |
| $\lambda = 10^{-5}$ | 3 | 99.77 | 99.83 | 99.86 | 98.42 | 96.18 |
| | 5 | 99.65 | 99.72 | 99.70 | 97.57 | 94.99 |
| $\lambda = 10^{-4}$ | 3 | 99.83 | 99.84 | 99.86 | 98.41 | 96.24 |
| | 5 | 99.66 | 99.73 | 99.71 | 97.52 | 95.03 |
| $\lambda = 10^{-3}$ | 3 | 99.85 | 99.85 | 99.86 | 98.56 | 96.84 |
| | 5 | 99.71 | 99.74 | 99.75 | 97.26 | 95.74 |
| Local RX | 3 | 99.27 | 99.30 | **99.73** | 97.89 | 93.92 |
| | 5 | 98.01 | 98.21 | 99.27 | 98.35 | 90.36 |

where $\mathbf{L}$ is a low-rank matrix and $\mathbf{S}$ is a sparse matrix of entries with a small fraction of nonzero entries. $\mathbf{L}$ and $\mathbf{S}$ are solved by the following convex optimization problem:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1, \quad \text{s.t. } \mathbf{X} = \mathbf{L} + \mathbf{S} \qquad (16)$$

where $\| \cdot \|_*$ and $\| \cdot \|_1$ denote the trace-nuclear norm (sum of the singular values) and the sum of the absolute values
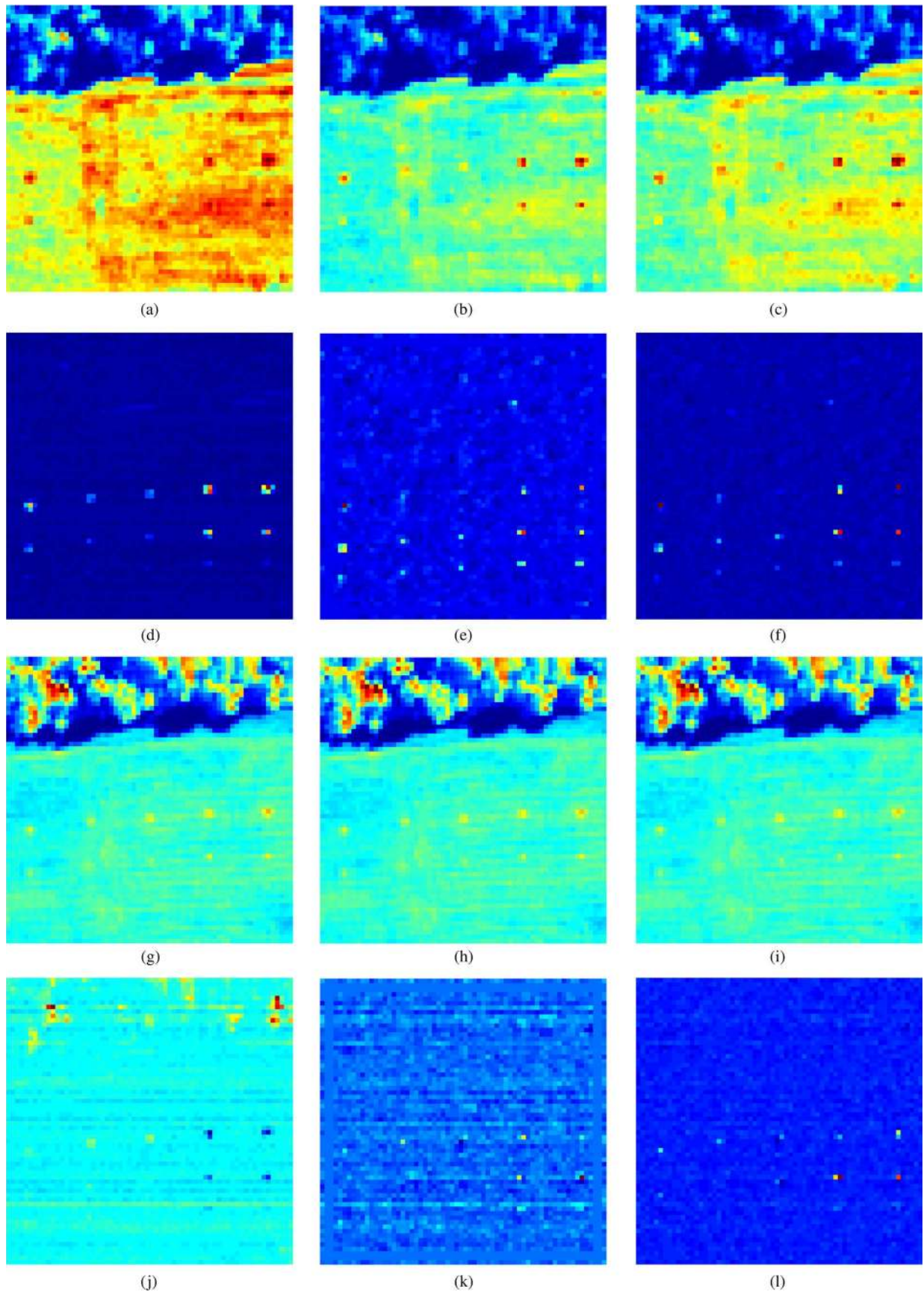
Fig. 8. For the HYDICE forest data, background prediction as well as residual images of the RPCA, SRD, and proposed CRD-DW-STO. (a) RPCA: background (band 100). (b) SRD: background (band 100). (c) CRD-DW-STO: background (band 100). (d) RPCA: residual (band 100). (e) SRD: residual (band 100). (f) CRD-DW-STO: residual (band 100). (g) RPCA: background (band 60). (h) SRD: background (band 60). (i) CRD-DW-STO: background (band 60). (j) RPCA: residual (band 60). (k) SRD: residual (band 60). (l) CRD-DW-STO: residual (band 60).
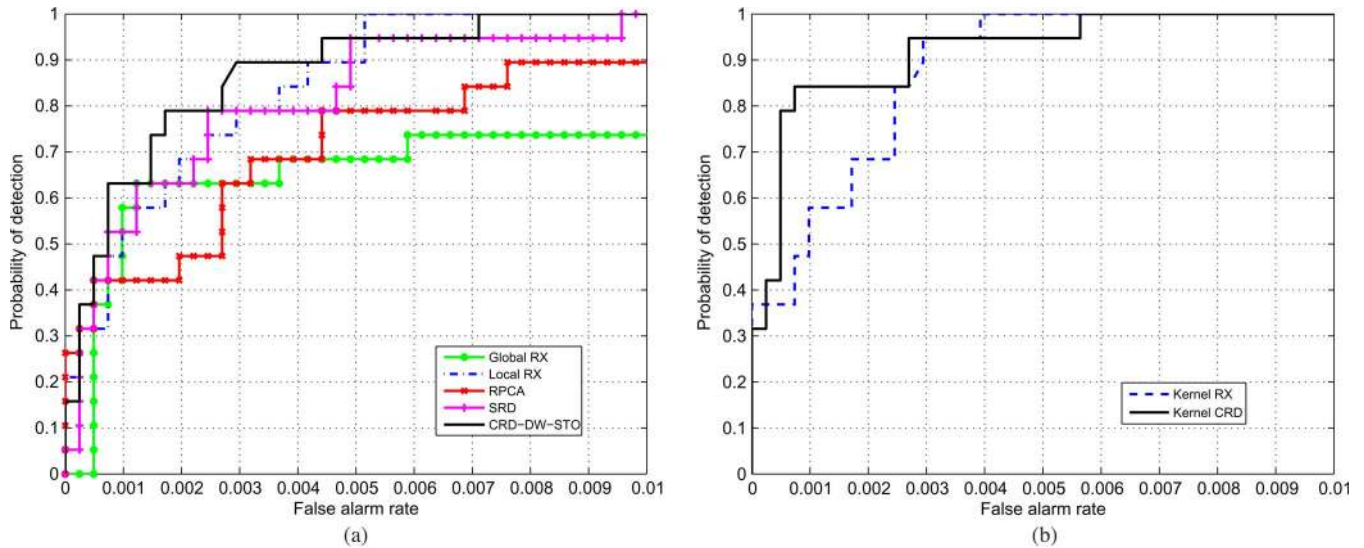
Fig. 9.    For the HYDICE forest data. (a) ROC evaluation of different anomaly detectors. (b) ROC evaluation of the kernel RX and kernel CRD.

of matrix entries, respectively, and $\lambda > 0$ is the regularization parameter. The optimization of (16) can be further solved by any off-the-shelf interior point solver after being reformulated as a semidefinite program. More detailed description of the optimization can be found in [28] and [29].

Unlike traditional PCA, RPCA can handle well the gross corruption of large noise (high magnitude). In RPCA, the trace-nuclear norm is for low-rank recovery, and the $\ell_1$-norm is for error correction as expressed in (16). Currently, RPCA has been employed to a variety of applications, such as removing the shadows of face images [28]. For the application of anomaly detection, $\mathbf{X}$ can be viewed as hyperspectral data with anomalous pixels, $\mathbf{L}$ is the prediction of the background image, and $\mathbf{S}$ is the residual (sparse) image which contains anomalies. RPCA-based anomaly detection can be achieved by calculating the Mahalanobis distance in the sparse matrix $\mathbf{S}$.

## III. EXPERIMENTAL RESULTS

### A. Hyperspectral Data

Four different hyperspectral data sets are used to evaluate the effectiveness of the proposed algorithm in the task of detecting anomalies. The first hyperspectral data set was acquired by the HYDICE image sensor [40] depicted in Fig. 2(a). This forest scene consists of size $64 \times 64$ pixels with 210 spectral bands of spectral coverage 0.4–2.5 $\mu$m for a forest area. In the experiments, a total of 169 bands are used after removal of water-absorption bands with a spatial resolution of 1.56 m. The ground-truth map includes 19 anomalous pixels corresponding to 15 target panels covered with different paints as illustrated in Fig. 2(b). Fig. 2(c) shows the spectral signatures of two randomly chosen background (normal) pixels and one anomalous pixel.

The second data were also collected by HYDICE airborne sensor [21]. This urban scene consists of $80 \times 100$ pixels for an urban area. The spatial resolution is approximately 1 m. Here, 175 bands remain after removal of water vapor absorption bands. There are approximately 21 anomalous pixels, repre-

senting cars and roof. The scene and the ground-truth map of anomalies are illustrated in Fig. 3.

The third data set[2] was acquired by the HyMap airborne hyperspectral imaging sensor, which provides 126 spectral bands spanning the wavelength interval of 0.4–2.5 $\mu$m. The image data set, covering one area of Cooke City, MT, USA, was collected on July 4, 2006, with the spatial size of $200 \times 800$. Each pixel has approximately 3 m of ground resolution. Seven types of targets, including four fabric panel targets and three vehicle targets, were deployed in the region of interest. In our experiment, we crop a subimage of size $100 \times 300$, including all of these targets (anomalies) as depicted in Fig. 4.

The fourth data set[3] was obtained from the Airborne Visible/Infrared Imaging Spectrometer covering the Moffett Field, CA, USA, at the southern end of the San Francisco Bay on August 20, 1992. This scene consists of $512 \times 512$ pixels [as shown in Fig. 5(a)] with 224 bands spanning the wavelength interval of 0.4–2.5 $\mu$m. The spatial resolution is approximately 20 m. Here, we implement global RX to find the anomalous pixels [41] in the scene [as shown in Fig. 5(b)], where 59 anomalies are identified when the threshold is set to 0.5 in Fig. 5(c), which are used as ground truth for local detectors.

### B. Detection Performance

In this section, we investigate the anomaly detection accuracy of the proposed collaborative-representation-based algorithms. The representation-based methods (and their acronyms) investigated in the experiments are summarized as follows:

1) representation without $\ell_1$-norm or $\ell_2$-norm minimization as RD;
2) representation with $\ell_2$-norm minimization as CRD;
3) representation with $\ell_2$-norm minimization and distance-weighted regularization matrix as CRD-DW;

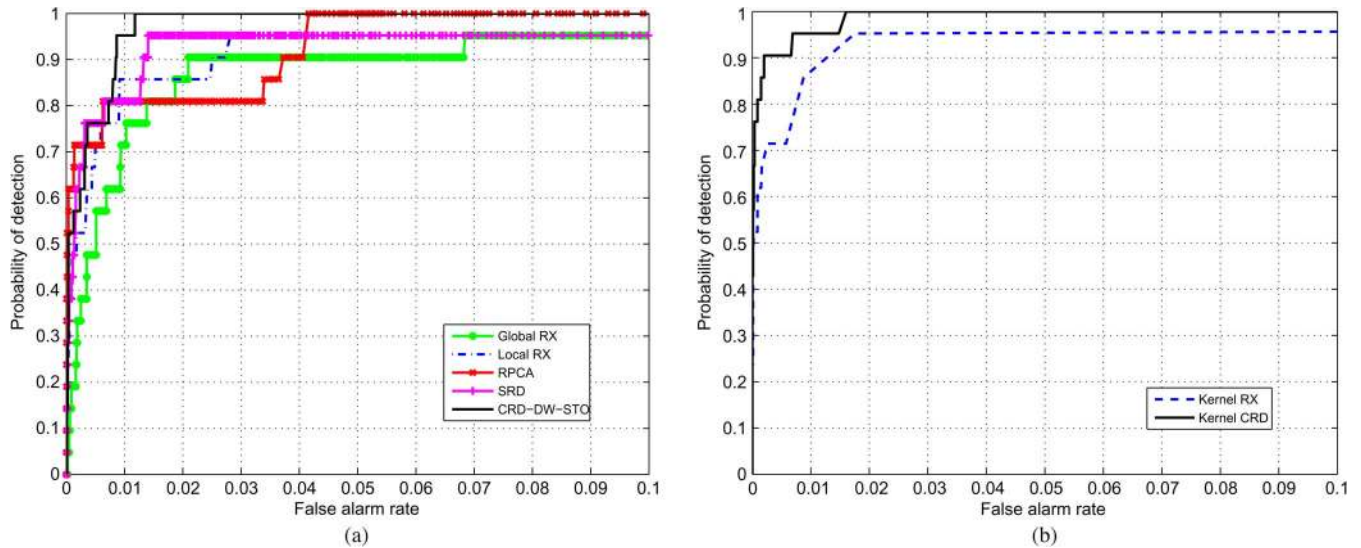[2]http://dirsapps.cis.rit.edu/blindtest/.
[3]http://aviris.jpl.nasa.gov/.

Fig. 10. For the HYDICE urban data. (a) ROC evaluation of different anomaly detectors. (b) ROC evaluation of the kernel RX and kernel CRD.

4) representation with $\ell_2$-norm minimization and distance-weighted regularization matrix and sum-to-one constraint as CRD-DW-STO.

First, we illustrate the effects of the regularization term and window size using the first experimental data. The receiver operating characteristic (ROC) [42] curves are employed. Fig. 6 shows the performance when the outer window size $w_{\text{out}} = 11$, the inner window size $w_{\text{in}} = 3$, and the regularization parameter is chosen as small as $\lambda = 10^{-6}$ for CRD, CRD-DW, and CRD-DW-STO. It shows that CRD-DW-STO outperforms all other detectors, which is expected because the former adopts the distance-weighting measurement, providing more flexibility to obtain the weights. Compared to CRD-DW, CRD-DW-STO provides more stable weight vectors, especially in heterogeneous areas. Fig. 7 further shows the parameters of window size $(w_{\text{out}}, w_{\text{in}})$ (e.g., (11, 3), (11, 5), etc.) for the proposed CRD-DW-STO as well as the local RX algorithm. For optimal parameter selection, we further compute the area under the ROC curve (AUC) to evaluate the performance of the proposed CRD-DW-STO and local RX with varying window size $(w_{\text{out}}, w_{\text{in}})$ and $\lambda$ as listed in Table I. From the results, it is obvious to observe that the detection performance is insensitive to the regularization parameter but sensitive to window size. The best window size for this experiment is (11, 3). Thus, the regularization parameter of the proposed CRD-DW-STO is fixed to $\lambda = 10^{-6}$ in the following experiments.

Then, we compare the estimated background images using the RPCA, the SRD method (the $\ell_1$-minimization[4] is implemented by ll_ls.m), and the proposed CRD-DW-STO as shown in Fig. 8. For CRD-DW-STO, we employ the same parameter setting as the previous experiment. The parameters, such as $\lambda$ for both RPCA and SRD, are optimized in our experiment. For visual comparison, Fig. 8 illustrates the background estimates as well as the residual images using these different approaches for the HYDICE forest data set. From the comparative results, it is evident that, for the 100th band,

[4]http://www.stanford.edu/boyd/software.html.

### TABLE II
AUC (IN PERCENT) PERFORMANCE OF THE PROPOSED CRD-DW-STO AND LOCAL RX WITH VARYING WINDOW SIZE ($w_{\text{out}}, w_{\text{in}}$) AS WELL AS $\lambda$ FOR THE HYDICE URBAN DATA

| | | $w_{out}$ | | | |
|---|---|---|---|---|---|
| | $w_{in}$ | 13 | 15 | 17 | 19 |
| $\lambda = 10^{-6}$ | 5 | 99.61 | 99.61 | 99.51 | 99.37 |
| | 7 | 99.56 | **99.69** | 99.49 | 99.47 |
| | 9 | 99.35 | 99.59 | 99.32 | 98.72 |
| $\lambda = 10^{-5}$ | 5 | 99.61 | 99.61 | 99.51 | 99.34 |
| | 7 | 99.56 | 99.69 | 99.49 | 99.47 |
| | 9 | 99.36 | 99.59 | 99.32 | 98.73 |
| $\lambda = 10^{-4}$ | 5 | 99.61 | 99.61 | 99.51 | 99.34 |
| | 7 | 99.56 | 99.69 | 99.48 | 99.47 |
| | 9 | 99.32 | 99.59 | 99.32 | 98.73 |
| $\lambda = 10^{-3}$ | 5 | 99.61 | 99.61 | 99.51 | 99.34 |
| | 7 | 99.55 | 99.69 | 99.48 | 99.47 |
| | 9 | 99.34 | 99.56 | 99.31 | 98.73 |
| Local RX | 5 | 94.16 | 94.63 | 93.32 | 92.93 |
| | 7 | 94.62 | **94.93** | 92.03 | 91.47 |
| | 9 | 94.34 | 94.36 | 92.73 | 91.29 |

### TABLE III
AUC (IN PERCENT) PERFORMANCE OF THE PROPOSED CRD-DW-STO AND LOCAL RX WITH VARYING WINDOW SIZE ($w_{\text{out}}, w_{\text{in}}$) FOR THE HYMAP IMAGE DATA (VIEWING "F1" TARGET AS ANOMALIES)

| | | $w_{out}$ | | | |
|---|---|---|---|---|---|
| | $w_{in}$ | 13 | 15 | 17 | 19 |
| $\lambda = 10^{-6}$ | 5 | 74.25 | 78.22 | 77.73 | 76.06 |
| | 7 | 79.68 | 81.61 | 81.06 | 79.40 |
| | 9 | 90.73 | **91.72** | 91.55 | 90.20 |
| | 11 | 87.74 | 90.47 | 88.95 | 87.48 |
| Local RX | 5 | 73.49 | 75.45 | 77.72 | 76.08 |
| | 7 | 75.52 | 78.99 | 81.34 | 79.01 |
| | 9 | 85.05 | 87.91 | **89.67** | 88.84 |
| | 11 | 82.83 | 84.67 | 86.94 | 86.03 |

the RPCA, SRD, and CRD-DW-STO methods exhibit similar performance; however, for the 60th band, the estimated residual images of RPCA and SRD are not as good as that of CRD-DW-STO. Specifically, our proposed CRD-DW-STO method can better suppress background in the upper area of the scene.

Next, we evaluate the detection performance of our proposed CRD-DW-STO detector, comparing with conventional global

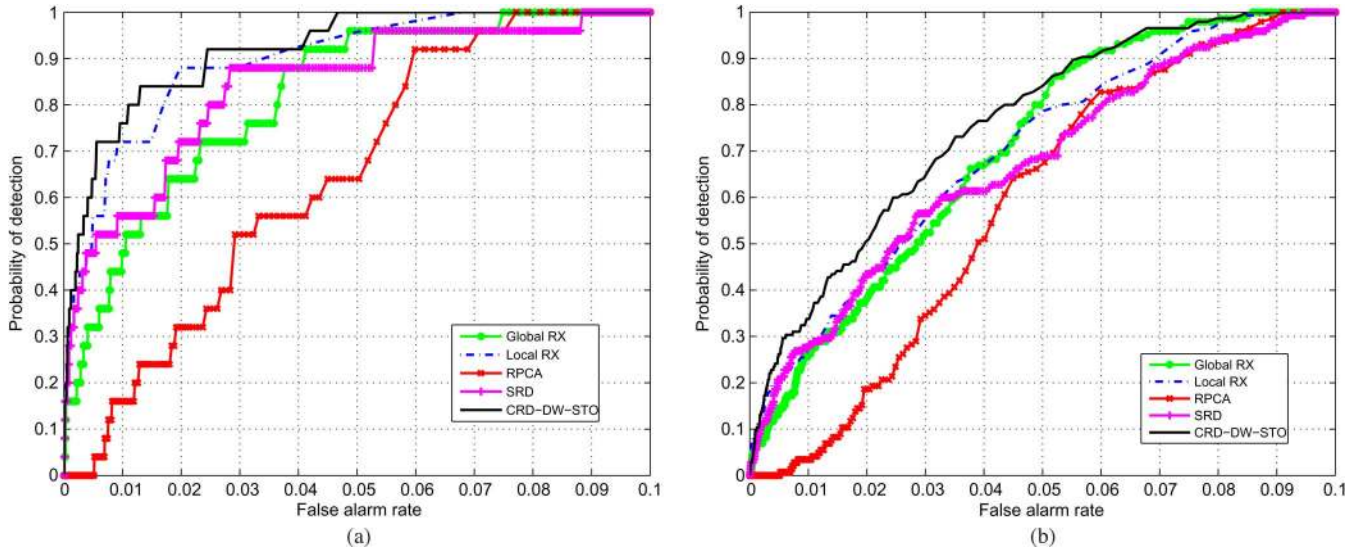Fig. 11. For the HyMap image data. (a) ROC evaluation of different anomaly detectors viewing "F1" target as anomalies. (b) ROC evaluation of different anomaly detectors viewing all of the targets as anomalies.

RX, local RX, RPCA, and SRD. According to empirical searching, for the HYDICE forest data, window size $(w_{out}, w_{in})$ is set to (11, 3). Fig. 9 illustrates their ROC curves for the HYDICE forest data set. In Fig. 9(a), an important observation is that the local RX outperforms the global RX and RPCA. Compared to the local RX, the proposed CRD-DW-STO exhibits a slightly lower probability of detection for a high false alarm rate (e.g., when the false alarm is 0.006–0.007); however, the overall detection performance of the proposed detector is still better. A nonlinear version of the proposed CRD (i.e., kernel CRD) has been further compared with the traditional kernel RX [22] as illustrated in Fig. 9(b). As indicated in [22], the value of parameter $\gamma$ is usually large for the RBF kernel, and in our work, the value of $\gamma$ is determined experimentally and is set to 20. We observe that the kernel CRD outperforms three other algorithms (i.e., the kernel RX, proposed CRD-DW-STO, and local RX), especially when the false alarm rate is from 0.0005 to 0.002. Furthermore, the AUC (in percent) values for these four detectors are 99.92, 99.87, 99.86, and 99.73, respectively.

For the HYDICE urban data, the detection performance is illustrated in Fig. 10. The proposed CRD-DW-STO exhibits a slightly lower probability of detection when the false alarm is lower than 0.005, while the AUC value is still the largest one among all of the detectors. According to Table II, the best performance of CRD-DW-STO can be achieved when $\lambda = 10^{-6}$ and the window size is (15, 7). The parameters of window size $(w_{out}, w_{in})$ are set to (15, 7) for the local RX and SRD after extensive searching. In Fig. 10(a), it is interesting to notice that, for the HYDICE forest data set, CRD-DW-STO has a lower probability of detection in the high false alarm rate region than the local RX; however, for the HYDICE urban scene, CRD-DW-STO has a lower probability of detection in the low false alarm rate region than RPCA and SRD but outperforms the local RX in general. The performance between the kernel CRD and the kernel RX is illustrated in Fig. 10(b). For these data, the value of kernel parameter $\gamma$ is set to 50. It is obvious that the kernel CRD outperforms CRD-DW-STO, the kernel RX
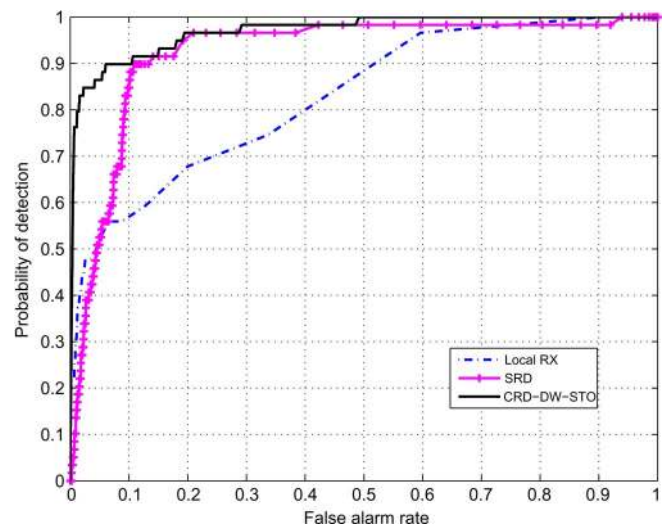


Fig. 12. For the Moffett Field data, ROC evaluation of different anomaly detectors.

outperforms the local RX, and the kernel CRD has the best detection performance. The AUC (in percent) values for these four detectors (i.e., the kernel CRD, kernel RX, proposed CRD-DW-STO, and local RX) are 99.87, 97.32, 99.69, and 94.93, respectively, which confirms that the proposed methods (i.e., CRD-DW-STO and kernel CRD) can outperform the traditional detectors.

For the third HyMap data set, anomaly detection is relatively more difficult. We evaluate the detection performance using the HyMap subimage (i.e., $100 \times 300$). According to Table III, the optimal window size is (15, 9) for CRD-DW-STO using these data. Here, the window size is set to (17, 9) for both local RX and SRD. The optimal $\lambda$ is $10^{-6}$. Fig. 11(a) illustrates the detection performances for the HyMap image viewing only "F1" target as anomalies. It can be seen that the proposed CRD-DW-STO offers a better result than all other detection methods; for example, in Fig. 11(a), when the false alarm rate

TABLE IV
AUC (IN PERCENT) PERFORMANCE OF THE PROPOSED CRD-DW-STO
AND LOCAL RX WITH VARYING WINDOW SIZE $(w_{out}, w_{in})$
FOR THE MOFFETT FIELD DATA

| | $w_{in}$ | $w_{out}$ | | | |
|---|---|---|---|---|---|
| | | 11 | 13 | 15 | 17 |
| $\lambda = 10^{-6}$ | 3 | 94.78 | 96.25 | **97.84** | 97.25 |
| | 5 | 93.57 | 95.94 | 97.72 | 97.10 |
| | 7 | 93.35 | 95.39 | 97.71 | 97.01 |
| | 9 | 90.07 | 93.84 | 97.56 | 96.42 |
| Local RX | 3 | 67.53 | 72.51 | 64.74 | 61.55 |
| | 5 | 74.09 | 77.79 | 72.46 | 65.41 |
| | 7 | 77.81 | 82.36 | 75.07 | 68.26 |
| | 9 | 79.47 | **82.82** | 78.01 | 72.48 |

TABLE V
EXECUTION TIMES (IN SECONDS) FOR ALL
OF THE EXPERIMENTAL DATA SETS

| | HYDICE Forest | HYDICE Urban | HyMap Image | Moffett Field |
|---|---|---|---|---|
| Global RX | 1.80 | 1.32 | 3.78 | – |
| Local RX | 142.43 | 228.83 | 480.16 | 2586.76 |
| RPCA | 14.58 | 14.44 | 38.72 | – |
| SRD | 1439.84 | 6701.62 | 6324.96 | 16734.73 |
| CRD-DW-STO | 62.84 | 324.23 | 375.46 | 2165.84 |

is 0.2, the probability of detection is 0.82, which is much higher than others. Fig. 11(b) further shows the detection performances for the same scene viewing all of the various targets as anomalies. Again, the proposed CRD-DW-STO apparently performs the best.

For the Moffett Field data, Fig. 12 illustrates the detection performance of three local detectors. The reason of only considering local detectors is that the ground-truth map is obtained by a global detector (i.e., global RX) [41]. According to Table IV, the best performance of CRD-DW-STO can be achieved when the window size is set to (15, 3). In this test, $\lambda$ is $10^{-6}$, and the window size is set to (13, 9) for the local RX and (15, 3) for SRD. In Fig. 12, the proposed CRD-DW-STO exhibits better performance than SRD and local RX. The AUC (in percent) values for these three detectors are 97.84, 92.32, and 82.82, respectively.

Finally, we report the computational complexity of the aforementioned detection methods. All experiments were carried out in MATLAB on an Intel Core 2 Duo CPU machine with 4 GB of RAM. The execution times for the experimental data are shown in Table V. At this point, it is important to notice that the local RX and CRD-DW-STO are computationally more expensive than the global RX as expected. Note that the downloaded RPCA code uses the MEX function which calls C program in MATLAB.

## IV. CONCLUSION

In this paper, we have proposed a novel collaborative-representation-based anomaly detection algorithm as well as its kernel version for a hyperspectral image. To estimate the background, each pixel is approximately represented via a linear combination of surrounding samples within a sliding dual window. The weight vector of combination, based on the distance-weighted Tikhonov regularization, has a closed-form solution under the $\ell_2$-norm minimization. The anomalies are calculated

from the residual image which is obtained by subtracting the predicted background from the original hyperspectral data. It is demonstrated that the proposed CRD-DW-STO and kernel CRD provide excellent detection performance with low computational cost and outperform other algorithms, such as the global RX, local RX, kernel RX, RPCA and SRD, for several real experimental data.

## REFERENCES

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

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

[3] 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.

[4] P. Bajorski, "Target detection under misspecified models in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 470–477, Apr. 2012.

[5] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[6] N. M. Nasrabadi, "Regularized spectral matched filter for target recognition in hyperspectral imagery," *IEEE Signal Process. Lett.*, vol. 15, no. 3, pp. 317–320, Mar. 2008.

[7] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[8] 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.

[9] J. Muñoz-Marí, F. Bovolo, L. Gomez-Chova, L. Bruzzone, and G. Camps-Valls, "Semisupervised one-class support vector machines for classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3188–3197, Aug. 2010.

[10] 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.

[11] 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.

[12] Z. Guo and S. Osher, "Template matching via $\ell_1$ minimization and its application to hyperspectral data," *Inv. Probl. Imaging*, vol. 5, no. 1, pp. 19–35, 2011.

[13] 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.

[14] O. Duran and M. Petrou, "A time-efficient method for anomaly detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3894–3904, Dec. 2007.

[15] J. M. Molero, E. M. Garzon, I. Garcia, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithms for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.

[16] S. Khazai, A. Safari, B. Mojaradi, and S. Homayouni, "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.

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

[18] 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.

[19] 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.

[20] 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.

[21] L. Ma, M. M. Crawford, and J. Tian, "Local manifold learning based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2010.

[22] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2010.

[23] 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.

[24] 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.

[25] 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.

[26] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 471–478.

[27] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, "Collaborative representation based classification for face recognition," *arXiv: 1204.2358*, 2012.

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

[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, Tech. Rep., Oct. 2010.

[30] B. Bao, G. Liu, C. Xu, and S. Yan, "Inductive robust principal component analysis," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3794–3800, Aug. 2012.

[31] W. Li and Q. Du, "Unsupervised nearest regularized subspace for anomaly detection in hyperspectral imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Melbourne, Vic., Australia, Jul. 2013, pp. 1055–1058.

[32] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2014.

[33] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC, USA: V. H. Winston & Sons, 1977.

[34] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

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

[36] W. Li, S. Prasad, and J. E. Fowler, "Decision fusion in kernel-induced spaces for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3399–3411, Jun. 2014.

[37] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 783 194-1–783 194-14, Mar. 2009.

[38] S. Prasad and L. M. Bruce, "Information fusion in kernel-induced spaces for robust subpixel hyperspectral ATR," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 572–576, Jul. 2009.

[39] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 894–898, Sep. 2011.

[40] C. I. Chang and S. S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.

[41] Q. Du, W. Zhu, and J. E. Fowler, "Anomaly-based JPEG2000 compression of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 696–700, Oct. 2008.

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

**Wei Li** (S'11–M'13) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-Sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Postdoctoral Researcher with the University of California, Davis, CA, USA. He is currently with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. His research interests include statistical pattern recognition, hyperspectral image analysis, and data compression.

Dr. Li is an active Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

**Qian Du** (S'98–M'00–SM'05) received the Ph.D. degree in electrical engineering from the University of Maryland Baltimore County, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. Her research interests include hyperspectral remote sensing image analysis, pattern classification, data compression, and neural networks.

Dr. Du served as the Cochair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society (2009–2013). She is currently the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition. She is a member of SPIE, ASPRS, and ASEE. She also serves as an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Applied Remote Sensing*. She received the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was the General Chair of the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) in Shanghai, China, in 2012.