Hyperspectral Anomaly Detection based on Anomalous Component Extraction Framework

Shangzhen Song[a], Huixin Zhou[a,*], Jun Zhou[b], Kun Qian[a], Kuanhong Cheng[a], Zhe Zhang[a]

Abstract—**Anomaly detection has become an important topic in Hyperspectral Imagery (HSI) analysis in the last two decades with the advantage of detecting the targets surrounding in diverse backgrounds without prior knowledge. HSIs usually have complex and redundant spectral signals due to the complicated land-cover distribution. Generally, it is difficult to estimate the background accurately, and distinguish the anomaly targets. The performances of traditional algorithms are difficult to meet the requirements. In this paper, we propose a novel anomalous component extraction framework for hyperspectral anomaly detection based on Independent Component Analysis (ICA) and Orthogonal Subspace Projection (OSP). In the proposed method, the brightest anomalous component is extracted to initialize the projection vector, by which the performance of ICA can be improved greatly. Moreover, the Independent Component (IC) containing the most abnormal information can be obtained according to the vector. Besides, The OSP algorithm is applied to suppress the background components in the remaining data. Then the data are iteratively processed by ICA to extract the anomalous component subtly. Therefore, in the initialization process, the possible situation of detecting the pixels in the same position can be effectively avoided, and the interference of the last iteration procedure can be cut down greatly, helping to optimize the detection. Finally, the experimental results show that the proposed framework achieves a superior performance compared to some of the state-of-the-art methods in the field of anomaly detection.**

Keywords—**Anomaly detection, Hyperspectral image, Locally linear embedding, Independent component analysis, Orthogonal subspace projection.**

I. INTRODUCTION

HYPERSPECTRAL Imagery (HSI) technology provides 3-D images in spatial and spectral dimensions, which contain much more valuable information than that of 2-D images. With the increasing spectral and spatial resolution of hyperspectral sensors, it becomes popular to detect target using HSIs [1][2]. Hyperspectral target detection plays an important role in military and civilian applications. Due to the rich spectral information contained in the hyperspectral image, it is easier to perform target detection than that utilizing the visible light band. However, practically, the samples obtained is distinct from true data collected since known spectra usually comes from laboratory calibration. Moreover, it is not simple to obtain a complete target spectral information database. All these factors make it difficult for us to obtain the prior information of the target for detection, which leads to poor detection performance. Therefore, hyperspectral anomaly detection (AD), a target detection technique without priori information, is widely applied in practical situations [3]-[5].

Researches of anomaly detection in HSI have been developed for over two decades. Among them, the Reed-Xiaoli (RX) algorithm [6] is known as a benchmark in hyperspectral anomaly detection, which utilizes the multivariate Gaussian model to characterize the background pixels. The Mahalanobis distance between the test pixel and background mean vector is computed. An anomaly is obtained by comparing the distance and a threshold. However, the multivariate Gaussian distribution is not suitable to describe the complicated background in a real HSI [7]. As a result, a number of improved RX-based anomaly detection methods were proposed, such as the cluster-based anomaly detector (CBAD) [8], which utilizes a clustering technique to model the HSI as a multivariate normal distribution in each cluster and then executes the detection process. The weighted-RX method [9] calculates the mean vector and covariance matrix with the Gaussian probability estimation as the reweighted vector. The subspace RX (SSRX) [10] applies principal component analysis (PCA) on the HSI and performed the RX detector on a limited number of PCA bands. The above mentioned methods belong to linear detection methods, the performance of which are not outstanding due to the limitation of linear model. Therefore, nonlinear algorithms have been successfully applied into detection and other fields, such as the Kernel-RX (KRX) [11], the robust nonlinear anomaly detector [12], and the Kernel-PCA (KPCA) [13]. These kernel-based methods map the original data into a higher dimensional feature space, enhancing the separability between background and targets. In addition, there are other nonlinear

a: School of Physics and Optoelectronic Engineering, Xidian University, Xi'an, China

b: School of Information and Communication Technology, Griffith University, Brisbane, Australia

* Corresponding author.

E-mail address: hxzhou@mail.xidian.edu.cn (H.Zhou).

methods like support vector data description (SVDD) detector [14], and PE-AD detector [15]. Nonlinear anomaly detector can effectively handle the situation that the background region cannot meet the normal distribution assumption. Applying the kernel function to anomaly detection is able to adapt to the complex features in the data and achieve better results.

Despite the high spectral resolution delivering rich information, its high spectral dimension and complicated spectral correlation commonly have great effects on the anomaly detection performance when a detector is directly implemented on the original whole spectra. Consequently, dimensionality reduction algorithms, such as PCA [16], maximum noise fraction (MNF) [17], and independent component analysis (ICA) [18] [19] are effective preprocessing methods before anomaly detection. Among them, ICA is a multivariate signal analysis method, which extracts potential independent components from multivariate data according to statistical independence. It has been applied in information coding, neurobiology and other fields [20], [21]. But the classic ICA algorithm initialized the projection vector at random, which limited the performance seriously. Compared with linear methods above, manifold learning assumes that high dimensional data can be mapped to a low-dimensional manifold, which can represent the intrinsic structure of data, and maintain some of the important relationship while eliminating the redundancy, such as Isometric Feature Mapping (Isomap) [22]. Locally Linear Embedding (LLE) [23] [24], Laplacian Eigenmaps (LE) [25], and Local Tangent Space(LTSA) [26]. However, the disadvantage of dimensionality reduction before AD is that some potential but significant information hidden in the whole HSI may be lost. Besides, most AD algorithms cannot extract all the abnormal information hidden in the dataset sufficiently, which results in low detection rate while avoiding high false alarm rate. Therefore, it is a problem that how to further improve the detection efficiency.

In this paper, the key step to improve the AD performance is to construct a framework of refinement for abnormal components extraction, instead of simple stack. A widely used algorithm, orthogonal subspace projection (OSP) [27] plays a crucial role to achieve this goal. It solves the problem of quantitative analysis of hyperspectral mixed pixels, and has been successfully applied to classification, target detection and other occasions. From this perspective, an effective hyperspectral anomaly detection framework with anomalous component extraction based on ICA and orthogonal subspace projection (IOACED) is proposed. For an original HSI dataset, the IOACED technique is first applied with four main processes. Firstly, an initialized projection vector is obtained based on the LLE algorithm. Secondly, the final projection vector and Independent Component (IC)s are extracted by the Orthogonalized-ICA (OICA) algorithm. Thirdly, the value of the potential anomaly signal ratio (PASNR) [28] on each IC will be calculated and compared with a threshold. The components containing the most abnormal information are picked out and the remaining components that consist of a great deal of background will be suppressed through the OSP method. Meanwhile, by iteration, the useless information can be re-moved to the greatest extent. Finally, the AD output is generated after background suppression. The main advantages of our proposed method include: (1) The AD performance of ICA is improved by extracting the brightest anomalous component as the initialized projection vector, which is different from initializing it randomly in original ICA method. (2) The process of abnormal components extraction is refined by OSP iteratively, which makes the detection result more accurate.

The rest of this paper is organized as follows. Section II briefly reviews the basic theories of the related methodologies. Section III describes the proposed anomaly detection method in detail. The experiment results of the proposed on several HSI datasets are presented in Section IV. Finally, the conclusion is drawn in Section V.

II. BRIEF REVIEW OF THE RELATED METHODOLOGIES

In this section, we introduce the ICA and OSP theory briefly. Both of them play important roles in our proposed theory.

A. Independent Component Analysis

ICA assumes that there are several internal factors in observation data, which are independent from each other. Observation data can be denoted as the linear combination of these factors, which can be represented as:

$$Y = A \cdot X \tag{1}$$

where $X=[x_1, x_2,...,x_N]^T \in R^{N \times D}$ has N independent components, D denotes the number of bands in the data. $Y=[y_1, y_2,...,y_N]^T \in R^{M \times D}$ has M variables collected by the sensor, and $A \in R^{M \times N}$ is coefficient matrix. ICA is utilized to solve X when only Y is known, A projection matrix $W = [w_1, w_2,...,w_N] \in R^{M \times N}$ is attempted to maximize its proximity to A. Then X^*, the estimated value of X, can be represented as follows:

$$X^* = W^{-1}Y = W^{-1}AX \tag{2}$$

where W and A are approximately equal. Therefore, $W^{-1}A$ is closely equal to I. The ICs need to be solved by using statistics theory to describe the non-Gaussian characteristics of the data according to the central limit theorems [29]. Negentropy [29 30] is utilized to describe the non-Gaussian characteristics. The increase in negentropy indicates the enhancement of non-Gaussianity. Furthermore, for ICA, the whitening operation can improve the extraction efficiency of the ICs and decorrelate the variables [30]. After whitening, the observation data Y is denoted as Z. A single IC x_i can be expressed as $w_i^T Z$, $1 \le i \le N$ by (2). In the process of solving W, Aapo Hyvarinen proposed the Fast ICA algorithm [31] which has the advantages of simple computation, fast speed, and strong robustness. It can be written as:

$$w_i \leftarrow E[Zg(w_i^T Z)] - E[g(w_i^T Z)]w_i$$
$$w_i^* \leftarrow w_i / \|w_i\| \tag{3}$$

where w_i^*, $1 \leq i \leq N$ is the column vector of W after normalization. $g(\cdot)$ is the derivative of a non-quadratic function $G(\cdot)$. Equation (5) is exploited to iterate w_i when the convergence condition is reached. Then an approximate estimate of an IC x^* can be obtained by $w_i^T Z$.

During the process of solving ICs, orthogonalization is a serviceable method to remove the correlation. The specific steps for orthogonalized ICA(OICA) are

(i) Remove redundant bands, whiten the data, and determine the number of ICs N;
(ii) Initialize W, the norm of which should be equal to 1 meanwhile;
(iii) Update each column of W through (3);
(iv) Use the following formula to ensure that w_n is orthogonal to the previous n-1 w_i, $1 \leq i \leq n$-1, $i < n \leq N$;

$$w_n = w_n - \sum_{i=1}^{n-1}(w_n^T w_i)w_i \tag{4}$$

(v) Utilize (5) to determine whether the result of (3) con-verges. κ denotes a small positive number. If it does not converge, return to (iii) to repeat the iterative process.

$$\left| w_n^T w_n^* \right| < \kappa \tag{5}$$

(vi) After estimating w_n, continue to solve w_{n+1}, until w_N is solved.

B. Orthogonal Subspace Projection

It is comprehensible that a pixel of HSI always consists of several endmember spectrums, and different endmembers corresponding to each material, including background and outliers. Hence endmember vectors can be divided into two matrices: background characteristic matrix U and anomaly characteristic matrix M_T. The data vector y in HSI can be represented as

$$y = M_T \alpha_T + U \alpha_B + n \tag{6}$$

where $U=[u_1,u_2,...,u_b]$, $M_T=[m_1,m_2,...,m_t]$. u_i is an eigenvector orthogonal to each other. U represents the orthogonal space of u_i. α_T and α_B are abundance coefficients corresponding to two matrices respectively.

The background components in HSI can be suppressed by OSP. After U is determined with a kind of suitable statistical quantity, such as the spectrum of detected pixels [32], the projection matrix can be obtained as follows:

$$P_U^\perp = I - UU^T \tag{7}$$

where I denotes the unit matrix. Since u_i denotes an eigenvector orthogonal to each other, it is simple to get $PU=U-U^TU=0$. The pixel vector after being projected by OSP can be represented as

$$y_{OSP} = P_U^\perp y$$
$$= P_U^\perp M_T \alpha_T + P_U^\perp U \alpha_B + P_U^\perp n \tag{8}$$
$$= P_U^\perp M_T \alpha_T + P_U^\perp n$$

From the above equation, OSP can effectively remove the interference of some uninteresting eigenvectors (background) in HSI.

III. PROPOSED METHOD

This section gives the details of the proposed IOACED method. In the first part, an initialized vector selection strategy is introduced. In the second part, the evaluation for the content of anomaly information is described. In the third part, a refined anomaly component extraction method is given. Finally, an overview of the process is discussed.

A. Initialization Vector Selection

When ICA is exploited to extract features, we need to get an initial W, each column of which is substituted into (3) for iteration. It is obviously known that different initialized projection vector w_0 will result in different iteration performance, though they may converge by (5). In general, w_0 is usually selected at random, which will restrict the performance of ICA. From this perspective, a specified projection vector initialization strategy is introduced in the first part of our proposed method. The LLE algorithm is applied to reduce the dimension of HSI and keep the components that contain only useful information. LLE assumes that the data is distributed on a manifold of a high-dimensional space, but linear in a small local area [23]. The data of HSI is represented as $X=[x_1,x_2,...,x_M] \in R^{D \times M}$, M denotes the number of pixels in a spectral band and D denotes the number of bands. LLE achieves the mapping of high dimensional space manifolds in low dimensional space [33]. Both the high and low dimensional spaces share the same topology structure. The algorithm flow can be conducted as follows

(i). Compute the Euclidean distance $dx(i,j)$ between any two samples x_i and x_j in X. Each point of the distance matrix D is expressed as $D_{ij}=dx(i,j)$.

(ii). Find the K nearest samples to x_i in X according to D_{ij}, expressed as $\{x_{ij}\}_{j=1}^K$

(iii). Take (9) as the objective function, where p_{ij} is a linear reconstruction coefficient corresponding to x_{ij}.

$$\min \sum_{i=1}^N \left\| x_i - \sum_{j=1}^K q_{ij}x_{ij} \right\|^2$$
$$s.t. \sum_{j=1}^K q_{ij} = 1 \tag{9}$$

(iv). Calculate y_i and y_{ij}, which denote the low dimensional mapping of x_i and x_{ij} respectively.

$$\min \sum_{i=1}^N \left\| y_i - \sum_{i=1}^K q_{ij}y_{ij} \right\|^2 \tag{10}$$

Define a matrix $Q \in R^{N \times N}$. If x_i and x_j are neighbor pixels, Q_{ij} is equal to q_{ij}, otherwise Q_{ij}=0. This problem is equivalent to

solving the eigenvectors from the second to the $(d+1)$th minimum eigenvalues of matrix M, from $M=(I\text{-}Q)^{\mathrm{T}}(I\text{-}Q)$. The eigenvector will be the output, expressed as $Y=[y_1, y_2,...,y_M] \in R^{d \times M}$. Following this, the data is whitened to derive $Z \in R^{d \times M}$ by $z_i=D^{-1/2}E^T y_i$ $1 \leq i \leq M$ [19] [24]. D and E are the eigenvalue matrix and the eigenvector matrix corresponding to the covariance matrix of y_i respectively.

Then, Z is substituted into the a locally adaptive kernel density estimation (LAKDE) algorithm [34]. LAKDE is remarked as a local AD strategy. It is employed within the likelihood ratio test (LRT) rule to model the local background, and information extracted from the image is injected into the bandwidth selection process. According to [30], the proposed strategy is obtained by

$$\Lambda_{AD}(y) = -\log\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{h^d} \cdot k(\frac{y-Y_B(i)}{h})\right\}$$

$$\begin{cases} y \in H_1 & if\ \Lambda_{AD}(y) \geq \lambda \\ y \in H_0 & if\ \Lambda_{AD}(y) \leq \lambda \end{cases} \quad (11)$$

where y is the test pixel. $Y_B(i)$ represents the ith background sample. N, d, and k denote the number of background samples, the matrix dimension, and the kernel function respectively (the Gaussian Radial Basis (GRB) is selected). h is bandwidth. If h is too large or too small, the result will all be affected [36]. So in [33], an adaptive method is proposed to specify h. The Euclidean distance between y and $Y_B(i)$ was calculated. After all N distances were calculated, the kth distance was sorted in ascending order, where k was usually taken the first 20 percent of N. Equation (13) was exploited to obtain the detection result. The pixel with the greatest brightness was picked out as the initialization projection vector w_0 in ICA. w_0 contains comparatively abundant anomaly information that may be a pure outlier in large probability. Due to the characteristic of ICA that the local maximum is calculated iteratively, the maximum projection of the anomaly will be derived by selecting w_0 to initialize the projection vector in ICA, which can distinguish the IC with most anomaly information.

B. Assessment of anomaly information content

In this subsection, we substitute aforementioned w_0 into OICA and derive the final w by (3), 4) and (5). Then the estimate of IC, demoted as x^*, can be obtained by $w^T Z$. Following this, Potential Anomaly Signal to Noise Ratio (PASNR) [28] is exploited to estimate the anomaly information content in each IC. PASNR is a feature selection method based on histogram statistic and threshold segmentation. The key is to find a threshold in the abundance image of ICs for segmentation. Pixels with grayscale values larger than the threshold are considered as potential anomalies, otherwise background pixels. Then utilize (14) to compute PASNR

$$PASNR = 10 \cdot \log_{10}\left(\frac{\mathrm{var}(y_P)}{\mathrm{var}(y_B)}\right) \quad (12)$$

where var (·) denotes the variance of the data. y_P and y_B are the grayscale value of potential anomaly and background pixel respectively. As seen in section B, the threshold between potential anomaly signals and the background plays the vital role in calculating the PASNR for each IC. According to Chiang *et al* [37], it can be specified that the potential anomaly pixels are those pixels that come after the first-empty zero-point bin from the histogram constructed from the signal, as shown in Fig. 1(a), where the horizontal axis is the gray value, and the vertical axis is the number of pixels. We can find this first-empty zero point as the threshold. The lighter pixels are usually represented as anomaly, however, sometimes the gray value of background pixels can be higher than anomaly pixels in gray scale distribution of ICs, as shown in Fig. 1(b). Because in some ICs there are very few anomaly pixels and the energy of background is much stronger than outliers, which are not light enough. Therefore, in the proposed method, we find the nearest zero point from the peak in both directions meanwhile, and segment the image according to the gray value corresponding to the two zero points respectively. Then the PASNR is calculated in both images, and select the greater one as the final result, expressed as VP. In contrast, the selection from both sides is more considerate than from one side, and the image can be segmented more clearly.

C. Refinement of abnormal components extraction

As described in section A, w_0 is obtained as the initialized projection vector for OICA and is utilized to solve the final projection vector w_i. Then calculate the first IC x_1^* by $w_1^T Z$. PASNR is utilized to estimate the abnormal information content in x_1^*. It is necessary to set a threshold γ to compare with the VP_1. Once VP_1 is larger than γ, it means x_1^* contains abundant abnormal information. According to the characteristics that ICA is able to calculate the local maximum iteratively, the projection direction of the maximum anomaly can be finally derived as w.

However, there will be another problem. After obtaining the first IC, the IC obtained later may still contain much abnormal information, which possibly dispersed to a series of subsequent ICs. This is not what we expect. In this article, the OSP algorithm and an iteration process are introduced to effectively remove the interference of uninteresting information. The specific implementation steps are as follows:

Firstly, the whitened matrix Z is processed by OSP according to (10), after the first w_0 processed by a simple eight-neighborhood average. it can effectively avoid the repeated detection of pixels with the same position in initializing the vector w_0. Then a new initial vector w_0, which distinguishes from the first one above, is obtained by LAKDE. Through OICA, w and the second IC are obtained as well. and meanwhile w will also be not similar with the previous one. The remaining abnormal information are concentrated on the front ICs as far as possible with the effects of suppressing the background by OSP.

After getting the second IC, its value of PASNR VP_2 is also compared with γ. If VP_2 is larger, there will be still considerable
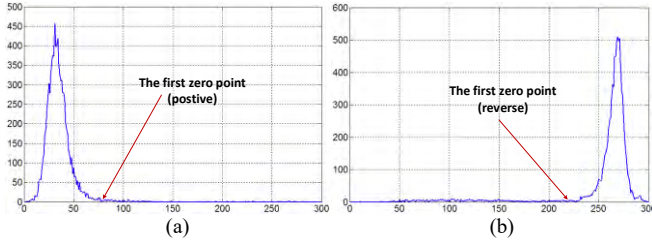
Fig. 1. Statistical histogram of independent components

abnormal information in the rest data. Then the iterative process is repeated until VP_n is less than γ, which indicates that the abnormal information in the image are all extracted, and the iteration is terminated.

When OSP is applied to anomaly detection in HSI, it is difficult to construct the background subspace without any prior information of the image. Some researchers [32] construct the background subspace U by modeling the local data and comparing the spectral differences between the center pixel and its neighborhoods. Then P can be obtained by using the eigenvector matrix B of U, which is complicated and may suppress both background and anomalies. In the article, we transform the whitened matrix Z obtained by dimensionality reduction for X with LLE in section A, containing d bands. Then the abnormal information content of each band is estimated by PASNR, one with the smallest VP is selected as the background matrix U. The orthogonal complement space P can be obtained according to (7). In this way, an approximate projection direction is found. The data is projected onto P by (8), which can eliminate quite a few background feature components in the image.

After iteration, we can obtain n ICs whose VP are all larger than γ, and construct them to a new dataset Z^*, which contains almost all of abnormal information. Finally, the kernel local RX algorithm is exploited to scan each pixel by double windows, in order to suppress the remaining background pixels and only leave outliers as the final result.

D. Overview of Proposed IOACED Framework

The specified projection vector initialization method, the assessment of anomaly information content, and the refinement of abnormal components extraction based on OSP are three main implementations in the proposed IOACED framework. The detailed process of the proposed IOACED framework is summarized in the following.

1) Set the parameters: the threshold γ.
2) For an HSI dataset X, a corresponding Y is obtained through the dimensionality reduction by the LLE algorithm. Use the whitening operation $z_i = D^{-1/2} E^T y_i$ to obtain Z.
3) Z is detected by the LAKDE algorithm in order to extract the pixel with the greatest brightness as the initialization vector w_0
4) Solve the final projection matrix w by OICA with w_0, and calculate the corresponding IC.

5) Utilize the PASNR method to estimate the abnormal information content of the IC, expressed as VP_i. Set the threshold γ to compare with VP_i. If VP_i is larger than γ, smooth w_0 by an eight-neighborhood average, and then the OSP algorithm is applied, as shown in (8), to process Z to further suppress the background and concentrate the abnormal information. Then return to step 3) for iteration. When VP_i is larger than γ, go to step 6).
6) Reconstitute all the remaining ICs to a new 3-D dataset Z^*. The KLRX algorithm is applied to suppress the remaining background pixels and only leave outliers as the final result δ_{IOACED}.

Algorithm 1 Hyperspectral Anomaly Detection via the IOACED Framework

---

Input: HSI data X, number of the nearest samples K in LLE, dimension of dataset d after dimensionality reduction, threshold γ

Obtain: the whitened matrix Z and the initialization vector w_0 extracted by LLE and LAKDE

Compute: 1) the final projection matrix w by OICA via (3), (4), and (5);
   2) the corresponding IC;
   3) the abnormal information content of the IC, VP_i via (12);

While $VP_i > \gamma$ **do**
   Smooth last w_0 by an 8-neighborhood average;
   Compute the projection matrix $P_U^\perp$ via (7);
   Update $Z_{OSP} = P_U^\perp Z$;
   Update w (3), (4), and (5);
   Compute the corresponding IC and VP_i
End While
Compute: δ_{IOACED} with Z^*.
Output: δ_{IOACED}

The main steps for implementing the proposed IOACED framework are listed in Algorithm 1. And the flowchart of the IOACED method is shown as Fig. 2.

IV. EXPERIMENTS AND ANALYSES

This section refers to a series of verified experiments undertaken to evaluate the performance of the proposed IOACED framework. The constituent parts are as follows.

1) Descriptions of the hyperspectral datasets used to investigate the effectiveness of the proposed method.
2) Detection performance evaluation of the proposed IOACED spectral anomaly detection methods on the experimental hyperspectral datasets, including detection maps, receiver operating characteristic (ROC) curves, and background-anomaly separation analyses
3) Discussion of the experimental results with the area under ROC curve (AUC) values.
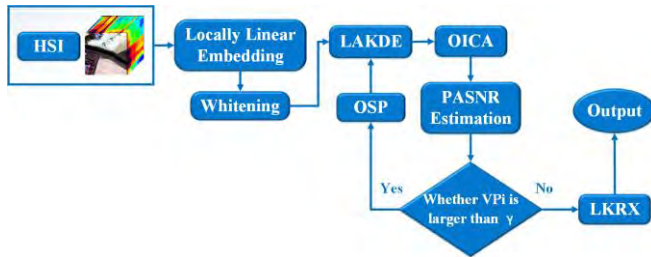4) Sensitivity analysis of the relevant parameters.

Fig. 2. Flowchart of the proposed IOACED method

A. Hyperspectral Dataset Description

In this subsection, three real-world hyperspectral datasets are utilized in the experiments to investigate the performance of the proposed IOACED method in the task of detecting anomaly targets. These three datasets possess different properties, including the data collection sensor, the clutter of the background land cover, the field of image coverage, and the size and spatial distribution of the targets. All of the detection algorithms and experiments are implemented in MATLAB v2017 on a personal computer with an Intel Core i3 3.40-GHz central processing unit, 8.0 GB of RAM, and 64-bit Windows 7.

The first hyperspectral dataset was collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) hyperspectral remote sensor, covering a scene of Moffett Field, California, at the southern end of the San Francisco Bay, with spatial resolution of 3.5m per pixel. The raw HSI consists of 224 spectral channels spanning the wavelength range of 0.4 to 2.5um. In our experiment, 178 bands of the raw HSI were retained after removing the bands corresponding to low signal-to-noise ratios and water absorption regions (1–98, 113–114, 129-153, and 172–224). In the original image, a scene about the Moffett Field airport is picked out for anomaly detection, which covers the area of 192×48 pixels. The main background land-cover types in this image are ground, road, and grass, as shown in Fig. 3(a). There are seven aircrafts and three maintenance vehicles in the image, whose spectra are different from the spectra of the background materials, as illustrated in Fig. 3(c), which consist of 80 pixels and account for 0.87% of the whole image, as shown in Fig. 3(b). Hence, these seven aircrafts and three maintenance vehicles are considered as the anomaly targets to be detected in this dataset.

The second dataset is the San Diego data, also collected by the AVIRIS, over San Diego, CA, USA [38] [39]. The initial images have 3.5 m spatial resolution and 224 spectral channels ranging from 370 nm to 2510 nm. In the experiment, a subset image with the image size of 100 × 100 pixels was selected, and the digital numbers were used as the input data. After removing the bad bands [1–6, 33–35, 97, 107–113, 153–166, 221–224] due to water absorption and low signal-to-noise ratio, the 189 bands were used in the experiment. In the image scene shown in Fig. 4(a), main ground objects of the background are roof, road, shadow and grass. Three planes occupying 58 pixels were commonly regarded as anomalies because they cover a very small number of pixels and are spectrally different from main

ground objects. Fig. 4(b) shows the ground objects of the anomalies and Fig. 4(c) plots spectral curves of anomalies and main ground objects.
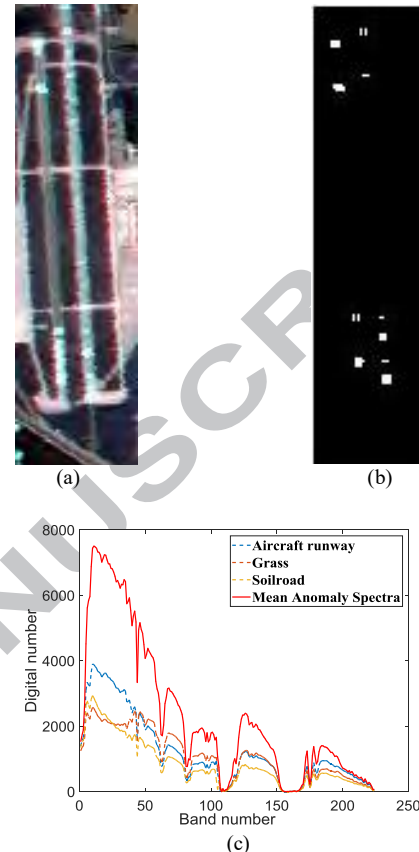

Fig. 3. AVIRIS-Moffett Field hyperspectral dataset. (a) Pseudo-color image. (b) Reference. (c) Spectra of the background and the anomalies.

The third hyperspectral dataset was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) hyperspectral remote sensor. This HSI, which consists of a suburban residential area with an approximately 3m spatial resolution, is publicly available [40]. It consists of 210 spectral channels from 0.4 to 2.5 um. After removing the bands of the water absorption regions, low signal-to-noise ratio, and poor quality (1–4, 76, 87, 101–111, 136–153, and 198–210), 162 bands remained. This image scene is cluttered with different background land-cover types of parking lot, soil, water, road I, and road II, as depicted in Fig. 5(a). There are 8 man-made vehicles, which consist of 17 pixels and account for 0.21% of the whole image, as shown in Fig. 5(b). The spectra of these pixels are different from the spectra of the background land-cover types, as illustrated in Fig. 5(c), and these pixels are considered as the anomaly targets to be detected.
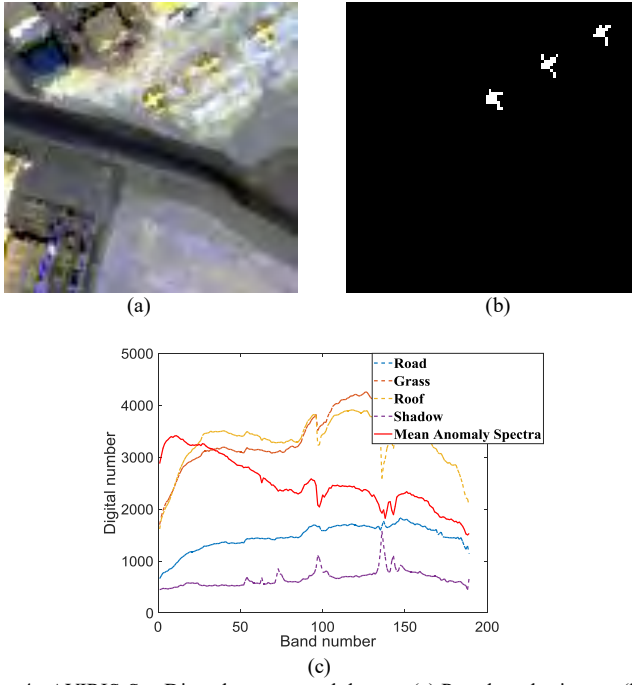
Fig. 4. AVIRIS-San Diego hyperspectral dataset. (a) Pseudo-color image. (b) Reference. (c) Spectra of the background and the anomalies.
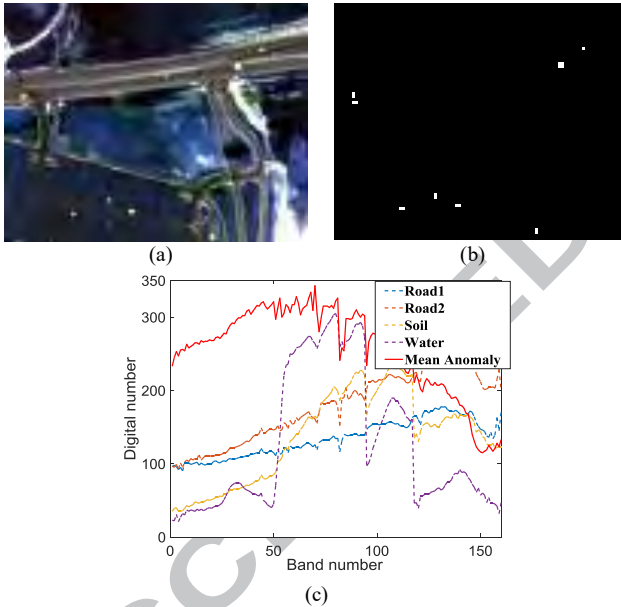

Fig. 5. HYDICE hyperspectral dataset. (a) Pseudo-color image. (b) Reference. (c) Spectra of the background and the anomalies.

B. Experimental Details

In this subsection, the proposed IOACED method is compared with a set of state-of-art anomaly detectors: RXD [6], LRX [42], KLRX [11], LRR [46], SRD [43], CRD [44] and ICA-KLRX [28]. The parameter analysis for IOACED will be in Section *C*. and the setup of these detectors are described as follows.

1) Parameter setup for Compared algorithms

Table I lists the parameters of five detectors on the three datasets. For the LRX and KLRX methods, the outer window (OW) and inner window (IW) of the AVIRIS-Moffett Field dataset are set to be 13×13 and 5×5, respectively; those of San Diego dataset are set to be 21×21 and 5×5, respectively; and those of HYDICE dataset are set to be 19×19 and 7×7, respectively. For the CRD and SRD methods, the OW and IW of the AVIRIS-Moffett Field dataset are set to be 15×15 and 7×7, respectively; those of San Diego dataset are set to be 17×17 and 7×7, respectively; and those of HYDICE dataset are set to be 13 ×13 and 7×7, respectively. For the ICA-KLRX method, HFC is utilized to estimate the virtual dimensionality(VD) [45]. The VD of two AVIRIS datasets are 11, and of the HYDICE dataset is 6. The size of IW and OW are the same as the KLRX.

2) Detection Performance

The detection performance is evaluated quantitatively by receiver operating characteristic (ROC) curves [41] and the normalized background-anomaly separation maps. Based on the ground truth, the ROC curve plots the relationship between the detection rate(DR) and false alarm rate(FAR); DR and FAR are defined as follows:

$$DR = \frac{R_{detection}}{R_{anomaly}} \quad FAR = \frac{R_{false}}{R_{image}} \tag{15}$$

where $R_{detection}$ is the number of correctly-detected anomaly pixels and $R_{anomaly}$ is the total number of anomaly pixels in the image. R_{false} is the number of falsely-detected anomaly pixels and R_{image} is the total number of pixels in the image. The ROC curve of an outstanding detector is located near the top left of the coordinate plane, since it can achieve a high detection rate with a low false alarm rate. Moreover, the AUC value is also used to measure the detection performance, which denotes the area under the ROC curve. A detector with better detection performance has a larger AUC value.

For the AVIRIS-Moffett Field dataset, the color detection maps of algorithms in comparison are shown in Fig. 6. ICA-KLRX obtains a superior performance in background suppression and anomaly extraction overall. However, for ICA-KLRX, the brightness of anomaly targets in the image is not very high, because it is affected by the background suppression. Furthermore, none of three small anomaly targets is detected. The detection probability is limited. In addition, RX and LRX have a poor background suppression, and only anomaly targets with high anomaly abundance fractions can be well detected. For KLRX, the background suppression is not very effective, but the detection performance is better than RX and LRX, and some anomaly targets with low anomaly abundance fractions are still distinguishable. For LRR, the performance of suppressing background is better than KLRX, and the false alarm in the middle right of the scene is suppressed as well, though there is still some background can be obviously seen, especially in the bottom left of the scene. For SRD, three small anomalies can be seen, but not obviously, and the background in the bottom left of the scene is suppressed badly. CRD can detect all anomalies, including seven small targets, and the background in the bottom left of the scene is suppressed well.

However, the suppression for the runway in the middle of the scene is relatively weak, which can increase the false alarm rate.

In the proposed IOACED algorithm, it can be seen that the background suppres-

TABLE I
THE LISTS OF PARAMETERS OF ALL DETECTORS ON THE THREE DATASETS

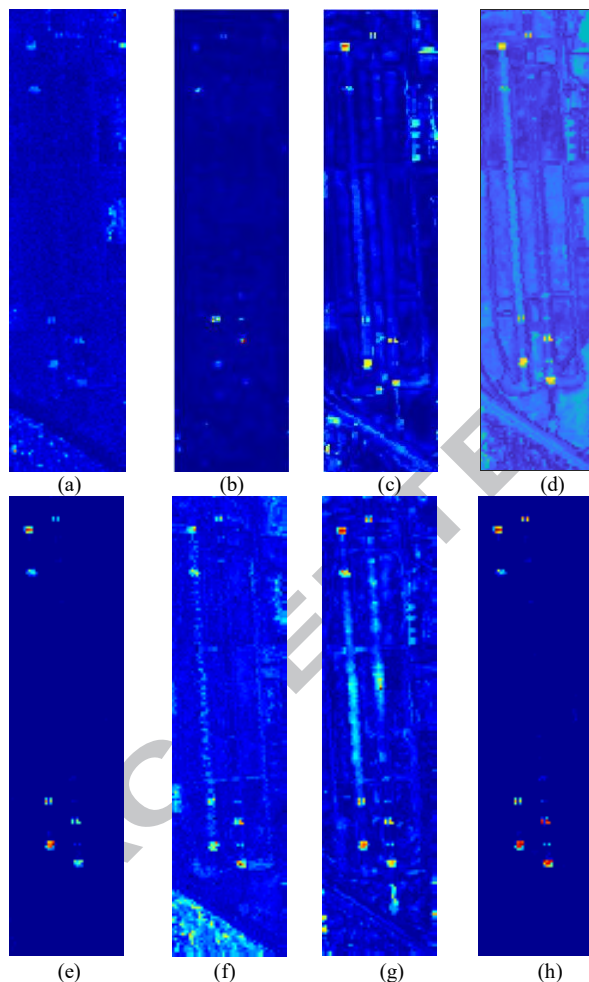| Datasets	Configuration Parameters							
	RXD	LRR	LRX	KLRX	CRD	SRD	ICA-KLRX	
AVIRIS-Moffett Field	-	-	OW=13×13	OW=13×13	OW=15×15	OW=15×15	OW=13×13	VD=11
			IW=5×5	IW=5×5	IW=7×7	IW=7×7	IW=5×5	
AVIRIS- San Diego	-	-	OW=21×21	OW=21×21	OW=17×17	OW=17×17	OW=21×21	VD=11
			IW=5×5	IW=5×5	IW=7×7	IW=7×7	IW=5×5	
HYDICE	-	-	OW=19×19	OW=19×19	OW=13×13	OW=13×13	OW=19×19	VD=6
			IW=7×7	IW=7×7	IW=7×7	IW=7×7	IW=7×7	

Fig. 6 Color detection maps of different algorithms for the AVIRIS-Moffett Field dataset. (a) RX. (b) LRX. (c) KLRX. (d) LRR. (e) ICA-KLRX. (f) SRD. (g) CRD. (h) IOACED.

sion performance is outstanding, and all anomaly targets are successfully extracted. A quantitative analysis of these different anomaly detection algorithms is illustrated in Fig. 7. Fig. 7(a)

presents the ROC curves of each algorithm. It can be observed that the IOACED algorithm achieves the best detection performance among all of the algorithms in the figure, since a high detection rate can be obtained when the false alarm rate is very low. ICA-KLRX also has good probabilities of detection with low false alarm rates. The probability of detection of ICA-KLRX reaches more than 0.9 when the false alarm rate is 0.005. Except for ICA-KLRX, LRR and KLRX achieve better ROC curves than the other compared methods.
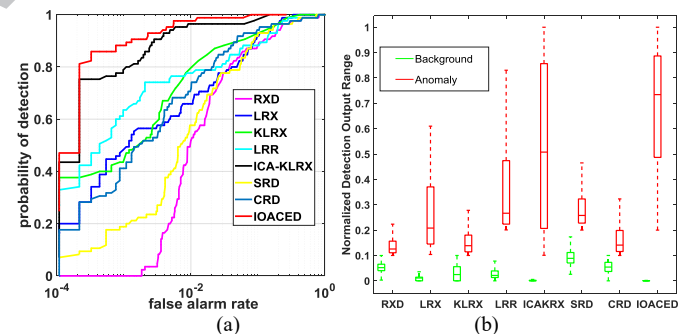

Fig 7. Quantitative analysis of compared algorithms for the AVIRIS-Moffett Field dataset. (a) ROC curves. (b) background-anomaly separation map.

Fig. 7(b) demonstrates the separability of background and anomaly targets in all investigated algorithms through the background-anomaly separation map. For the convenience of comparison, values of all detection results are normalized to 0-1. Each detector has two boxes: the green one represents the distributions of background pixels' values, and the red one is the distributions of anomaly pixels' values. The central mark in each box indicates the median, while the bottom and top edges indicate the lower quartile and the upper quartile, with the whiskers indicate the extreme values within 1.5-times the interquartile range from the end of the box. From Fig. 7(b), it is proven that the IOACED algorithm and ICA-KLRX can suppress the background information to a small range, and the gap between the anomaly box and the background box is large when compared with other detectors. But the normalized de-

tection output range of the anomaly box of IOACED is 0.85, which is higher than ICA-KLRX, 0.82. Therefore, the IOACED algorithm has the best background-anomaly separation performance in all compared algorithms.

For the AVIRIS-San Diego dataset, the color detection maps of all compared algorithms are illustrated in Fig. 8. It can be seen that KLRX, LRR, and the IOACED algorithm have relatively better performance in background suppression, and IOACED is the best one among them. Besides, the anomaly targets extraction of KLRX, ICA-KLRX and IOACED are all effective, but ICA-KLRX has more false-alarm rate. In addition, RX suppresses most of the background, but anomaly targets are affected as well. In LRX, the performance is very weak, the image is scattered, and the targets are unseen. For CRD, the performance of anomaly target extraction is poor, and a part of the anomalies is indistinguishable from the background. For SRD, it has better detection performance than LRX and CRD. Fig. 9 shows the quantitative comparisons of these algorithms for the AVIRIS- San Diego dataset via ROC curves and background-anomaly separation map. From Fig. 9(a), it can be seen that IOACED and KLRX obtain superior ROC curves, and are always higher than other algorithms on the probability of detection. When the false alarm rate is less than 0.01, the performance of IOACED is much better than KLRX. When the false alarm rate is more than 0.005, all anomaly pixels can be detected by IOACED. In addition, the background-anomaly separation map is demonstrated in Fig. 9(b). It shows that IOACED has a more advanced performance when compared with other algorithms.

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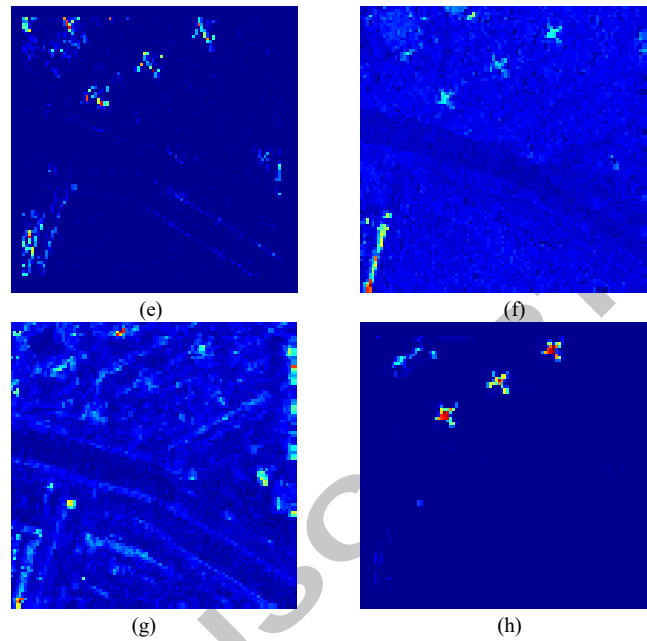

(e) (f) (g) (h)

Fig. 8. Color detection maps of different algorithms for the AVIRIS- San Diego dataset. (a) RX. (b) LRX. (c) KLRX. (d) LRR. (e) ICA-KLRX. (f) SRD. (g) CRD. (h) IOACED.
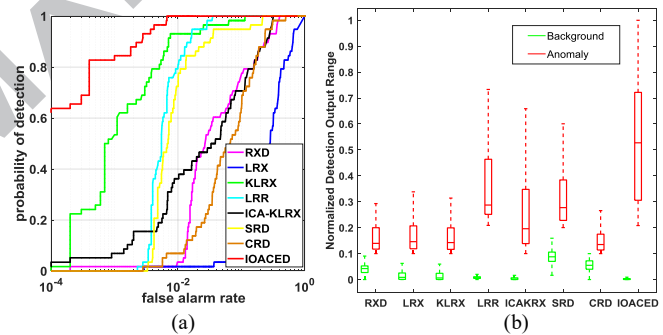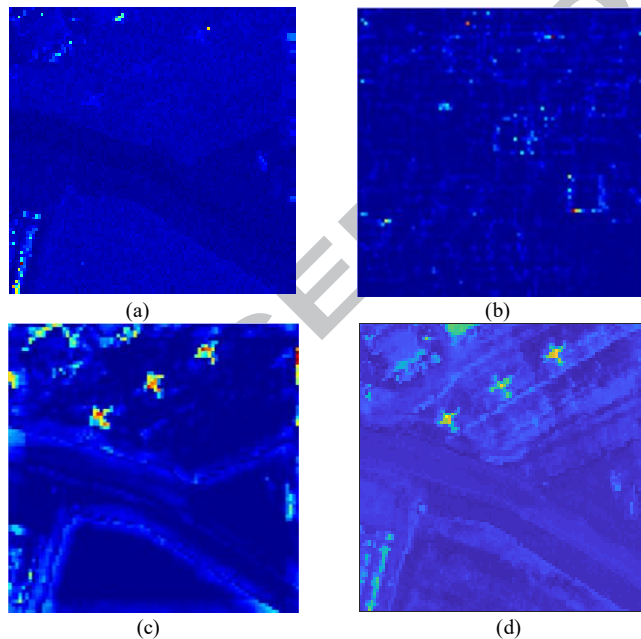

(a) (b)

Fig 9. Quantitative analysis of compared algorithms for the AVIRIS- San Diego dataset. (a) ROC curves. (b) background-anomaly separation map.

For the HYDICE dataset, Fig. 10 shows the color detection maps of all compared algorithms. It can be observed that the proposed IOACED algorithm extrudes the anomalies by suppressing most of the background land-cover types to a low detection output, though there are still two false alarm pixels in the up right of the scene. Comparing the details in the detection maps, IOACED achieves a superior performance in suppressing the background than other algorithms. RX and SRD obtain high false alarm rates. For LRX, it has good performance for background suppression, but some anomaly targets are suppressed as well, only three targets in the bottom can be seen clearly. For LRR, all anomalies can be found, and the performance of background suppression is better than KLRX, but worse than IOACED. Fig. 11 gives a quantitative comparison between the different anomaly detectors with ROC curves and background-anomaly separation map. From Fig. 11(a), it can be observed that LRR, CRD, and IOACED all obtain high probabilities of detection with low false alarm rates. The probability of detection of IOACED is best when the false alarm rate is less

than 0.001, and more than 0.9 when the false alarm rate is 0.02. Compared to LRR, and CRD, IOACED exhibits a slightly lower probability of detection when the false alarm rate is [0.003, 0.01]. However, the overall detection performance of IOACED is still better. From Fig. 11(b), it can be seen that the proposed IOACED algorithm achieves a better background-anomaly separation than the other methods.
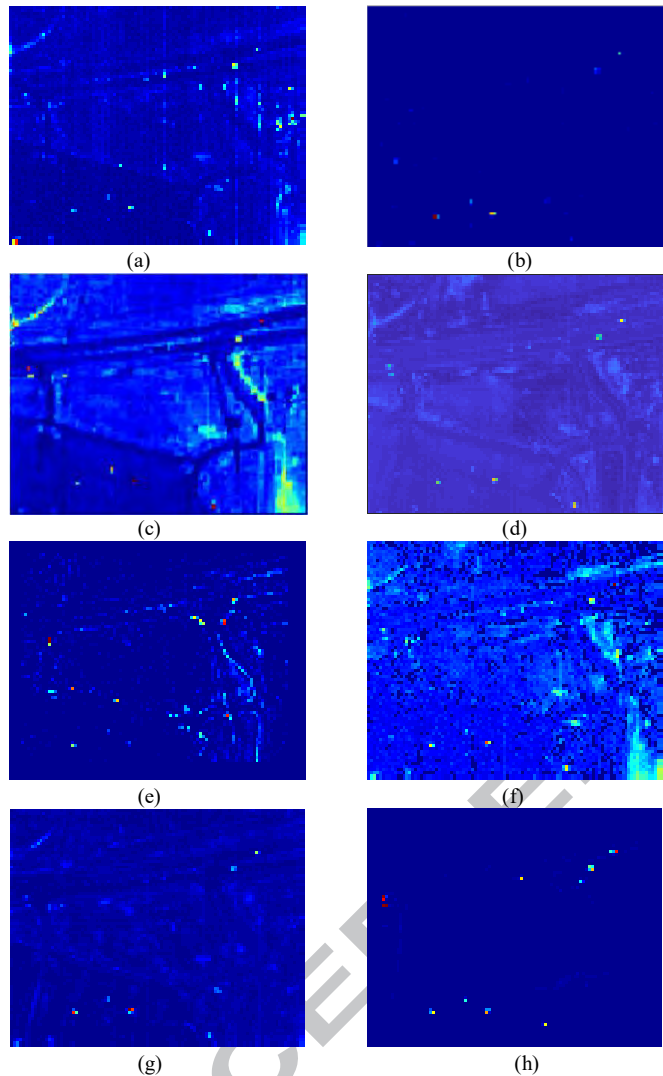

Fig 10. Color detection maps of different algorithms for the HYDICE dataset. (a) RX. (b) LRX. (c) KLRX. (d) LRR. (e) ICA-KLRX. (f) SRD. (g) CRD. (h) IOACED.
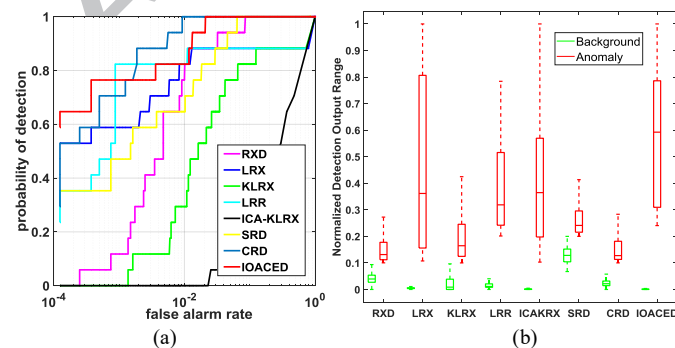

Fig. 11. Quantitative analysis of compared algorithms for the HYDICE dataset. (a) ROC curves. (b) background-anomaly separation map.

C. Discussion and Analysis

A discussion of the proposed IOACED framework and the parameter analysis of its detection performance are provided in the following.

TABLE II
AUC VALUES OF DIFFERENT ALGORITHMS FOR THE AVIRIS-MOFFETT FIELD, AVIRIS-SAN DIEGO AND HYDICE DATASETS

Dataset	Algorithm/AUC value			
	RXD	LRX	KLRX	LRR
AVIRIS-Moffett Field	0.9540	0.9720	0.9791	0.9675
	ICA-KLRX	SRD	CRD	IOACED
	0.9957	0.9609	0.9706	**0.9986**
	RXD	LRX	KLRX	LRR
AVIRIS-San Diego	0.9202	0.6788	0.9944	0.9924
	ICA-KLRX	SRD	CRD	IOACED
	0.9271	0.9817	0.9037	**0.9994**
	RXD	LRX	KLRX	LRR
HYDICE	0.9692	0.8932	0.8760	0.9257
	ICA-KLRX	SRD	CRD	IOACED
	0.7194	0.9789	0.9870	**0.9887**

TABLE III
RUNNING TIMES OF THE ALGORITHMS OBTAINED WITH THE HYPERSPECTRAL DATASETS

/second	AVIRIS-Moffett Field	AVIRIS-San Diego	HYDICE
RXD	0.13	0.20	0.11
LRX	32.61	40.47	24.76
KLRX	33.88	39.00	33.29
LRR	43.80	52.73	47.53
ICA-KLRX	16.71	12.51	12.94
SRD	55.96	137.86	60.18
CRD	28.41	43.50	36.04
IOACED	**28.42**	**33.79**	**25.35**

First, a quantitative index, AUC, is utilized for the investigation of the experimental results of the three datasets. A detector with a larger area under its ROC curve will obtain a larger AUC value, which means that the detector obtains a better detection performance, and the best AUC performances are highlighted for each experimental dataset in Table II. The statistics in Table II further indicate that the IOACED algorithm yields the best detection performance in all the compared algorithms. In fact, three datasets in our experiments have different characteristics. For the experimental results of the datasets, it can be found as follows.

1) For the AVIRIS- Moffett Field dataset, it contains three different types of backgrounds, which can make the detection more difficult. However, the IOACED algorithm achieves very high AUC values (more than 0.99), which proves the advantage of the proposed algorithm.

2) For the AVIRIS-San Diego hyperspectral dataset, the proposed IOACED framework perfectly suppresses the background land-cover types of roof and shadow, whose spectra are somewhat correlated with the spectra of the anomaly targets, and obtains a relatively high level of anomaly extrusion.

3) For the hyperspectral dataset acquired by the HYDICE sensor, great differences between the spectra of the background land-cover types lead to difficulty in implementing effective background suppression. This also makes the anomaly targets difficult to detect. However, the proposed IOACED framework suppresses the background land-cover types in this dataset to a satisfactory level and extrudes the anomaly targets.

Moreover, the computational costs of the proposed IOACED method and the compared algorithms are shown in Table III. It can be seen that the SRD algorithm performs the highest computational cost, and the proposed IOACED algorithm has the similar cost with CRD, which is lower than SRD. When the size of HSI dataset is large, the computational cost of algorithms increases as well. The time cost is an important factor for the practical application of an anomaly detector, so our proposed IOACED algorithm has an advanced performance in efficiency.

Second, we evaluate the influences of the different parameters settings of the IOACED algorithm. There are three important parameters in our proposed method. The first parameter analysis was undertaken with the AVIRIS-Moffett Field dataset, and the other two were with the three datasets. We implemented this analysis by the use of AUC values calculated for the whole area under the curve.

Here the three parameters are: (1) the number of the nearest samples K in the LLE algorithm; (2) the dimension of dataset d after dimensionality reduction; and (3) the threshold γ in IOACED. Fig. 12 shows the ROC curves of IOACED on three datasets with the changing number of the nearest samples K from 20 to 45. For the Moffett Field dataset of Fig. 12(a), when the K increases from 20 to 45, the ROC curves of IOACED have similar trend with small fluctuations. Particularly, when the false alarm rates are between 0.0004 and 0.01, the probability of detection with K=20 is higher than others. The ROC curves of the HYDICE dataset in Fig. 12(c) have similar observations with that of Fig. 12(a), but the probability of detection with K=40 is the best. In Fig. 12(b), we can obtain the same conclusion more observably.
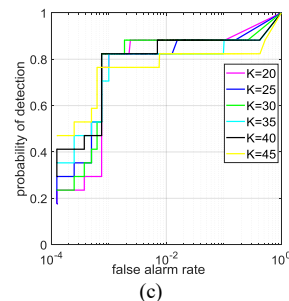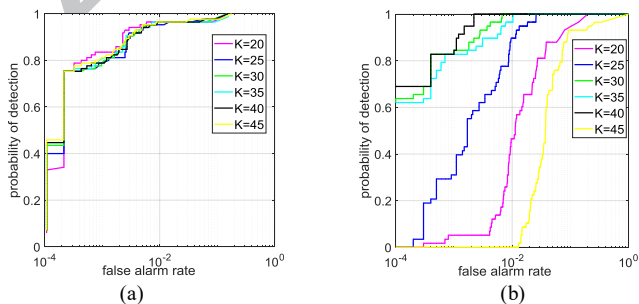

(c)

Fig. 12. The ROC curves of IOACED with different choice of the number of the nearest samples K for LLE on the three datasets. (a) AVIRIS- Moffett Field. (b) AVIRIS- San Diego. (c) HYDICE.

Furthermore, Fig. 13(a) shows the result of AUC curves with the dimension of dataset d after dimensionality reduction on the three datasets. In the figure, the AUCs of IOACED on the Moffett Field and the San Diego datasets have similar trend and almost keep stable and smooth with d from 6 to 11, and the values of AUC are all higher than 0.99. When d is 12, the AUC curves of the San Diego has an obvious fall. Particularly, compared to two AVIRIS datasets, the AUC curve on the HYDICE dataset has the highly distinguishing performance. It has the greatest fall decreasing from 0.9123 to 0.8814 as the d increases from 9 to 11. The reason is that LLE achieves the mapping of high dimensional space manifolds in low dimensional space. Both the high and low dimensional spaces share the same topology structure. The sizes of anomaly targets in the first and second datasets are large, and the background is relatively simple. The topological structure of the reconstructed matrix after dimensionality reduction is quite similar to that of the original data matrix, and has strong robustness to different band numbers after LLE. While the scene of the third dataset is more complex and the sizes of the anomaly are quite small. The background disturbs the abnormal object greatly. This leads to the difference of the topological structure between dimension-reduced matrix and the original matrix. to this end, when dimension D changes, the w_0 may change as well, resulting in different detection performance. When d is 12, the value of AUC reaches 0.9186, which shows the best performance. Fig. 13(b) plots the curves of computational time with respect to the increasing dimension number d. The results show that computational time of IOACED increases with the rising d. Nevertheless, a larger value of d does not necessarily correspond to a better detection performance, shown as Fig. 13(a), and significantly increase computing time. Therefore, a smaller value of d, e.g., 9, is preferred.
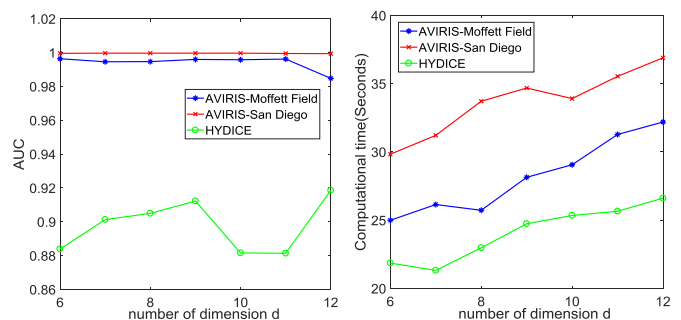

(a)

(b)

(a)   (b)

Fig. 13. The curves of (a) AUC and (b) computation time with different choices of dimension d on the three HSI datasets.

Fig. 14 illustrates histogram of AUC and the line chart of iteration number of IOACED on the three datasets with the changing threshold γ. The histogram corresponds to the AUC value on the left, and the line chart corresponds to the number of iterations on the right side. It can be seen that the same algorithm has different threshold ranges for different datasets due to discrepant geological types and imaging characteristics of spectrometer. The range of the threshold we select in (a) is [0.2, 2]. When γ =0.2, the threshold is too small. As the result, the number of iteration reaches to 15, the time consuming is high, and the performance is not obviously improved. When γ is between [0.3, 0.4], the value of AUC is 0.9986, the algorithm iterates 3 times, of which both the performance and efficiency are the highest. In (b), the selection of the threshold range is [4.5, 9]. When γ is between [4.5, 5.5], the algorithm iterates 4 times and the value of AUC is 0.9997. As γ increases, both the iteration number and the detection performance decreases. The threshold range selection in (c) is [0.04, 8], but only when γ is 0.04 does the proposed method have the maximal iteration times 5, and the best detection performance 0.9887. the value of AUC with γ between 0.1 and 3 is 0.9864, which is a little lower than γ=5. Moreover, the iteration number is only 2 when γ is [4, 8], and the performance is greatly reduced as well.
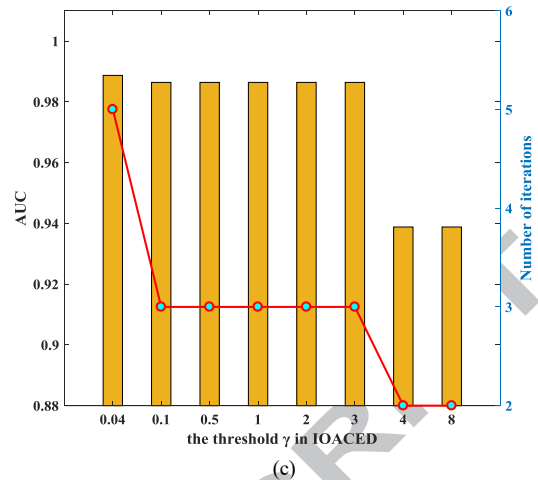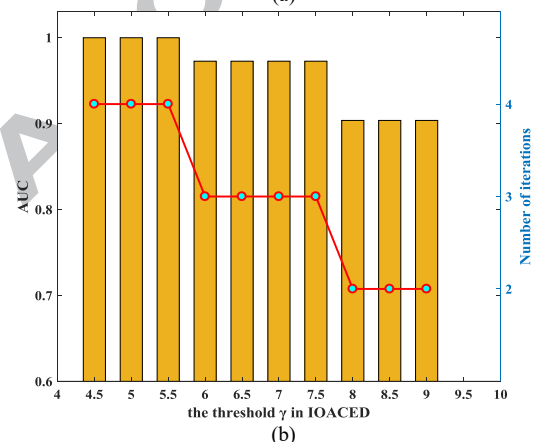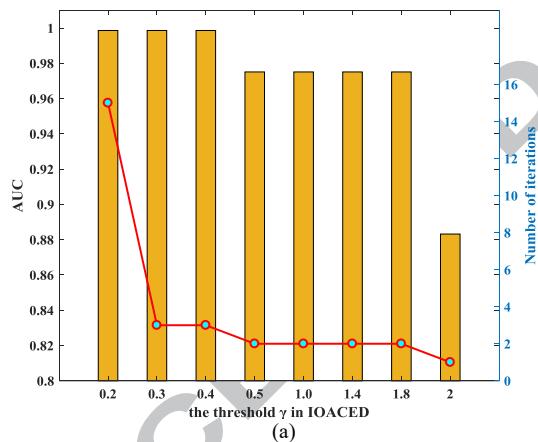
(c)

Fig. 14. The histogram of AUC and the line chart of iteration number on three HSI datasets with different choices of the threshold γ (a) AVIRIS- Moffett Field. (b) AVIRIS- San Diego. (c) HYDICE.

From the aforementioned analyses, the number of the nearest samples K, the dimension of dataset *d* after dimensionality reduction and the threshold γ are three parameters in the IOACED algorithm that have a relatively obvious influence on the detection performance. On the whole, the IOACED algorithm is quite robust to the settings of parameters if they are in reasonable ranges. The change of parameters will greatly affect the computational costs of the algorithm. Moreover, currently the parameters are set empirically and artificially. In the future, we will focus on how to determine all parameters automatically and adaptively, and we believe this will further improve the practicality of this method.

V. CONCLUSIONS

In this paper, a novel hyperspectral anomalous component extraction framework based on ICA and OSP is proposed. The LLE and the LAKDE algorithm are utilized to extract the brightest anomalous component as the initial projection vector, which improve the anomaly detection performance of the ICA algorithm. Furthermore, the OSP method is exploited to suppress the background components in the surplus data. In this way, the anomalous components in the iteration operation are refined, and the discrimination between anomalies and background is enhanced. Extensive experiments on real datasets confirm that the proposed anomaly detection algorithm has superior detection performance compared to the other detection methods. Moreover, the parameter sensitivity analysis proves the robustness of the proposed method with reasonable parameters settings.

VI. ACKNOWLEDGEMENTS

(a)

(b)

Conflict of interest statement

We declared that they have no conflicts of interest to this work. We do not have any commercial or associative interest that represents a conflict of interest in connection with the manuscript submitted.

REFERENCES

[1] Stefanou, Marcus S., and John P. Kerekes. "Image-derived prediction of spectral image utility for target detection applications." *IEEE Transactions on Geoscience and Remote Sensing* 48.4 (2010): 1827-1833.

[2] Manolakis, Dimitris, and Gary Shaw. "Detection algorithms for hyperspectral imaging applications." *IEEE signal processing magazine* 19.1 (2002): 29-43.

[3] X. Jia, B.-C. Kuo, and M. Crawford, "Feature mining for hyperspectral image classification," *Proceedings of the IEEE,* 101(3): 676-697, 2013.

[4] J. Yin, C. Gao and X. Jia, "A new target detector for hyperspectral data using cointegration theory," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6(2): 638 – 643, 2013.

[5] Reed, Irving S., and Xiaoli Yu. "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38.10 (1990): 1760-1770.

[6] Reed, I.S.; Yu, X. Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution. IEEE Trans. Acoust. Speech Signal Process. 1990, 38, 1760–1770.

[7] Eismann, Michael T., Alan D. Stocker, and Nasser M. Nasrabadi. "Automated hyperspectral cueing for civilian search and rescue." *Proceedings of the IEEE* 97.6 (2009): 1031-1055.

[8] Carlotto, Mark J. "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.

[9] Guo, Qiandong, et al. "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.6 (2014): 2351-2366.

[10] Schaum, A. P. "Hyperspectral anomaly detection beyond RX." *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII*. Vol. 6565. International Society for Optics and Photonics, 2007.

[11] Kwon, Heesung, and Nasser M. Nasrabadi. "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery." *IEEE Transactions on Geoscience and Remote Sensing* 43.2 (2005): 388-397.

[12] Zhao, Rui, Bo Du, and Liangpei 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.

[13] Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller. "Kernel principal component analysis." *International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 1997.

[14] Banerjee, Amit, Philippe Burlina, and Chris Diehl. "A support vector method for anomaly detection in hyperspectral imagery." *IEEE Transactions on Geoscience and Remote Sensing* 44.8 (2006): 2282-2291.

[15] Veracini, Tiziana, et al. "Nonparametric framework for detecting spectral anomalies in hyperspectral images." *IEEE Geoscience and Remote Sensing Letters* 8.4 (2011): 666-670.

[16] Shlens, Jonathon. "A tutorial on principal component analysis." *arXiv preprint arXiv:1404.1100* (2014).

[17] Green, Andrew A., et al. "A transformation for ordering multispectral data in terms of image quality with implications for noise removal." *IEEE Transactions on geoscience and remote sensing* 26.1 (1988): 65-74.

[18] Hyvarinen, Aapo. "Fast and robust fixed-point algorithms for independent component analysis." *IEEE transactions on Neural Networks* 10.3 (1999): 626-634.

[19] Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." *Neural networks* 13.4-5 (2000): 411-430.

[20] Wang, Jing, and Chein-I. Chang. "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis." *IEEE transactions on geoscience and remote sensing* 44.6 (2006): 1586-1600.

[21] Zhang, Junping, and Fengyang Zhu. "Target detection approach for hyperspectral imagery based on independent component analysis and local singularity." *Natural Computation, 2009. ICNC'09. Fifth International Conference on*. Vol. 2. IEEE, 2009.

[22] Tenenbaum, Joshua B., Vin De Silva, and John C. Langford. "A global geometric framework for nonlinear dimensionality reduction." *Science* 290.5500 (2000): 2319-2323.

[23] Roweis, Sam T., and Lawrence K. Saul. "Nonlinear dimensionality reduction by locally linear embedding." *Science* 290.5500 (2000): 2323-2326.

[24] Saul, Lawrence K., and Sam T. Roweis. "Think globally, fit locally: unsupervised learning of low dimensional manifolds." *Journal of machine learning research* 4.Jun (2003): 119-155.

[25] Belkin, Mikhail, and Partha Niyogi. "Laplacian eigenmaps for dimensionality reduction and data representation." *Neural computation* 15.6 (2003): 1373-1396.

[26] Zhang, Zhenyue, and Hongyuan Zha. "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment." *SIAM journal on scientific computing* 26.1 (2004): 313-338.

[27] Harsanyi, Joseph C., and C-I. Chang. "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach." *IEEE Transactions on geoscience and remote sensing* 32.4 (1994): 779-785.

[28] Johnson, Robert J., Jason P. Williams, and Kenneth W. Bauer. "AutoCAD: An improved ICA-based hyperspectral anomaly detection algorithm." *IEEE Transactions on Geoscience and Remote Sensing* 51.6 (2013): 3492-3503.

[29] He, Lin, et al. "Anomaly detection in hyperspectral imagery based on maximum entropy and nonparametric estimation." *Pattern Recognition Letters* 29.9 (2008): 1392-1403.

[30] Hyvärinen, Aapo, and Erkki Oja. "Independent component analysis: algorithms and applications." *Neural networks* 13.4-5 (2000): 411-430.

[31] Hyvarinen, Aapo. "Fast and robust fixed-point algorithms for independent component analysis." *IEEE transactions on Neural Networks* 10.3 (1999): 626-634.

[32] DONG, Chao, et al. "Hyperspectral image anomaly detection based on local orthogonal subspace projection [J]." *Optics and Precision Engineering* 8 (2009): 037.

[33] Han, Tian, and David G. Goodenough. "Nonlinear feature extraction of hyperspectral data based on locally linear embedding (LLE)." *Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. Proceedings. 2005 IEEE International*. Vol. 2. IEEE, 2005.

[34] Matteoli, Stefania, et al. "A locally adaptive background density estimator: An evolution for RX-based anomaly detectors." *IEEE geoscience and remote sensing letters* 11.1 (2014): 323-327.

[35] Cremers, Daniel, Timo Kohlberger, and Christoph Schnörr. "Shape statistics in kernel space for variational image segmentation." *Pattern Recognition* 36.9 (2003): 1929-1943.

[36] Matteoli, Stefania, et al. "Models and methods for automated background density estimation in hyperspectral anomaly detection." *IEEE Transactions on Geoscience and Remote Sensing* 51.5 (2013): 2837-2852.

[37] Chiang, S-S., C-I. Chang, and Irving W. Ginsberg. "Unsupervised target detection in hyperspectral images using projection pursuit." *IEEE Transactions on Geoscience and Remote Sensing* 39.7 (2001): 1380-1391.

[38] Taghipour, Ashkan, and Hassan Ghassemian. "Hyperspectral anomaly detection using attribute profiles." *IEEE Geoscience and Remote Sensing Letters* 14.7 (2017): 1136-1140.

[39] Liu, Yongjian, Guoming Gao, and Yanfeng Gu. "Tensor matched subspace detector for hyperspectral target detection." *IEEE Transactions on Geoscience and Remote Sensing* 55.4 (2017): 1967-1974.

[40] U.S. Army Corps of Engineers. [Online]. Available: http://www.tec.army.mil/Hypercurbe

[41] Kerekes, John. "Receiver operating characteristic curve confidence intervals and regions." *IEEE Geoscience and Remote Sensing Letters* 5.2 (2008): 251-255.

[42] Borghys, Dirk, et al. "Comparative evaluation of hyperspectral anomaly detectors in different types of background." *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*. Vol. 8390. International Society for Optics and Photonics, 2012.

[43] Zhang, Yuxiang, Bo Du, and Liangpei Zhang. "A sparse representation-based binary hypothesis model for target detection in hyperspectral images." *IEEE Transactions on Geoscience and Remote Sensing* 53.3 (2015): 1346-1354.

[44] Li, Wei, and Qian Du. "Collaborative representation for hyperspectral anomaly detection." *IEEE Transactions on geoscience and remote sensing* 53.3 (2015): 1463-1474.

[45] Chang, Chein-I., and Qian Du. "Estimation of number of spectrally distinct signal sources in hyperspectral imagery." *IEEE Transactions on geoscience and remote sensing* 42.3 (2004): 608-619.

[46] Xu, Yang, et al. "Anomaly detection in hyperspectral images based on low-rank and sparse representation." *IEEE Transactions on Geoscience and Remote Sensing* 54.4 (2016): 1990-2000.

HIGHLIGHTS

1. A hyperspectral anomaly detection framework with anomalous component extraction based on ICA and OSP is proposed, which discriminate the anomaly from backgrounds efficiently.

2. The anomaly detection performance of ICA is improved by extracting the brightest anomalous component as the initialized projection vector w_0, which is different from initializing w_0 randomly in original ICA method.

3. The process of abnormal components extraction is refined by OSP iteratively, which makes the detection result more accurate.

4. This algorithm yields an outstanding performance over other methods.