



# Hyperspectral anomaly detection using ensemble and robust collaborative representation

Shaoxi Wang<sup>a</sup>, Xintao Hu<sup>a</sup>, Jialong Sun<sup>b</sup>, Jinzhuo Liu<sup>c,d,e,\*</sup>

<sup>a</sup> School of Microelectronics, Northwestern Polytechnical University, Xi'an 710072, China

<sup>b</sup> School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

<sup>c</sup> School of Software, Yunnan University, Kunming 650106, China

<sup>d</sup> Key Laboratory in Software Engineering of Yunnan Province, Kunming 650106, China

<sup>e</sup> Engineering Research Center of Cyberspace, Yunnan University, Kunming 650106, China

## ARTICLE INFO

### Article history:

Received 12 December 2020

Received in revised form 7 October 2022

Accepted 29 December 2022

Available online 2 January 2023

### Keywords:

Anomaly detection (AD)  
collaborative representation  
hyperspectral images (HSIs)  
machine learning

## ABSTRACT

In this paper, we propose a novel ensemble and robust anomaly detection method based on collaborative representation-based detector. The focused pixels used to estimate the background data are randomly sampled from the image. To soften the outliers' contributions among the selected pixels, we assign low weights to the outliers by adopting a robust norm regression. Consequently, the estimation result is less sensitive to the presence of outliers, as the experiment results attest. However, the algorithm performance is unstable due to the randomness of pixel sampling. To eliminate the instability and boost the detection performance, an ensemble learning method is employed. We repeat modeling background based on random pixel selection, and the detection result is an ensemble of all batches. We show that in most datasets, the proposed method outperforms the traditional algorithms. Moreover, batch processes for detection boosting secure future advances in performance utilization with parallel computing applied.

© 2023 Elsevier Inc. All rights reserved.

## 1. Introduction

Hyperspectral image (HSI) has come to the fore in the past three decades, due to its application in distinguishing the anomalies from the natural ground objects spectrally. Hyperspectral images are captured by hyperspectral detectors, which contain both the 2D geometry spatial image and spectral information with high spectral resolution. Different ground objects possess unique characteristics, which are reflected in the corresponding spectral information. Therefore, the spectral information can be regarded as the signature of the ground object and used to distinguish a certain kind of ground objects from the others, making the hyperspectral image an intelligence resource of primary importance. The hyperspectral image's significance has been appreciated in various remote sensing applications, such as object classification in [1–5], hyperspectral unmixing in [6,7], anomaly detection in [8], change detection in [9,10], et cetera. One of the most renowned and widely used applications is anomaly detection, a method to separate the objects of interest from the background, which are usually rare in the data. Anomaly detection has been the principal subject studied in system health monitoring, target detection, and fault detection, to name a few.

\* Corresponding author.

E-mail addresses: [shxwang@nwpu.edu.cn](mailto:shxwang@nwpu.edu.cn) (S. Wang), [hxt\\_tg@mail.nwpu.edu.cn](mailto:hxt_tg@mail.nwpu.edu.cn) (X. Hu), [jlsun@mail.nwpu.edu.cn](mailto:jlsun@mail.nwpu.edu.cn) (J. Sun), [jinzhuo.liu@hotmail.com](mailto:jinzhuo.liu@hotmail.com) (J. Liu).

Efforts have been made to propose an anomaly detection method with high performance and strong robustness, ensuring that the algorithm functions correctly in the presence of outliers, i.e. noises that disturb the performance of anomaly detection. Various robustness enhancement approaches have been made based on traditional anomaly detection algorithms, such as adaptive Reed-Xiaoli (RX) [11], compressed columnwise robust principal component analysis (CCRPCA) for hyperspectral anomaly detection [12], low-rank representation sum-to-one (LRRSTO) [13] and robust nonlinear anomaly detection (RNAD) [14]. However, the time consumption of these algorithms is either longer than the traditional ones or not mentioned in their works. Some related work brings new perspectives and novel approaches, such as background joint sparse representation (BJSR) [15], graph-based anomaly detection in [16], detection algorithm via spectral unmixing and dictionary-based low-rank decomposition in [17], and anomaly detection algorithm based on background estimation and adaptive weighted sparse representation proposed in [18].

Intrigued to develop an efficient and robust anomaly detection algorithm that retains low time consumption and complexity, we employ an  $\ell_{2,1}$ -norm approach to enhance the detection robustness under the assumption that the background of the hyperspectral data is normally distributed. The introduction of  $\ell_{2,1}$ -norm regression is enlightened by [19], where Nie proposed an efficient algorithm to accelerate the weight matrix solving progress of joint  $\ell_{2,1}$ -norm regression problems.

By introducing  $\ell_{2,1}$ -norm regression and random sampling strategy into background modeling, we successfully improved the robustness of the traditional CRD algorithm proposed in [20] with complexity and time consumption maintained at a lower level and performance improved, as the experiment results derived from five real hyperspectral datasets attest. The proposed method can efficiently enhance the robustness and accuracy of hyperspectral anomaly detection in the presence of outliers.

The rest of this paper is structured as follows. In the second part, we review related work in the field of anomaly detection. A detailed description of the proposed robust anomaly detection algorithm based on  $\ell_{2,1}$ -norm regression and random background modeling are illustrated in the third part. Dataset description, experiment results, and conclusions are shown in the following parts.

## 2. Related Work

Global/Local RX (GRX/LRX) and collaborative representation-based detector (CRD) are milestones in hyperspectral anomaly detection. RX algorithms were introduced in [21], in which they proposed convincing theoretical and practical reasons in terms of the general likelihood ratio test (GLRT). It carried out an assumption that background pixels and anomaly pixels follow Gaussian distributions with the same covariance but with different means. Each pixel is tested respectively to judge whether it follows the background's distribution. A pixel that deviates from the background's distribution could be considered an anomaly.

CRD has been a popular method to detect anomalies. It exhibits the local characteristics by the frame of two windows. Each surrounding pixel between two shifting windows, whose matrix is denoted as  $\mathbf{X}_w$ , is sampled to estimate the central pixel  $\mathbf{y}$  to be measured. The original version employs a ridge regression in order to find a weight vector  $\alpha$ :

$$\arg \min_{\alpha} \|\mathbf{y} - \mathbf{X}_w \alpha\|_2^2 + \lambda \|\alpha\|_2. \quad (1)$$

Here  $\lambda$  controls the penalty of weight values. Then, taking derivative to  $\alpha$  and setting it to zero, we can calculate  $\alpha$  by

$$\hat{\alpha} = (\mathbf{X}_w^T \mathbf{X}_w + \lambda I)^{-1} \mathbf{X}_w^T \mathbf{y}. \quad (2)$$

Afterwards, it calculates Euclidean distance  $\|\mathbf{y} - \mathbf{X}_w \hat{\alpha}\|_2$  between the actual pixel ( $\mathbf{y}$ ) and collaborative represented pixel ( $\mathbf{X}_w \hat{\alpha}$ ) to determine anomalies.

There are challenges in applying CRD:

- The size of the outer window and the inner window is difficult to determine since it is related to the size of specific anomalous objects. If a dataset contains both large anomalies and small ones, a fixed window size may lead to poor performance.
- If the radiation intensity of some anomalies highly deviates from the mean, low bias anomalies may be misclassified.
- The estimation process above is applied to each pixel from the image, and the majority is to evaluate the matrix inverse, which has the complexity of  $O(N \times s^3)$ . Therefore, it is a challenge to apply CRD to a high-speed scene.

The proposed method employs random sampling to estimate the background in one-off calculating. In the next section, we will introduce its details.

## 3. Methods

Here, we propose a global collaborative representation to model the background pixels we choose randomly (as the Algorithm 1 shows). Then anomaly pixels could be found out by assigning each pixel to the trained model respectively. Our

method is outlined in Fig. 2. Here a 3-D hyperspectral image is denoted as  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ , in which  $p$  represents the number of image bands and  $n$  is the number of image pixels. (ERC RD<sup>1</sup> below stands for the method we proposed.).

### 3.1. Robust Collaborative Representation

For solving the objective problem, we introduce the  $\ell_{2,1}$ -norm. The  $\ell_{2,1}$ -norm involves a norm of a matrix  $\mathbf{W} \in \mathbb{R}^{m \times n}$  to

$$\|\mathbf{W}\|_{2,1} = \sum_{i=1}^m \sqrt{\sum_{j=1}^n \mathbf{w}_{ij}^2}. \quad (3)$$

Then take the derivative of  $\mathbf{W}_{ij}$  to Eq. 3, we have

$$\frac{\partial \|\mathbf{W}\|_{2,1}}{\partial \mathbf{W}_{ij}} = \mathbf{w}_{ij} \left( \sum_{j=1}^n \mathbf{w}_{ij}^2 \right)^{-1/2} = \frac{\mathbf{w}_{ij}}{\|\mathbf{W}^i\|_2}, \quad (4)$$

where  $\|\mathbf{W}^i\|_2$  is the  $\ell_2$ -norm of the  $i$ -th row vector of the matrix  $\mathbf{W}$ . Therefore, the Eq. 4 can be written as a matrix equation form

$$\frac{\partial \|\mathbf{W}\|_{2,1}}{\partial \mathbf{W}} = \Sigma \mathbf{W}, \quad (5)$$

where  $\Sigma \in \mathbb{R}^{d \times d}$  is a diagonal matrix and the  $i$ -th diagonal element  $\Sigma_{ii}$  equals to  $1/\|\mathbf{W}^i\|_2$ .

Anomalies are more likely to be the minority, for which the totality of pixels generally models the background objects in numerous algorithms. Different from the traditional CRD method of sampling pixel's neighborhood, we sample only a few pixels randomly as  $\mathbf{X}_s \in \mathbb{R}^{p \times s}$  from the whole image  $\mathbf{X}$ , in order to reduce the impact of anomalies' scale. Then a global background representation can be estimated with parameter matrix  $\mathbf{W}$  (of size  $s \times n$ ) via joint  $\ell_{2,1}$ -norm least square regression. The objective can be described as below:

$$\widehat{\mathbf{W}} = \arg \min_{\mathbf{W}} \|\mathbf{X} - \mathbf{X}_s \mathbf{W}\|_{2,1} + \lambda \|\mathbf{W}\|_{2,1}, \quad (6)$$

where  $\lambda$  is a Lagrange multiplier. The first term constraints the difference between the raw image and the estimated one. Meanwhile, since the  $\ell_{2,1}$ -norm has row rotational invariance, and also the weight of outliers in the  $\ell_{2,1}$ -norm is less than in the  $\ell_2$ -norm, background estimation becomes robust. Furthermore, the constraint of the second regularization term makes the weight matrix  $\mathbf{W}$  row-sparse and brings more attention to the key samples in  $\mathbf{X}_s$  during background estimation. This constraint can be interpreted as applying feature selection to the sampled pixels, which further improves robustness.

Consider the derivative to Eq. 6 of  $\mathbf{W}$  and set it to zero, we have

$$\mathbf{X}_s^T \mathbf{D} (\mathbf{X} - \mathbf{X}_s \mathbf{W}) + \lambda \mathbf{H} \mathbf{W} = 0. \quad (7)$$

In Eq. 7, the  $i$ -th element of diagonal matrix  $\mathbf{D}$  is  $1/\|\mathbf{X} - \mathbf{X}_s \mathbf{W}\|_2$  and the  $i$ -th element of diagonal matrix  $\mathbf{H}$  is  $1/\|\mathbf{W}\|_2$ . Therefore, the parameter matrix  $\mathbf{W}$  of collaborative background representation can be solved:

$$\widehat{\mathbf{W}} = (\mathbf{X}_s^T \mathbf{D} \mathbf{X}_s + \lambda \mathbf{H})^{-1} \mathbf{X}_s^T \mathbf{D} \mathbf{X}. \quad (8)$$

Then we have a background estimation  $\mathbf{X}_s \widehat{\mathbf{W}}$  which is insensitive to outliers.

### 3.2. Pixel Scores Boosting

We score the deviation from background for each pixel  $i$  by

$$\mathbf{S}(i) = \|\mathbf{X} - \mathbf{X}_s \widehat{\mathbf{W}}\|_2^{(i)}. \quad (9)$$

Due to the sampling randomness, the algorithm performance is unstable. The more background pixels are sampled, the better detection result we can get, as well as the fewer anomalous pixels. In most cases,  $\mathbf{X}_s$  contains only a few anomalies, and thence we can boost the detection performance by repeating random background modeling, scoring and calculating the average of tries:

$$\mathbf{S}(i) = \frac{1}{T} \sum_{t=1}^T \mathbf{S}_t(i). \quad (10)$$

As a result, a stable performance could be obtained.

<sup>1</sup> The source code of our method could be accessed at <https://github.com/hxt-tg/ensemble-and-robust-crd>.

## 4. Experiments Results

### 4.1. Verifying Methodology

All the implementations of both our novel methods and contradistinctive ones utilize the MATLAB R2019a, and run on the AMD Ryzen7 2700 CPU with parallel calculation disabled. We employ the receiver operating characteristic (ROC) curves [22] and the area under the ROC curve (AUC) to measure the performance of mentioned methods. If a detector is better than the other one, it lies upper than that one. Also, the corresponding AUC value will be larger.

---

**Algorithm 1.:** ERCRD algorithm procedure
 

---

**Data:** Image matrix  $\mathbf{X} \in \mathbb{R}^{p \times n}$ , number of sampling points  $s \in \mathbb{R}$ , number of repeats  $\mathbf{T} \in \mathbb{R}$

**Result:** Scoring vector  $\mathbf{S} \in \mathbb{R}^n$

Initialize scoring matrix  $\mathbf{S} \leftarrow \mathbf{0}$ ;

**for**  $r \leftarrow 1$  **to**  $\mathbf{T}$  **do** Ensemble

Randomly sample pixels  $\mathbf{X}_s \in \mathbb{R}^{p \times s}$  from image  $\mathbf{X}$ ;

Calculate the diagonal matrix  $\mathbf{D}$ , where the  $i$ -th element

$\mathbf{D}_i = 1/||(\mathbf{X} - \mathbf{X}_s \mathbf{W})^i||_2$ ;

Calculate the diagonal matrix  $\mathbf{H}$ , where the  $i$ -th element  $\mathbf{H}_i = 1/||\mathbf{W}^i||_2$ ;

Calculate the parameter matrix  $\widehat{\mathbf{W}} = (\mathbf{X}_s^\top \mathbf{D} \mathbf{X}_s + \lambda \mathbf{H})^{-1} \mathbf{X}_s^\top \mathbf{D} \mathbf{X}$ ;

Calculate the global background representation  $\mathbf{X}_s \widehat{\mathbf{W}}$ ;

**for**  $i \leftarrow 1$  **to**  $n$  **do**

Add the  $\ell_2$ -norm score of  $(\mathbf{X} - \mathbf{X}_s \widehat{\mathbf{W}})^i$  to  $\mathbf{S}(i)$ ;

**end**

**end**

Average scoring matrix  $\mathbf{S} \leftarrow \mathbf{S}/\mathbf{T}$ ;

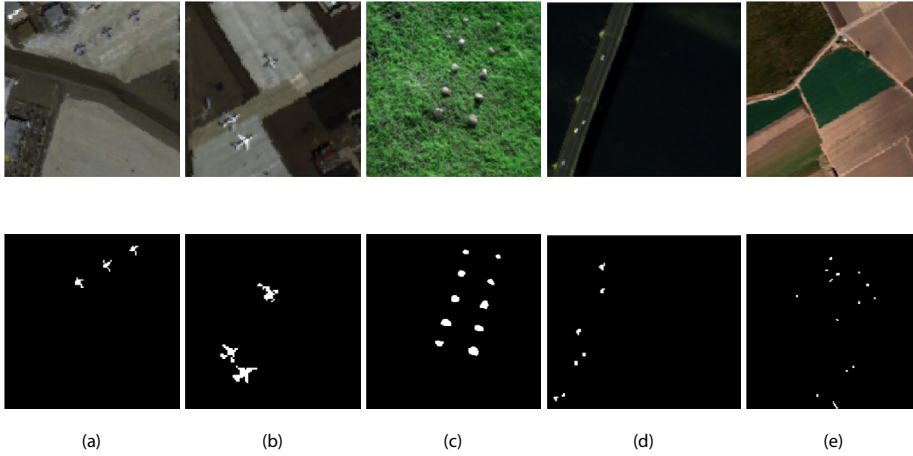
---

### 4.2. Dataset Descriptions

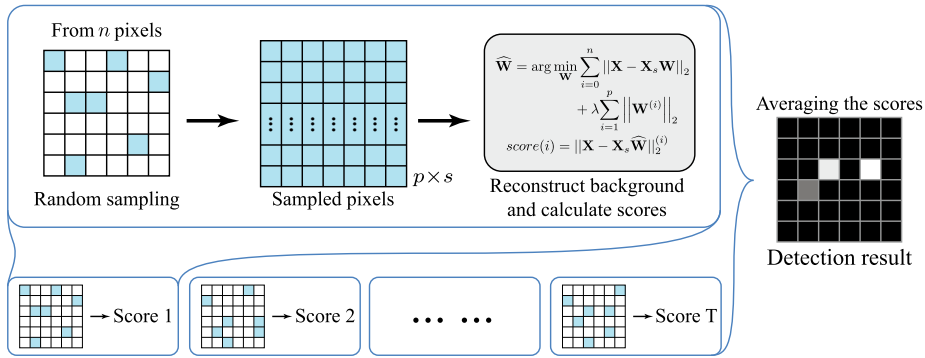
After analysis and comparison, five real HSI datasets, AVIRIS, San Diego airport, Cri, PaviaC and Salinas are used to validate the proposed method in this paper.

The first two hyperspectral datasets, denoted by AVIRIS and San Diego airport separately, are two different subimages generated from the San Diego Airport dataset, which is obtained in [23,24]. The San Diego Airport dataset is captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, covering a scene at the location of the airport in San Diego, CA, USA. It has a spatial resolution of 3.5m per pixel with 224 spectral channels with the wavelength ranging from 370 to 2510nm. After removing the bands corresponding to low signal-to-noise ratios and water absorption regions, 189 bands were retained in our experiment. Next, the AVIRIS dataset (shown in Fig. 1a) and the San Diego dataset (shown in Fig. 1b) are cropped, covering a scene of  $150 \times 150$  pixels and another one of  $100 \times 100$  pixels separately. In the AVIRIS dataset, three aircraft located at the upper side of the image are composed of 58 pixels and accounts for 0.25% in the whole image. Thus, the three aircraft are considered to be anomaly objects, as shown in Fig. 1a. Similarly, in the San Diego dataset, pixels of the three aircraft in the image are spectrally different, and thus the aircraft are regarded as objects of interest in anomaly detection, as shown in Fig. 1b.

The third dataset was collected by Nuance Cri sensor in [25], covering an area of  $400 \times 400$  with a spectral resolution of 10nm. It is composed of 46 spectral bands, and the bandwidths range from 0.65 to 1.1  $\mu\text{m}$ . As shown in Fig. 1c, the grass-like objects constitute the background, with ten rocks located in them, arranged in two columns. The rocks are comprised of 1254



**Fig. 1.** Pseudocolor image of data (a) AVIRIS. (b) San Diego airport. (c) Cri. (d) PaviaC. (e) Salinas.



**Fig. 2.** (This figure is updated.) Method Outline of ERCRD. The main idea is ensemble learning. In each epoch, pixels ( $s$  column vectors) are sampled randomly to construct a matrix of size  $p \times s$ . Then taking advantage of robust  $\ell_{2,1}$ -norm regression, we can estimate a global background weight matrix  $\mathbf{X}, \hat{\mathbf{W}}$  of size  $p \times n$ . We can acquire a scoring matrix in this epoch according to the Euclidean distance between the actual pixel  $\mathbf{X}$  and the estimated one  $\mathbf{X}, \hat{\mathbf{W}}$  at each pixel. By repeating the same procedure, we can calculate the average of these scores at each pixel.

pixels and account for 0.78% of the scene. Therefore, the rocks in the scene are identified to be anomaly objects in the experiments, as illustrated in Fig. 1c.

The fourth hyperspectral dataset used in this paper is cropped from the PaviaC dataset, collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor [26], which has been widely used in related publications. It was captured over the city center of Pavia, northern Italy. The original PaviaC dataset contains images of  $1096 \times 1096$  pixels, with noise bands included. A subimage of  $108 \times 120$  pixels is selected after removing low signal-to-noise ratio (SNR) bands, with 102 spectral bands ranging the spectrum of  $430$  to  $860nm$ . The spatial resolution is  $1.3$  meters per pixel. As shown in Fig. 1d, the background is formed by bridge and water, and the vehicles on the bridge and the bare soil near the bridge are considered anomalies, as they cover 43 pixels, accounting for only 0.33% of the image, and are spectrally different from principle ground objects. The ground objects of the anomalies are shown in Fig. 1d.

Another dataset used is a subimage cropped from the Salinas dataset obtained in [27], which was captured by AVIRIS over Salinas Valley, CA, USA in 1998. The original dataset is composed of images of  $512 \times 217$  pixels with 224 spectral bands. The spatial resolution of the Salinas dataset is  $3.7m$  per pixel. After removing the water absorption and low-SNR bands in the original dataset, a subimage of  $180 \times 180$  pixels is cropped. Vegetables, bare soils, and vineyard fields form the background of this scene. Fig. 1e shows the ground objects of the anomalies, whose spectral signatures differ from those of main ground objects.

#### 4.3. Algorithm Performance

Concerning the parameter settings, we set the penalty of  $\ell_{2,1}$ -norm in experiments for each data. As described in [20], the choice of this parameter is always empirical. Here for the convenience of the comparison, we set  $\lambda = 10^{-6}$  in the traditional

CRD methods and ours. For the local methods, we choose the window sizes which correspond to the best performance in the relevant works. The random sampling process in the proposed method follows a uniform distribution.

First, we illustrate the effects of the sampling size  $s$  in a part of data sets. We use the box figures to reflect the statistic characteristics. The median of AUC values is plotted as the short center line in the box. The notches represent the confidence interval (CI) around the median. The upper boundary of the box is the 75th percentile (Q3), and the lower 25th percentile (Q1). The upper whisker extends to the last datum less than the 75th percentile over 1.5 times the height of the box (Q3 – Q1). So does the lower whisker. Beyond the whiskers, AUC values are considered outliers and are plotted as individual circles. The smaller the box is, the more stable results we get.

Fig. 3 shows the statistic range of area under ROC curves along with the increment of the random samples' size for PaviaC. According to the pseudocolor image of PaviaC in Fig. 1d, the background pixels contain a large area of water and one bridge with bare soil near its pier. The cars lying on the bridge are considered anomalies. The location of these pixels signifies that the bridge pixels constitute the background of car pixels in this situation. When the sampling size  $s$  is small, there is a low chance to sample the bridge for constructing the background. So the background estimation  $\mathbf{X}_s \mathbf{W}$  mainly contains the information of the water, and it is hard to tell apart the cars and the bridge correctly. With the increment of  $s$ , the bridge can be modeled and described as background, and the AUC value goes up. However, after  $s$  grows to a certain extent, the background is estimated by the anomalies, which makes the AUC value slightly falls. A similar phenomenon appears in the experiment for the Salinas data in Fig. 4. The difference lies in the number of background types. Salinas data contains several types of fields, including vegetables, bare soils, and vineyard fields. Thus, we need to sample more pixels in order to estimate all kinds of backgrounds integrally. There is an example that differs from the previous. Fig. 5 illustrates the experiment for the Cri data. As drawn in Fig. 1c, the background comprises grass-like objects only and can be readily modeled from a few sample pixels. When the sample size  $s$  increases, anomalies are far more likely to be sampled and then misjudged.

The time consumption of ERCRD mainly depends on the execution of computing the matrix inverse, especially computing the Moore–Penrose inverse. Hence, it is hard to discuss the variation of time consumption. But in most situations, with a certain band amount and sensor resolution, the complexity accords with a consistent law.

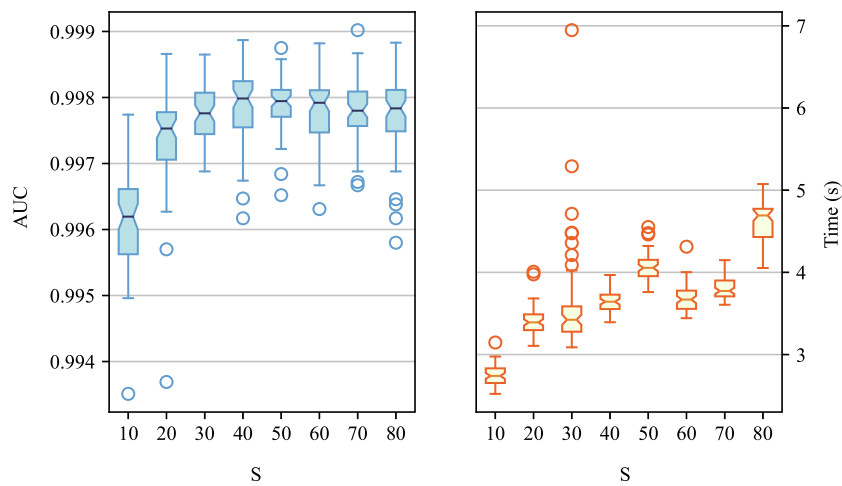
On the other hand, we have compared ERCRD with several local and global methods, including LRaSMD, LSMAD, global RX, local RX and CRD. For the Cri data, the amount of anomalies is not too small, and these pixels deviate a great deal from the background. We illustrate the detection performance in Fig. 6a. Associated with the legend in (a), the AUC (in percent) is 96.80, 91.34, 67.76, 96.21, 83.19 and 91.60, respectively. Also, the heat map in Fig. 6b depicts the scores (e.g., Euclidean distance of residuals, Mahalanobis distance, etc.) in each method. If a pixel has a high score, it will be lighter on the map. We can find that for this data, it is difficult to filter out the anomalies in the LRaSMD map and the local RX map. The anomalous pixels can be distinguished in the rest map. However, the scores of the anomalies in CRD and GRX are not higher than the background pixels, which is likely to cause misjudgement at the anomalies' boundary. The anomaly scores are high in LSMAD, but the entire background also possesses high scores. Thus, it is not robust enough for the detection procedure. ERCRD exhibits the best contrast between these objects and reflects the robustness in terms of the background modeling with outliers. In the heat map, the background part is dark, and the anomalies are light. Hence, the ERCRD estimated background is the closest to the original one among these methods and hardly affected by the outliers, which confirms that our methods can outperform the traditional detectors.

For PaviaC, as shown in Fig. 7, the AUC (in percent) is 99.73, 99.82, 88.90, 95.12, 91.17, 99.76, respectively. ERCRD, global RX and LSMAD perform better than the others. As we discussed earlier, the scores of the water background are distinctively lower than the anomalies since the water pixels stand a high chance of being sampled. CRD, LSMAD and GRX can also distinguish the anomalies from the water pixels. Moreover, the bridge scores in ERCRD are lower than those in the other methods, which is conducive to the robust detection procedure. For the other datasets, as the ROC plotted in Fig. 8, ERCRD outperforms the others for AVIRIS, the San Diego airport and the Salinas data as well.

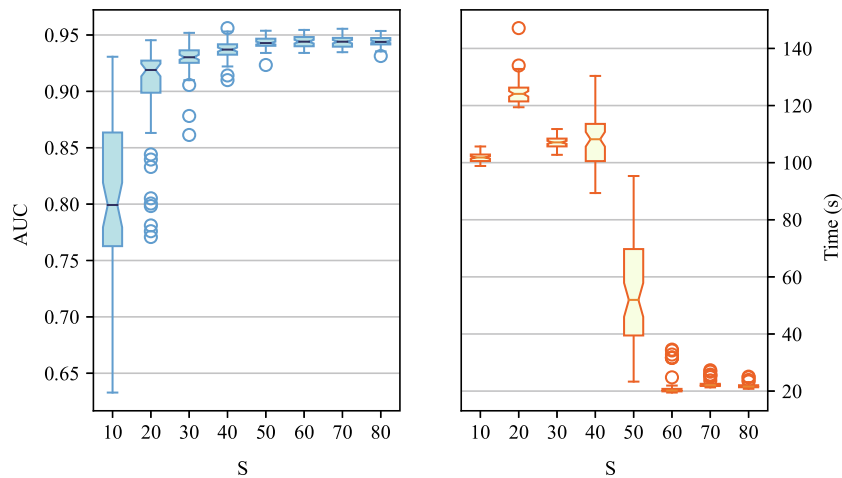
Finally, Table 1 summarizes the performance and time consumption of the six algorithms separately for all five data sets. Here we take the average AUC over 64 experiments as the result, as well as the time consumption. We find that global RX executes faster than the other methods in AVIRIS, San Diego airport, PaviaC and Salinas data. However, it does not perform well on some data. The running time is 6.03, 4.34, 3.57 and 21.39 s less than ERCRD. ERCRD executes faster than the global RX for the Cri data, which is the largest among the experimental data sets. For PaviaC, global RX has the highest AUC value and is so close to ERCRD. For the rest, ERCRD outperforms the other five methods significantly. Compared with the traditional CRD method, ERCRD surpasses in terms of both execution times and performance.

#### 4.4. Parameter Settings

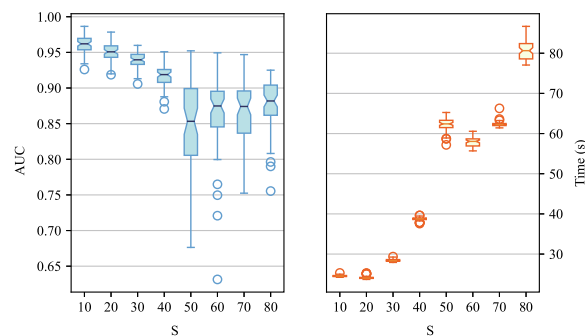
The parameter settings for main experiment (as shown in Fig. 6, 7 and 8) are shown in Table 2. Furthermore, we carry out an experiment on the effect of the ensemble size  $T$  on the performance of the proposed method. The ensemble size  $T$  is varied from 1 to 10 for the PaviaC data. Meanwhile, the other parameters are set as the corresponding optimal values. As Fig. 9 shows, the AUC increases with the growth of  $T$ , and the result would be more stable. Empirically,  $T = 10$  is large enough for most scenarios. For all the experiments in this paper, we set the ensemble size  $T$  as 10 for comparison.



**Fig. 3.** The statistic ranges of area under ROC curves (blue) and time consumptions (orange) for the PaviaC data. The proposed ERCRD algorithm repeats 64 times in order to calculate a AUC value, as well as to track the time consumptions. These box plots depict the 25th percentiles, median, 75th percentiles and outliers of the corresponding values.

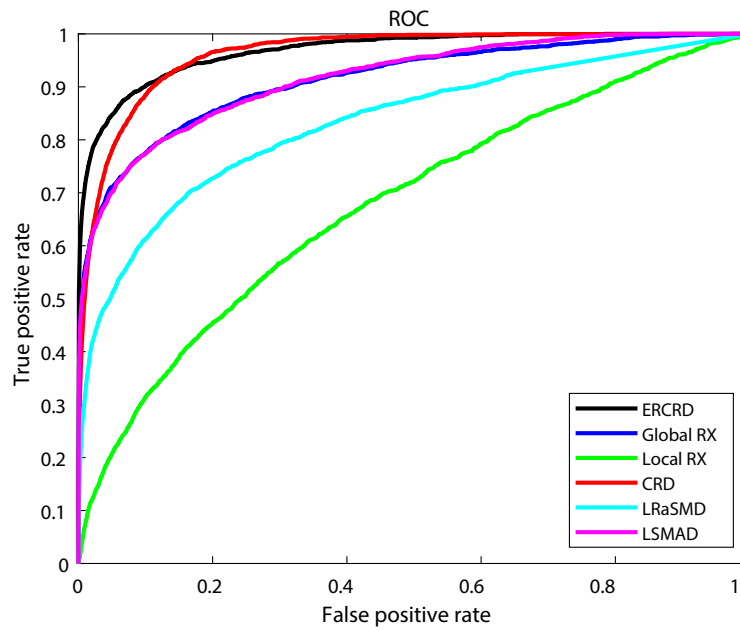


**Fig. 4.** The statistic ranges of area under ROC curves (blue) and time consumptions (orange) for the Salinas data.

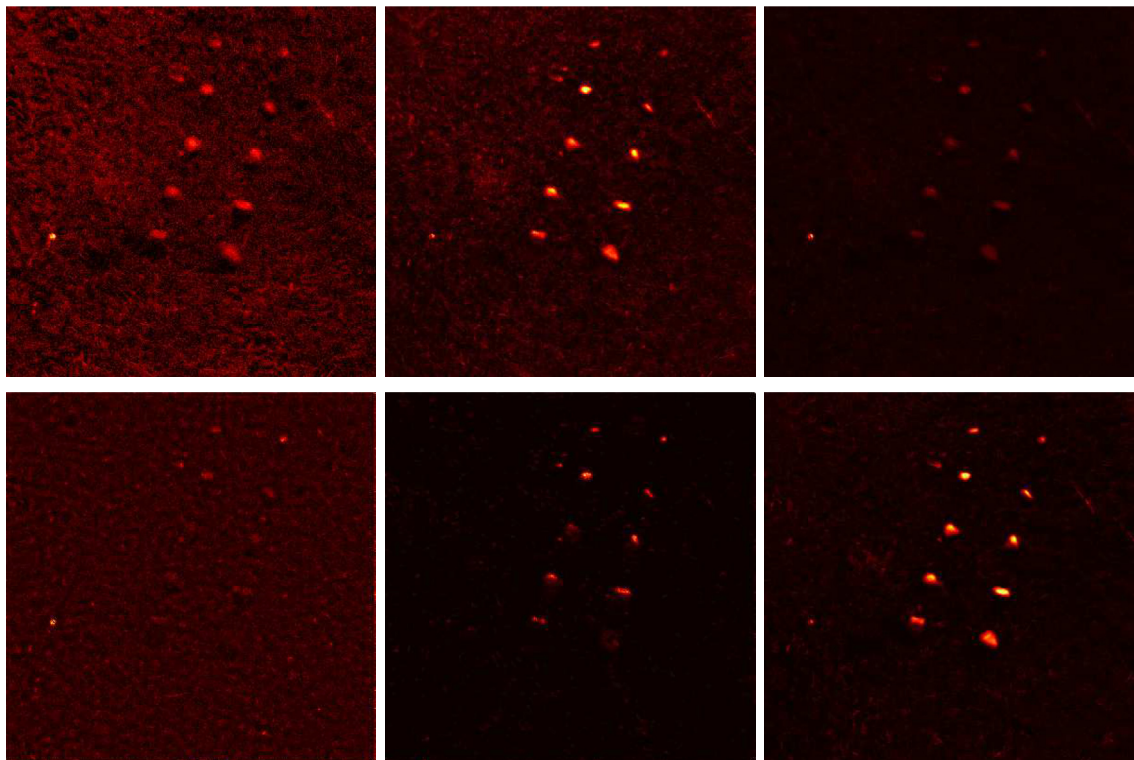


**Fig. 5.** The statistic ranges of area under ROC curves (blue) and time consumptions (orange) for the Cri data.





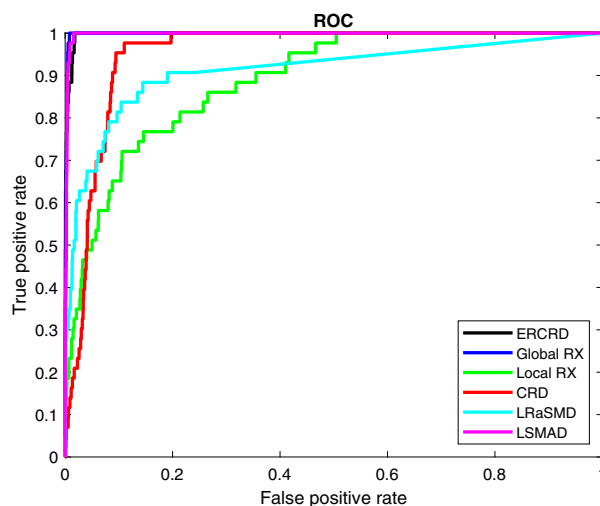
(a)



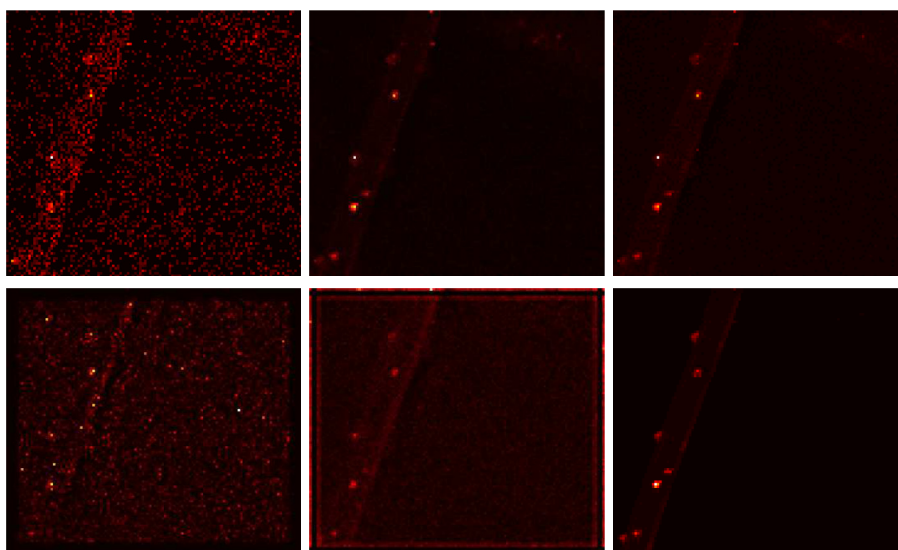
(b)

**Fig. 6.** For the Cri data, (a) ROC evaluation of different anomaly detection methods. (b) two-dimension figure of the detection scores. The employed methods in heat map is LRaSMD, LSMAD, GRX, LRX, CRD and ERCRD by rows, respectively.





(a)

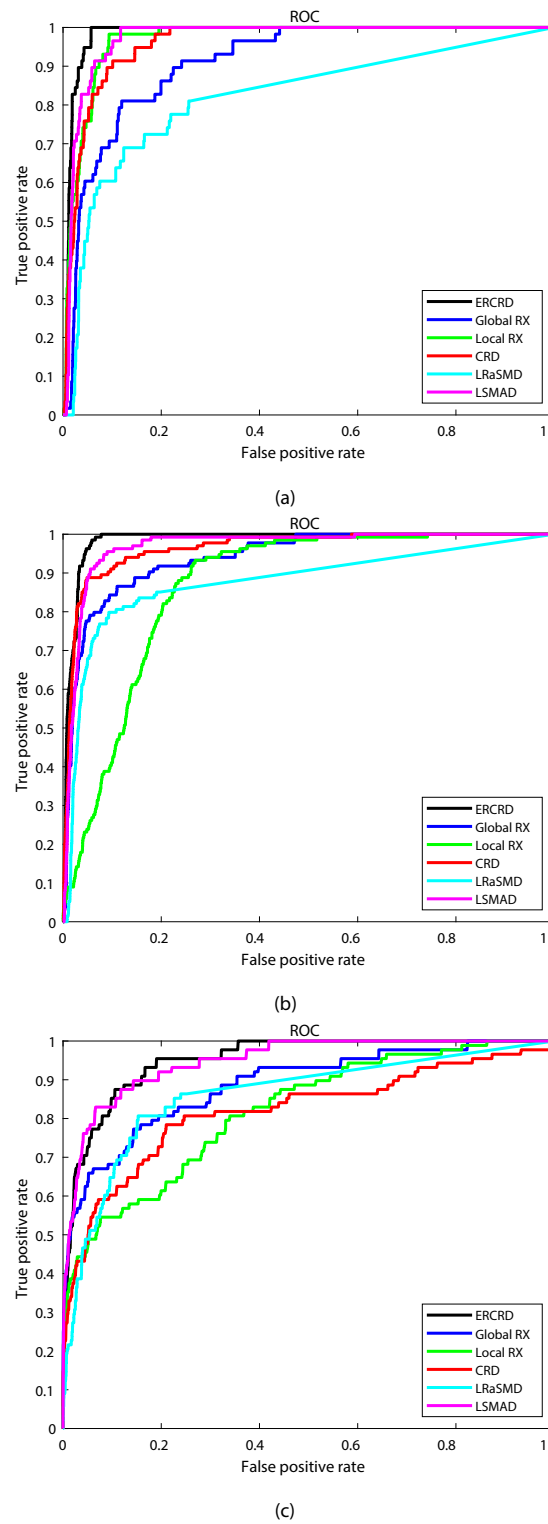


(b)

**Fig. 7.** For the PaviaC data, (a) ROC evaluation of different anomaly detection methods. (b) two-dimension figure of the detection scores. The employed methods in heat map is LRaSMD, LSMAD, GRX, LRX, CRD and ERCRD by rows, respectively.

## 5. Conclusions

In this paper, we propose a novel anomaly detection method based on CRD. In detail, a global background is estimated from randomly selected pixels via a joint  $\ell_{2,1}$ -norm regression, which has an iteration approach to be solved. The  $\ell_{2,1}$ -norm promises that outliers weigh less in regression. Hence, it reflects the robustness of estimation. Moreover, batch processes are applied for stable scoring and detection boosting, which stands a good chance of high-performance parallel computing optimization in future studies. We show that the robust random boosting CRD outperforms the classical algorithm, and the complexity maintains at a lower level. The future research we are striving for is to weight the boosting batches according to the statistical characteristics.



**Fig. 8.** ROC evaluation of different anomaly detection methods for the data (a) AVIRIS. (b) San Diego airport. (c) Salinas.

**Table 1**

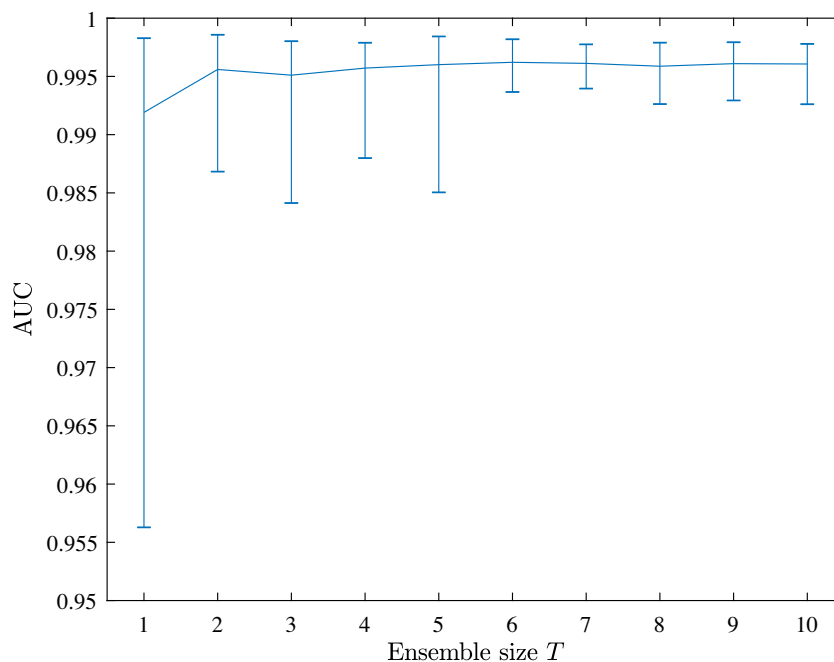
The AUC (in percent, top) performance and execution times (in seconds, bottom) for all of the experimental data sets including AVIris, SAN Diego airport, CRI, PAViaC and SALinas. The bold font highlights the highest AUC performance and the least execution time among the data sets. For local RX and CRD here, we choose 11 and 3 as the outer window size and the inner window size. In CRD, we set the penalty of the norm to  $10^{-6}$ . For comparison, the parameter of the penalty term in ERCRD is also  $10^{-6}$ .

Methods	AVI	SAN	CRI	PAV	SAL
LRaSMD	82.21	87.77	83.19	91.17	86.01
	13.53	10.17	50.84	9.07	36.08
LSMAD	97.17	96.84	91.60	99.76	94.72
	13.98	10.66	48.19	8.72	36.73
GRX	91.11	94.03	91.34	<b>99.82</b>	88.72
	<b>0.20</b>	<b>0.15</b>	<b>0.44</b>	<b>0.07</b>	<b>0.42</b>
LRX	96.85	86.56	67.76	88.90	81.88
	147.88	68.50	124.91	31.24	190.29
CRD	95.99	96.44	96.21	95.12	81.90
	147.62	17.96	556.77	34.14	22.03
ERCRD	<b>98.32</b>	<b>98.44</b>	<b>96.80</b>	99.73	<b>95.10</b>
	7.82	8.83	40.99	3.05	43.76

**Table 2**

The parameter settings for all the experimental datasets.

Methods	Parameters	Datasets				
		AVI	SAN	CRI	PAV	SAL
LRaSMD	Cardinality $k$	0.445	0.45	1.5	0.45	0.45
LSMAD	Cardinality $k$	0.005	0.005	0.015	0.005	0.005
GRX	–	–	–	–	–	–
LRX	$\text{win}_i$	$15 \times 15$	$15 \times 15$	$11 \times 11$	$5 \times 5$	$3 \times 3$
	$\text{win}_o$	$29 \times 29$	$17 \times 17$	$17 \times 17$	$11 \times 11$	$7 \times 7$
CRD	$\text{win}_i$	$13 \times 13$	$15 \times 15$	$11 \times 11$	$5 \times 5$	$3 \times 3$
	$\text{win}_o$	$19 \times 19$	$17 \times 17$	$17 \times 17$	$11 \times 11$	$7 \times 7$
	Regularization $\lambda$			$10^{-6}$		
ERCRD	Sample size $s$	10	10	10	10	60
	Ensemble size $T$			10		
	Regularization $\lambda$			$10^{-6}$		



**Fig. 9.** The impact of ensemble size  $T$  on performance. Each experiment of different ensemble size  $T$  is repeated over 64 times on data PaviaC. The upper dash of each line indicates the highest AUC value, so as the lower one.

## CRediT authorship contribution statement

**Shaoxi Wang:** Methodology, Software, Validation. **Xintao Hu:** Visualization. **Jialong Sun:** Validation. **Jinzhuo Liu:** Supervision, Writing - review & editing.

## Data availability

The authors do not have permission to share data.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This work was financially supported by Technology Development Program of Taicang (No.TC2019DYDS07), Technology Development Program of xi'an (No.201805042YD 20CG26), Technology Development Program of Shaanxi (No.2020GY-097) and National Natural Science Foundation of China (No. 61866039).

## References

- [1] L. Zhang, L. Zhang, B. Du, J. You, D. Tao, Hyperspectral image unsupervised classification by robust manifold matrix factorization, *Information Sciences* 485 (2019) 154–169.
- [2] Z. Zhu, S. Jia, S. He, Y. Sun, Z. Ji, L. Shen, Three-dimensional gabor feature extraction for hyperspectral imagery classification using a memetic framework, *Information Sciences* 298 (2015) 274–287.
- [3] J. Ham, Y. Chen, M.M. Crawford, J. Ghosh, Investigation of the random forest framework for classification of hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing* 43 (3) (2005) 492–501.
- [4] C. Shi, C.-M. Pun, 3d multi-resolution wavelet convolutional neural networks for hyperspectral image classification, *Information Sciences* 420 (2017) 49–65.
- [5] M. Khodadadzadeh, J. Li, A. Plaza, H. Ghassemian, J.M. Bioucas-Dias, X. Li, Spectral-spatial classification of hyperspectral data using local and global probabilities for mixed pixel characterization, *IEEE Transactions on Geoscience and Remote Sensing* 52 (10) (2014) 6298–6314.
- [6] X. Li, J. Huang, L.-J. Deng, T.-Z. Huang, Bilateral filter based total variation regularization for sparse hyperspectral image unmixing, *Information Sciences* 504 (2019) 334–353.
- [7] S. Jia, Y. Qian, Constrained nonnegative matrix factorization for hyperspectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing* 47 (1) (2008) 161–173.
- [8] J. Frontera-Pons, M.A. Veganzones, F. Pascal, J.-P. Ovarlez, Hyperspectral anomaly detectors using robust estimators, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (2) (2015) 720–731.
- [9] M.J. Carlotto, A cluster-based approach for detecting man-made objects and changes in imagery, *IEEE Transactions on Geoscience and Remote Sensing* 43 (2) (2005) 374–387.
- [10] Q. Wang, Z. Yuan, Q. Du, X. Li, Getnet: A general end-to-end 2-d CNN framework for hyperspectral image change detection, *IEEE Transactions on Geoscience and Remote Sensing* 57 (1) (2018) 3–13.
- [11] M.A. Veganzones, J. Frontera-Pons, F. Pascal, J. Chanussot, Binary partition trees-based robust adaptive hyperspectral RX anomaly detection, in: 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 5077–5081, <https://doi.org/10.1109/ICIP.2014.7026028>.
- [12] W. Sun, G. Yang, J. Li, D. Zhang, Hyperspectral anomaly detection using compressed columnwise robust principal component analysis, in: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 6372–6375. doi:10.1109/IGARSS.2018.8518817.
- [13] Y. Xu, Z. Wu, Z. Wei, H. Liu, X. Xu, A novel hyperspectral image anomaly detection method based on low rank representation, in: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015, pp. 4444–4447, <https://doi.org/10.1109/IGARSS.2015.7326813>.
- [14] R. Zhao, B. Du, L. Zhang, A robust nonlinear hyperspectral anomaly detection approach, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (4) (2014) 1227–1234, <https://doi.org/10.1109/JSTARS.2014.2311995>.
- [15] J. Li, H. Zhang, L. Zhang, L. Ma, Hyperspectral anomaly detection by the use of background joint sparse representation, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (6) (2015) 2523–2533.
- [16] Y. Yuan, D. Ma, Q. Wang, Hyperspectral anomaly detection by graph pixel selection, *IEEE transactions on cybernetics* 46 (12) (2015) 3123–3134.
- [17] Y. Qu, W. Wang, R. Guo, B. Ayhan, C. Kwan, S. Vance, H. Qi, Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition, *IEEE Transactions on Geoscience and Remote Sensing* 56 (8) (2018) 4391–4405.
- [18] L. Zhu, G. Wen, Hyperspectral anomaly detection via background estimation and adaptive weighted sparse representation, *Remote Sensing* 10 (2) (2018) 272.
- [19] F. Nie, H. Huang, X. Cai, C.H. Ding, Efficient and robust feature selection via joint  $\ell_{2,1}$  norms minimization, in: *Advances in neural information processing systems*, 2010, pp. 1813–1821.
- [20] W. Li, Q. Du, Collaborative representation for hyperspectral anomaly detection, *IEEE Transactions on geoscience and remote sensing* 53 (3) (2014) 1463–1474.
- [21] J.Y. Chen, I.S. Reed, A detection algorithm for optical targets in clutter, *IEEE Transactions on Aerospace and Electronic Systems* 1 (1987) 46–59.
- [22] J. Kerekes, Receiver operating characteristic curve confidence intervals and regions, *IEEE Geoscience and Remote Sensing Letters* 5 (2) (2008) 251–255.
- [23] Y. Zhang, B. Du, L. Zhang, Regularization framework for target detection in hyperspectral imagery, *IEEE Geoscience and Remote Sensing Letters* 11 (1) (2013) 313–317.
- [24] Z. Zhou, X. Li, J. Wright, E. Candes, Y. Ma, Stable principal component pursuit, in: 2010 IEEE international symposium on information theory, IEEE, 2010, pp. 1518–1522.
- [25] T. Zhou, D. Tao, X. Wu, Manifold elastic net: a unified framework for sparse dimension reduction, *Data Mining and Knowledge Discovery* 22 (3) (2011) 340–371.

- [26] N. Billor, A.S. Hadi, P.F. Velleman, Bacon: blocked adaptive computationally efficient outlier nominators, *Computational statistics & data analysis* 34 (3) (2000) 279–298.
- [27] Y. Yuan, Q. Wang, G. Zhu, Fast hyperspectral anomaly detection via high-order 2-d crossing filter, *IEEE Transactions on Geoscience and Remote Sensing* 53 (2) (2014) 620–630.