

# Hyperspectral Anomaly Detection With Kernel Isolation Forest

Shutao Li<sup>✉</sup>, *Fellow, IEEE*, Kunzhong Zhang, Puhong Duan<sup>✉</sup>, *Student Member, IEEE*, and Xudong Kang<sup>✉</sup>, *Senior Member, IEEE*

**Abstract**—In this article, a novel hyperspectral anomaly detection method with kernel Isolation Forest (iForest) is proposed. The method is based on an assumption that anomalies rather than background can be more susceptible to isolation in the kernel space. Based on this idea, the proposed method detects anomalies as follows. First, the hyperspectral data are mapped into the kernel space, and the first  $K$  principal components are used. Then, the isolation samples in the image are detected with the iForest constructed using randomly selected samples in the principal components. Finally, the initial anomaly detection map is iteratively refined with locally constructed iForest in connected regions with large areas. Experimental results on several real hyperspectral data sets demonstrate that the proposed method outperforms other state-of-the-art methods.

**Index Terms**—Anomaly detection, hyperspectral image (HSI), Isolation Forest (iForest), kernel method.

## I. INTRODUCTION

HYPERSPECTRAL remotely sensed images contain hundreds or even thousands of spectral bands, which provides a powerful tool for many remote sensing applications, such as anomaly detection [1], [2], image visualization [3], [4], and classification of the land covers [5], [6]. Among these applications, hyperspectral anomaly detection has received considerable interest in recent years due to its importance in military defense, search-and-rescue, mine detection, and environmental monitoring. Hyperspectral anomaly detection aims at distinguishing the interesting targets that are very different spatially or spectrally from their surrounding background with no a prior knowledge about the spectral signatures of background and target. In essence, anomaly detection is usually regarded as a binary classification problem, which classifies the pixel under test as the anomaly class or the background class.

Manuscript received March 12, 2019; revised June 19, 2019 and July 11, 2019; accepted August 9, 2019. Date of publication September 11, 2019; date of current version December 27, 2019. This work was supported in part by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001, in part by the National Natural Science Foundation of China under Grant 61871179 and Grant 61601179, in part by the Science and Technology Plan Project Fund of Hunan Province under Grant CX2018B171, Grant 2017RS3024, and Grant 2018TP1013, and in part by the Science and Technology Talents Program of the Hunan Association for Science and Technology under Grant 2017TJ-Q09. (*Corresponding author: Xudong Kang*)

The authors are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Hunan University, Changsha 410082, China (e-mail: shutao\_li@hnu.edu.cn; kunzhong@hnu.edu.cn; puhong\_duan@hnu.edu.cn; xudong\_kang@163.com).

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

Digital Object Identifier 10.1109/TGRS.2019.2936308

To achieve this objective, various algorithms have been proposed for detecting anomaly targets in hyperspectral images (HSIs) over the last 20 years. Statistical model-based probability distribution is the first and the most prevalent category in hyperspectral anomaly detection [7]–[10]. One of the most well-known methods is Reed–Xiaoli (RX) algorithm [11], developed by Reed and Xiaoli, which is considered the benchmark method. The RX detector assumes that the background is well modeled with a Gaussian multivariate distribution. Based on the assumption, the background can be characterized with the mean vector and the covariance matrix of the pixels in the whole image or a local window surrounding the test pixel. Then, the anomaly pixels are detected by measuring the Mahalanobis distance between the test pixel and the local background. It is worth mentioning that the pixels modeled the background can include the whole image or local background, thus leading to global detector [12] or local detector [13], respectively. However, most real-world HSIs cannot be modeled by a homogeneous distribution, which often cover different classes of materials and exhibit complex background. To address this problem, a variety of improved RX-based anomaly detectors have been proposed. For example, the subspace RX anomaly detector [14] models the background with representative eigenvectors of the HSI covariance matrix, which eliminates the undesired background signatures in HSI. The kernel-RX anomaly detector is a nonlinear version of the RX detector developed by Kwon and Nasrabadi [15], which characterizes more adequately the normal model in a higher dimensional feature space by using the kernel method. The cluster-based anomaly detector (CBAD) [16] first performs clustering methods on the HSI, thus yielding several Gaussian clusters that are spectrally homogeneous. Then, the anomalies are detected in each cluster.

Besides the probability distribution-based methods, there are many other types of anomaly detection methods in the literature. For example, the support vector data description (SVDD)-based anomaly method [17] is a deterministic support vector approach that avoids prior assumptions about the data distribution. The SVDD is capable of estimating the support region where most of the data lie, which avoids the much more enormous challenge of estimating the underlying probability density function. The method is based on the assumption that a pixel in background can be approximately represented by a linear combination of its surrounding pixels, while the anomaly cannot. The contribution of each surrounding pixel is adjusted by a distance-weighted regularization matrix. Recently, a novel attribute and edge-preserving filtering-based

detection method (AED) [18] has been proposed. Different from previous works, the method only requires local filtering operations (i.e., attribute and edge preserving), and thus, it can be implemented very efficiently. Moreover, a multi-scale version of the AED has been developed in [19], which considered the multi-scale information in HSIs.

Additionally, sparse representation has received considerable attention for hyperspectral anomaly detection. In [20], the anomaly detector is modeled by the use of background joint sparse representation (BJSR). The most representative background bases are adaptively selected for the local region, and then, the background pixels can be well represented with the selected background bases while the anomaly cannot. Based on a similar idea, sparse representation, collaborative representation, and tensor representation-based anomaly detection methods have been well investigated in recent years [21]–[26]. Recently, deep learning has been a very active research topic in the field of HSI processing [27], [28], and deep learning-based anomaly detector [29] also has been proposed.

Different from the aforementioned methods, a novel Kernel Isolation Forest-based hyperspectral anomaly Detection method (KIFD) is proposed in this article. From a data description aspect, the anomalies usually have different spectral values with respect to the background, and thus, these pixels are usually more susceptible to isolation than the background. By constructing Isolation Forest (iForest) using randomly selected samples of the HSI, those pixels that can be isolated easily in the kernel feature space are detected as anomalies. Furthermore, in this article, the detection is performed with a novel recursive detection framework, which can well make full use of the local and global information of HSIs. It should be mentioned that an earlier version of this article has been submitted to the 2019 IEEE International Geoscience and Remote Sensing Symposium. Specifically, the major contributions of this article can be concluded as follows.

- 1) A recursive anomaly detection framework is developed which can well make use of the global and local information of HSIs.
- 2) A KIFD is proposed for the first time which can well capture the isolation property of anomaly objects in HSIs. Experiments performed on several real hyperspectral data sets demonstrate the state-of-the-art performance of the proposed method.

The rest of this article is organized as follows. The iForest theory and kernel method is briefly reviewed in Section II. The proposed method is introduced in Section III. Experimental results are given in Section IV. Finally, conclusions are given in Section V.

## II. BACKGROUND

### A. Isolation Forest

iForest is firstly introduced by Liu *et al.* [30] for outlier detection in data mining. The advantage of iForest is that it can identify outliers directly based on the concept of isolation without using any metrics, this eliminates computational cost

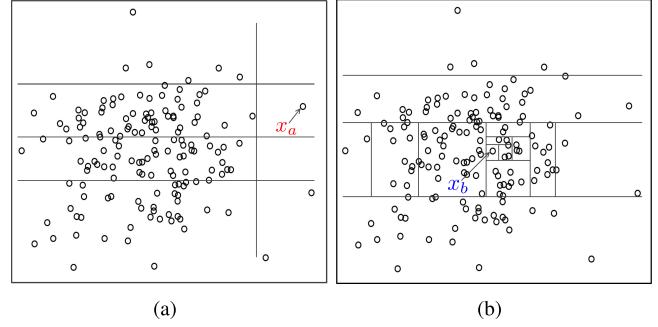


Fig. 1. Graph example illustrating the principle of iForest. Given a Gaussian distribution (a) anomalous instance is isolated through four random partitions and (b) normal instance is isolated through thirteen random partitions.

significantly compared to the distance-based [31] and density-based methods [32]. Here, the principle of iForest is briefly reviewed. The details of the iForest algorithm, we refer the reader to [30].

Specifically, iForest is based on the assumption that the **anomaly instances are usually rare and different from those of normal instances in a given data set, which makes them more susceptible to isolation in a number of binary tree structures than the normal instances**. A tree structure is constructed from the root to leaves by dividing the data at each node with a random selected feature and threshold. Each tree is grown until each instance is isolated in a leaf. The path length for an instance, also known as the isolation depth, is defined as that the instance traverses a binary tree from the root node until the traversal is terminated at a leaf node. Therefore, in randomly generated binary tree where instances are recursively partitioned, anomalous instances have a quick arrival to leaf nodes, while nominal instances require many more splits to finally reach leaf nodes. As a whole, anomalous instances have noticeable shorter average path lengths than those of normal instances over a number of binary trees. The **anomaly score assigned to an instance can be calculated using the average isolation depth across the binary trees**.

To demonstrate the fact that anomalies have higher susceptibility to be isolated under random partitioning, a graphical interpretation for a 2-D normally distributed data set is represented as in Fig. 1. It can be seen that the inlier \$x\_b\$ (i.e., normal instance) generally requires more separating lines to be isolated, while the outlier \$x\_a\$ (i.e., anomalous instance) generally requires less separating lines to be isolated. In this example, every partitioning process is generated by randomly selecting an attribute and randomly selecting a cut-point between the minimum and maximum values of the selected attribute. Since a binary tree is generated by recursive partitioning, the path length from the root node to a leaf node can represent the number of partitions required to isolate a point. Thus, the path length of \$x\_a\$ is shorter than the path length of \$x\_b\$.

iForest was proposed almost ten years ago. In recent years, the algorithm has been successfully applied in real applications, such as detecting anomalous taxi trajectories or detecting the anomalous data collected from wireless sensor networks [33]–[35]. To our best knowledge, the method has not been

researched in remote sensing applications. In this article, we develop an iForest method to detect anomalies in HSIs.

### B. Kernel Methods

In recent years, kernel methods have received considerable attention in the machine-learning community, and various powerful kernel-based methods such as kernel support vector machines (SVMs) [36], kernel fisher discriminant (KFD) [37] analysis, and kernel principal component analysis (KPCA) [38], [39], have been researched. The major reason is that kernel methods can be used to project the input data into a higher dimensional feature spaces so as to increase the computational capability of linear machines in an easy way. There is no constraint on the form of the mapping function, which would lead to infinite-dimensional linear spaces.

Given a data set  $A = \{a_1, a_2, \dots, a_n\} \in \Omega$ , if the classes in the data set are not linearly separable, then the kernel method can map the data into a potentially much higher dimensional feature space  $\mathcal{F}$  in the hope that in this higher dimensional space the classes become linearly separable. The kernel function  $\mathbf{k}$  is defined as an inner product in the feature space  $\mathcal{F}$  and is usually defined as follows:

$$\mathbf{k}(a_i, a_j) = \langle \Phi(a_i), \Phi(a_j) \rangle, \quad i, j = 1, \dots, n \quad (1)$$

where  $\Phi$  is a nonlinear mapping (or sometimes linear mapping) from the original input space  $\Omega$  to the feature space  $\mathcal{F}$ , and  $\langle \cdot, \cdot \rangle$  is an inner product. Commonly, the kernel function  $\mathbf{k}$  is defined directly, e.g., the radial basis function (RBF) kernel  $\mathbf{k}(a_i, a_j) = \exp(-\gamma \|a_i - a_j\|^2)$  with an adjustable parameter  $\gamma$  controlling the width of the RBF. The map  $\Phi$  and the feature space  $\mathcal{F}$  is defined implicitly, which is replaced with the kernel function  $\mathbf{k}$ . Through using the kernel function, the original input data can be mapped into the feature space of a higher dimension, which enables the algorithm to more satisfy the model requirements. Therefore, by providing a bridge between linear and nonlinear, kernel functions provide a more efficient solution for any algorithm that can be written in inner product. Please refer to [40] for a complete theoretical description of the kernel approach.

## III. PROPOSED APPROACH

### A. Constructing Isolation Forest

Here,  $\mathbf{X} \in \mathbb{R}^{D \times N}$  denotes the input HSI, where  $D$  represents the number of spectral bands and  $N$  is the number of pixels. In order to construct the iForest globally,  $M$  pixels are first randomly selected from the hyperspectral data. Then, the selected pixels can be divided into two child nodes, i.e., left node  $\mathbf{M}_l$  and right node  $\mathbf{M}_r$ , based on a simple decision rule. Specifically, if  $\mathbf{X}_s^m$  is smaller than  $\theta$ , the  $m$ th selected pixel will be divided into the left node, and vice versa. Here,  $s$  is a number randomly selected from 1 to  $D$ ,  $\theta$  is a value randomly selected between the minimum value and maximum value of the  $s$ th band, i.e.,  $\mathbf{X}_s$ . Next, the child nodes could be further divided by performing the operations above iteratively, until either: 1) the tree reaches a height limit  $H_{\max}$ ; 2) the number of pixels in each child node equals 1; and 3) the pixels in each child node have the same values. It should be mentioned that

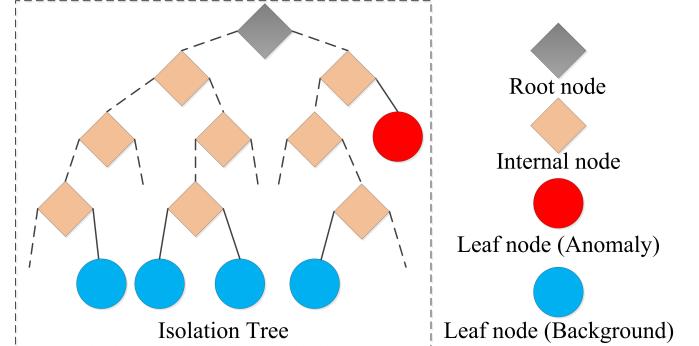


Fig. 2. Graph example illustrating the structure of an Isolation Tree.

$s$  and  $\theta$  are both randomly selected for all the internal nodes. In this article,  $H_{\max} = \log_2 M$  is set as the default parameter setting. The tree construction process can be repeated  $q$  times so as to construct the iForest, and  $q$  is set as 1000, which means that the forest consists of 1000 trees.

Fig. 2 gives a graph example illustrating the structure of an Isolation Tree. It can be seen that the Isolation Tree is actually a tree structure representation of the randomly selected samples, in which each node represents a single hyperspectral pixel or a number of hyperspectral pixels with similar spectral values. Furthermore, it is observed that most of the selected samples actually belonging to the background, which are usually located in the bottom of the tree. The reason is that background pixels appear much more frequently with respect to anomaly pixels, and thus, hard to be isolated. This observation also has been illustrated in Fig. 1.

### B. Detecting the Anomalies Using Isolation Forest

In this step, each pixel in the HSI is fed into the constructed iForest so as to detect the anomalies. Specifically, in an isolation tree, the path length of a pixel is defined as the number of edges that the pixel travels from the root to an external node. Since anomaly pixels usually appear with small area and distinct spectral signatures which are different from background, they are easily isolated to external nodes, and have shorter path lengths. By contrast, background pixels will tend to have longer path lengths in the constructed isolation trees. Based on this fact, path length can be employed as a measure to detect anomalies. It should be noted that a pixel may have different path lengths on different isolation trees. Therefore, the proposed isolation-based anomaly detection method is based on an ensemble learning mechanism, in which the final path length for each pixel is obtained by averaging path lengths of different isolation trees.

An iForest consists of  $q$  isolation trees  $\{Q_1, Q_2, \dots, Q_q\}$ . Let  $h_i(x)$  denotes the length of a test pixel  $x \in \mathbf{X}$  in isolation tree  $Q_i$ , the average path length over  $q$  isolation trees is calculated as follows:

$$E(h(x)) = \frac{1}{q} \sum_{i=1}^q h_i(x). \quad (2)$$

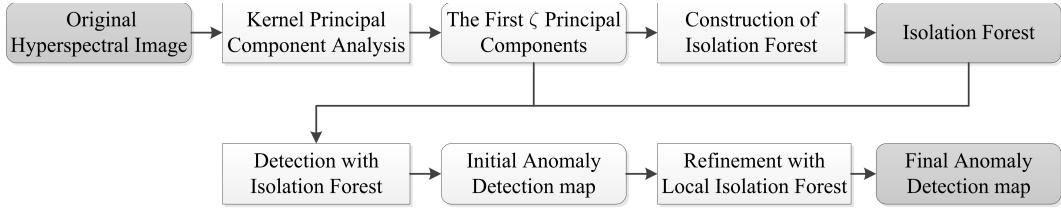


Fig. 3. Flowchart of the proposed KIFD method.

Then, the anomaly score  $s \in (0, 1]$  of a test pixel  $x$  is defined as follows:

$$s(x) = 2^{-\frac{E(h(x))}{c(M)}} \quad (3)$$

where  $M$  denotes the subsampling size.  $c(M) = 2H(M - 1) - 2(M - 1)/M$ , in which  $H(M)$  is the harmonic number that calculated by  $\ln(M) + 0.5772156649$  (Euler's constant). By performing the operations mentioned above to each hyperspectral pixel, an anomaly map  $\mathbf{D}$  can be obtained.

#### C. Refining the Detection Map With Local Isolation Forest

The global iForest-based detection may produce a number of false alarms since the forest is constructed using the randomly selected pixels in the whole scene. To make full use of the local information, the initial anomaly detection map  $\mathbf{D}$  is further refined by performing the IFD in a local processing way. The specific steps are concluded as follows.

- 1) *Binarization*: The anomaly detection map  $\mathbf{D}$  is converted into a binary image

$$\mathbf{B}_i = \begin{cases} 1, & \text{if } \mathbf{D}_i > \delta \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

where the threshold value  $\delta$  is selected using the Otsu's method [41].

- 2) *Local iForest Detection*: Generally, anomalies usually appear with small areas. Given the binary image  $\mathbf{B}$ , the connected components (CCs) in  $\mathbf{B}$  can be obtained, which represent the detected anomaly objects. Given an area threshold  $\alpha$ , if the area of a CC is larger than  $\alpha$ , a local iForest will be constructed with half of the randomly selected pixels in this CC, and the anomaly scores of pixels in this CC could be reevaluated by the constructed local iForest. In this article, the optimal value for  $\alpha$  is set to be  $N/120$ , which is adjusted with the number of pixels in the HSI.
- 3) *Termination of the Detection Process*: By performing steps 1 and 2 iteratively, the anomaly scores could be refined until there are not CCs with an area larger than  $\alpha$ , and the final anomaly detection map  $\hat{\mathbf{D}}$  can be obtained.

The proposed method could be directly performed in the data space or in a nonlinear kernel space. Specifically, instead of performing the detection on the original HSI. The hyperspectral data set  $\mathbf{X}$  is first transformed using the KPCA method. The first  $\zeta$  principal components are then used as the input of the IFD detection algorithm. Experimental results demonstrate that this preprocessing step makes an important

role in further improving the detection performance. For better understanding, the flowchart of the proposed KIFD method is shown in Fig. 3. As shown in this figure, the proposed method consists of the following major steps. First, the dimension of the input image is reduced with KPCA. Then, the iForest is constructed so as to obtain an initial anomaly detection map. Finally, the final anomaly detection map is obtained by refining the initial anomaly detection map with the local iForest.

## IV. EXPERIMENTS

### A. Data Sets

In this article, the proposed method is evaluated on four real hyperspectral data sets captured at different scenes, which are listed as follows.

- 1) *San Diego Data Set*: The first hyperspectral data set was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) covering the San Diego airport area, CA, USA. This scene covers an area of  $100 \times 100$  pixels with 224 spectral channels in wavelengths ranging from 370 to 2510 nm. In the experiments, a total of 189 bands are used after removing the water-absorption bands. The spatial resolution is approximately 3.5 m. The three airplanes are the anomalies to be detected in the image. The sample image and ground truth map for this data set are shown in Fig. 4(a) and (b), respectively.
- 2) *Hyperspectral Digital Imagery Collection Experiment (HYDICE) Data Set*: The second data set was collected by the HYDICE airborne sensor over an urban area, CA, USA. This urban scene consists of  $80 \times 100$  pixels, with 175 spectral channels in wavelengths ranging from 400 to 2500 nm. The image has a spatial resolution of 1 m. The scene mainly consists of a vegetation area, a construction area, and several roads including some cars. The man-made objects, i.e., cars and roofs are regarded as anomaly objects. The sample image and ground truth map for this data set are shown in Fig. 5(a) and (b), respectively.
- 3) *El Segundo Data Set*: The third data set was captured by the AVIRIS sensor, which has 224 spectral channels in wavelengths ranging from 366 to 2496 nm. This urban scene covers an area of El Segundo, CA, USA, with the spatial size of  $250 \times 300$ , as shown in Fig. 6(a). Each pixel has 7.1 m of ground resolution. The image data set is mainly composed of an area of oil refinery, several residential areas, parks, and one school zone. The constructions of oil refinery such as storage tanks and

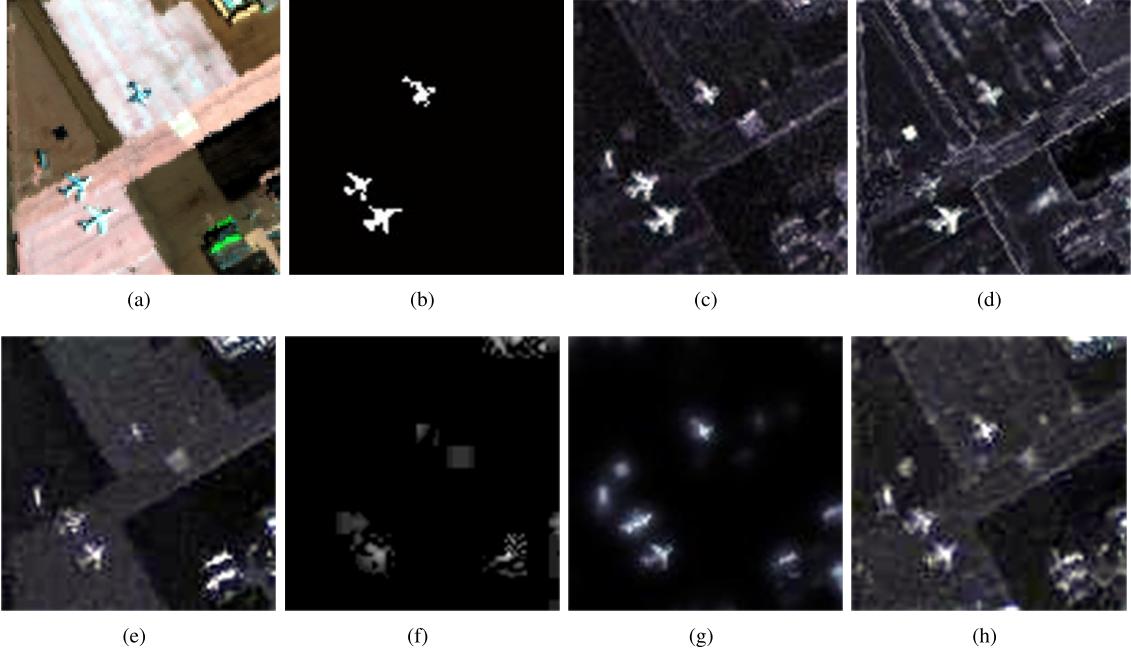


Fig. 4. San Diego data set. (a) Pseudocolor image, (b) ground truth map, and the detection maps of (c) KIFD, (d) IFD, (e) RX [11], (f) LRX [11], (g) AED [18], and (h) CRD [22].

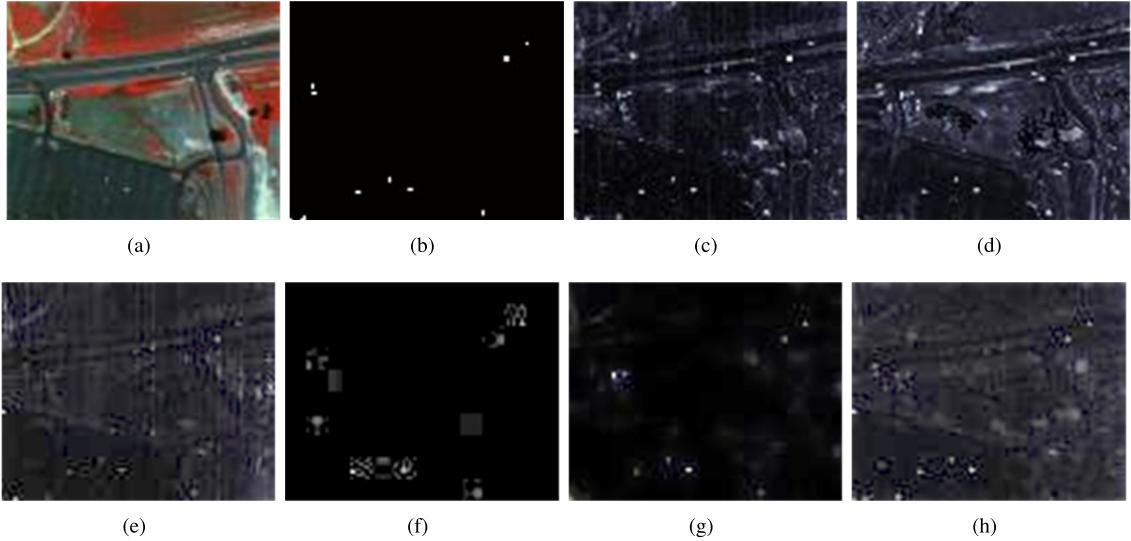


Fig. 5. HYDICE data set. (a) Pseudocolor image, (b) ground truth map, and the detection maps of (c) KIFD, (d) IFD, (e) RX [11], (f) LRX [11], (g) AED [18], and (h) CRD [22].

the towers are considered as anomaly targets, as shown in Fig. 6(b).

- 4) *Grand Isle Data Set*: The fourth data set is an AVIRIS image, which was captured at the location of Grand Isle on the Gulf Coast, part of Jefferson Parish, LA, USA. The data set consists of  $300 \times 480$  pixels with 224 spectral channels in wavelengths ranging from 366 to 2496 nm. The spatial resolution is approximately 4.4 m. The main background materials are island and water in the scene. Those man-made objects in the water are selected as the anomalies to be detected. The false color image of the four data set and its ground truth are shown in Fig. 7(a) and (b), respectively.

Furthermore, all data sets have been made available on corresponding author's homepage.<sup>1</sup>

#### B. Parameter Analysis

In this section, we analyze the influence of the parameter  $\zeta$  on the detection performance of the proposed KIFD method. The parameter  $\zeta$  controls the number of principal components. Fig. 8 shows the effect of the parameter  $\zeta$  over the area under the receiver operating characteristic (ROC) curve (AUC) values of the KIFD on each data set. We can observe that AUC values tend to grow gradually when  $\zeta$  varies from

<sup>1</sup><http://www.escience.cn/people/xudongkang>

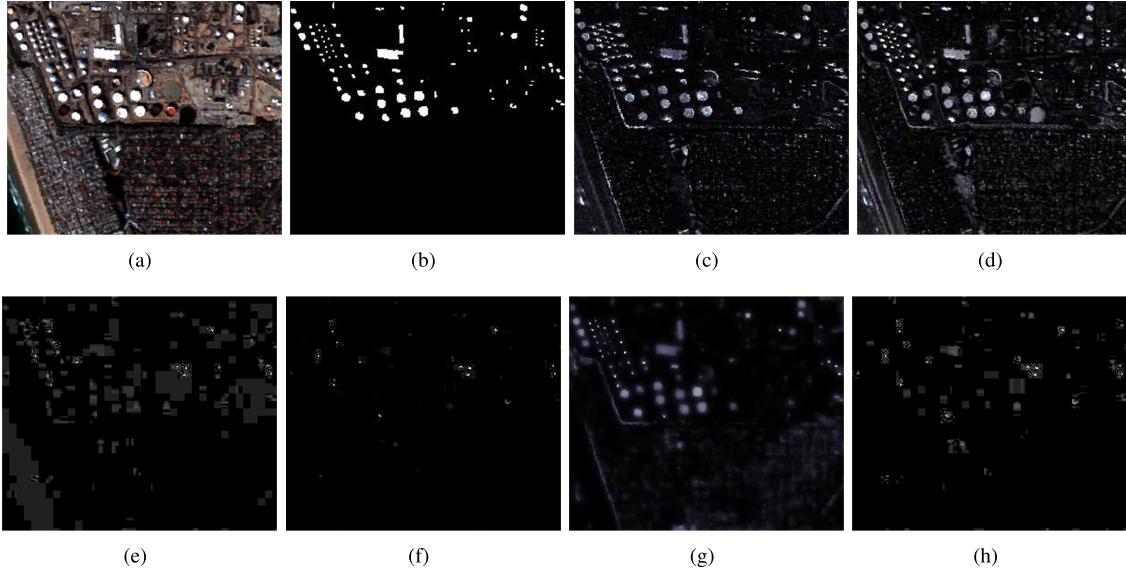


Fig. 6. EI Segundo data set. (a) Pseudocolor image, (b) ground truth map, and the detection maps of (c) KIFD, (d) IFD, (e) RX [11], (f) LRX [11], (g) AED [18], and (h) CRD [22].

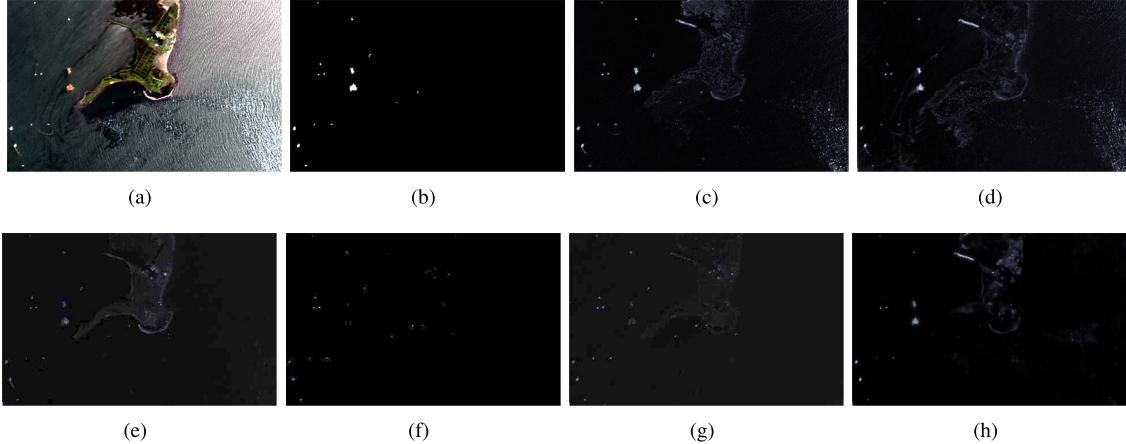


Fig. 7. Grand Isle data set. (a) Pseudocolor image, (b) ground truth map, and the detection maps of (c) KIFD, (d) IFD, (e) RX [11], (f) LRX [11], (g) AED [18], and (h) CRD [22].

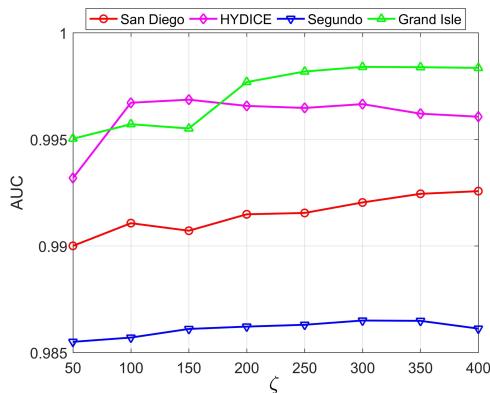


Fig. 8. Influence of the parameter  $\zeta$  on the detection performance of the proposed KIFD on each data set.

50 to 300 for all the data sets, except for the HYDICE data set. For the HYDICE data set, it can be seen that when  $\zeta$  increases from 50 up to 150, AUC values tend to

grow gradually. When  $\zeta$  is greater than 150, the AUC values tend to be stable and may decrease slightly. In a nutshell, when  $\zeta = 300$ , the proposed KIFD can obtain a satisfactory detection performance. Therefore, in this article, the default parameter of the KIFD is set as  $\zeta = 300$ .

### C. Analysis the Influence of the Global and Local Isolation Forest-Based Detection Steps

In this section, we analyze the influence of the global and local iForest-based detection steps on the performance of the proposed method. As shown in Fig. 9(a), it can be observed that the Global iForest detects the suspicious anomaly pixels in HSI, and the AUC score of the initial detection result is 0.9835. To make full use of the local information in the HSI, the initial anomaly detection map is further refined with iterative Local iForest. By this step, some false alarms could be effectively removed. Fig. 9(b) shows the final detection map, and the AUC score of the final detection map is 0.9917. Based on this experiment, it can be concluded that both the

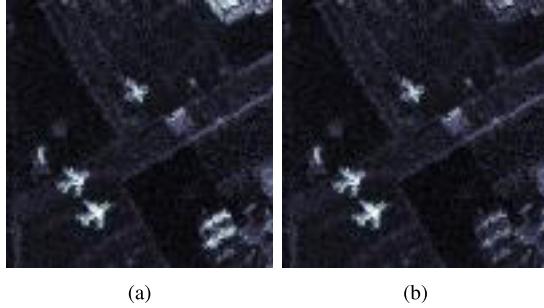


Fig. 9. San Diego data set. (a) Initial detection map of Global iForest. (b) Refining detection map with Local iForest.

global and local iForests make an important role in detecting the anomaly pixels.

#### D. Detection Performance

In this section, the anomaly detection performance of the proposed IFD and KIFD are evaluated and compared with four state-of-the-art detectors: the RX detector [11], local RX detector (LRX) [11], the AED [18], and Collaborative Representation-based Detector (CRD) [22]. The RX and LRX detectors are most widely used as the benchmark methods for comparing the performance of a new anomaly detector. The CRD and AED were recently developed AD methods, which have leading performances on several real hyperspectral data sets.

In the experiments, three of the most widely used metrics for anomaly detection evaluation are exploited to qualitatively and quantitatively evaluate the detection performances of those compared methods. The first metric is the ROC curve, which describes the relationship between the probability of detection (PD) and the false alarm rate (FAR) at various threshold settings based on the ground truth. Specifically, given a detection map and a ground truth map, the PD and FAR can be calculated as follows:

$$PD = \frac{N_D}{N_T} \quad FAR = \frac{N_F}{N} \quad (5)$$

where  $N_D$  is the number of detected target pixels under a certain threshold for the detection map and  $N_T$  is the total number of real target pixels in the image,  $N_F$  is the number of false alarm pixels and  $N$  is the number of pixels in the HSI. By comparing with the ROC curves of different methods, if a method has a higher PD than the other methods at the same false alarm, it demonstrates that the method outperforms the other methods. The second metric is the AUC, which is derived from the ROC curve and calculated with the whole area under ROC curve. A larger AUC value means the detector obtains a better detection performance. More detailed illustration of these two evaluation indexes can be found in [42]. The third metric is the separability range, which describes the ability of detector to separate anomalies from the background [43]. A better detector would obtain a more efficient separation of the anomalies and background.

TABLE I  
OPTIMAL PARAMETERS OF THE LRX, AED, AND CRD METHODS FOR THE EXPERIMENTAL DATA SETS

	LRX [11]		AED [18]	CRD [22]	
San Diego	$\omega_{in}$ 37x37	$\omega_{out}$ 55x55	$\kappa$ 25	$\omega_{in}$ 3x3	$\omega_{out}$ 31x31
HYDICE	$\omega_{in}$ 3x3	$\omega_{out}$ 5x5	$\kappa$ 5	$\omega_{in}$ 7x7	$\omega_{out}$ 15x15
Segundo	$\omega_{in}$ 9x9	$\omega_{out}$ 31x31	$\kappa$ 250	$\omega_{in}$ 15x15	$\omega_{out}$ 31x31
Grand Isle	$\omega_{in}$ 9x9	$\omega_{out}$ 11x11	$\kappa$ 250	$\omega_{in}$ 9x9	$\omega_{out}$ 11x11

In order to evaluate the detection performances of compared methods, we perform experiments on the above-mentioned four data sets and generate the results reported in Figs. 4–7. In the experiments, the optimal parameters of the LRX, AED, and CRD methods are selected for each data set according to the corresponding AUC performances. For example, the detection performance of AED method is sensitive to the area parameter  $\kappa$ . Then, based on the AUC metric, the optimal parameter  $\kappa$  ranging from 5 to 300 is selected optimally. In addition, it is known that the detection performances of the LRX and CRD methods are sensitive to the window sizes ( $\omega_{in}$  and  $\omega_{out}$ ). Similarly, the optimal window sizes  $\omega_{in}$  ranging from 3 to 41 and  $\omega_{out}$  ranging from 5 to 71 are selected optimally. The detailed optimal parameters of those compared methods are presented in Table I. For the proposed IFD and KIFD methods, the default parameters are set as  $M = 3\% \times N$  ( $N$  is the number of pixels in the HSI,  $q = 1000$ ). Furthermore, for the KIFD method, the number of principal components  $\zeta$  is set as 300.

First, the detection maps of the compared anomaly detection methods on the four real hyperspectral data sets are shown in Figs. 4–7. It can be seen that the proposed IFD method shows competitive performances with respect to the RX method, and the KIFD method can detect anomalies more accurately and clearly than the IFD method and other compared methods. For example, in Fig. 4, the KIFD method can accurately detect all airplanes and obtain a good extraction of the airplanes. The IFD and RX methods both have ability to detect all airplanes, but the shapes of some airplanes are missing. The LRX method can penalize background effectively, but cannot detect all anomalies. The AED and CRD methods also recognize the locations of airplanes, while the shapes of the airplanes are blurring and some false anomalies are also detected. It should be mentioned that some buildings that are quite different spectrally from their surrounding are considered as anomalies for all the compared methods. The reason is that some background pixels that have spectral difference with respect to their neighboring pixels would be also detected as anomalies. Moreover, the proposed KIFD and IFD methods show robust detection performance in large size scenes. As shown in Fig. 6, most of the anomaly pixels can be detected by the KIFD method. The reason is that, with the kernel method, the KIFD can achieve better background suppression and derive a detection result with less false alarms.

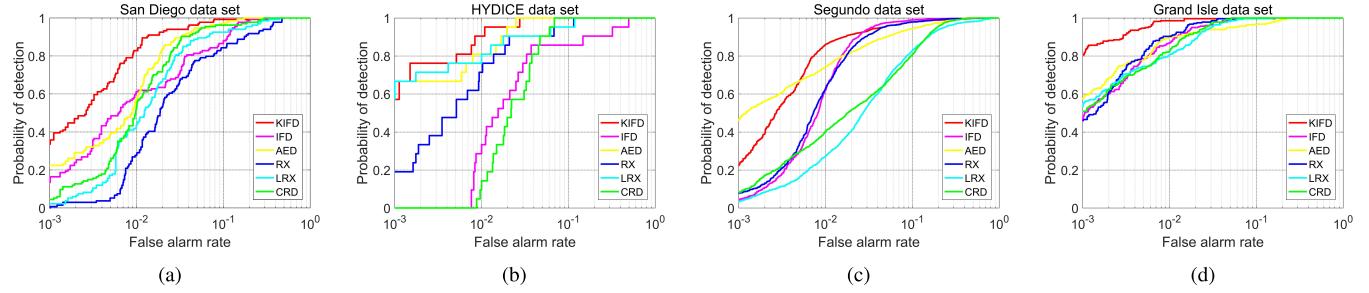


Fig. 10. ROC curves of the algorithms for (a) San Diego data set, (b) HYDICE data set, (c) Segundo data set, and (d) Grand Isle data set.

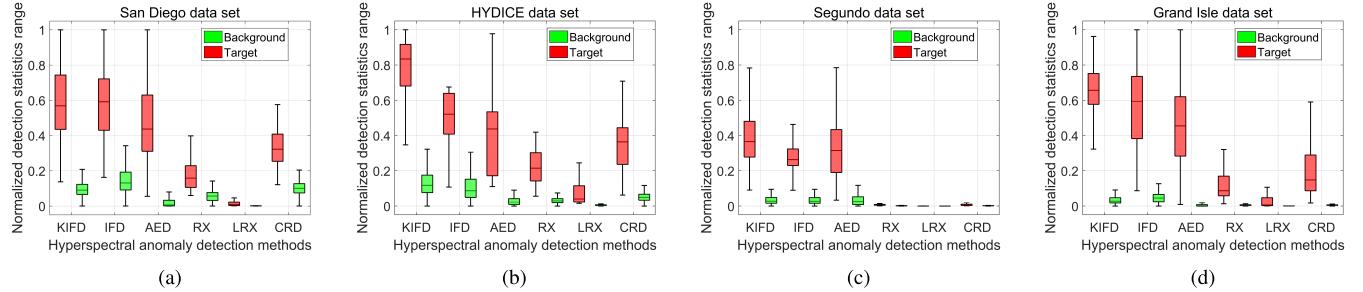


Fig. 11. Background-target separability maps of the algorithms for (a) San Diego data set, (b) HYDICE data set, (c) Segundo data set, and (d) Grand Isle data set.

TABLE II

AUC SCORES OF THE ALGORITHMS FOR THE EXPERIMENTAL DATA SETS

Dataset	KIFD	IFD	RX [11]	LRX [11]	AED [18]	CRD [22]
San Diego	<b>0.9917</b>	0.9713	0.9403	0.9678	0.9850	0.9781
HYDICE	<b>0.9965</b>	0.9417	0.9857	0.9890	0.9951	0.9725
Segundo	<b>0.9864</b>	0.9859	0.9841	0.9365	0.9804	0.9484
Grand Isle	<b>0.9983</b>	0.9951	0.9963	0.9936	0.9900	0.9949

By observing the detection results on other three real data sets, the following observations can be found. First, the RX usually has stable detection performances, but fail to detect all the anomalies due to the contamination of background statistical estimation. Second, since the detection performances of the LRX and CRD methods are sensitive to the window sizes, when the sizes and shapes of anomaly targets are relatively different and irregular, some of the anomalies cannot be detected [see Fig. 6(f) and (h)]. Third, due to the blurring effect produced in the filtering operation, those small-scale anomaly targets may be failed to be detected [see Figs. 5(g) and 6(g)].

Second, we also calculate the AUC scores of the compared algorithms as shown in Table II, in which the best AUC score is highlighted for each experimental data set. According to Table II, we can find that the KIFD achieves the best scores on all experimental data sets. The IFD also obtains satisfactory AUC scores, except for the HYDICE data set. The RX and CRD achieve the lowest detection accuracy for the San Diego data set and EI Segundo data set, respectively. Although the LRX and AED methods have relatively stable detection performances, it does not in any case obtain the highest AUC scores. In addition, Fig. 10 presents a quantitative comparison of different methods with the ROC curves. Note that for the ROC curve, a better detector would lie nearer the upper left corner (0, 1) and result in achieving higher

detection accuracy at the same false alarm. As shown in Fig. 10 ( $\log_{10}$  coordinate is used), the KIFD method is superior to the IFD, RX, LRX, AED and CRD methods under most conditions. It can be seen from Fig. 10(a) and (d) that the proposed KIFD method obtains much better ROC curves than the other compared methods since its probabilities of detection are always higher than others when FAR varies from 0.001 to 1. As shown in Fig. 10(b), it can be observed that the proposed KIFD method obtains higher probabilities of detection than the other compared methods under most conditions. As shown in Fig. 10(c), compared with the AED method, the proposed KIFD method obtains slightly lower probabilities of detection when the FAR varies from 0.001 to 0.005. However, the general performance of the proposed KIFD method is still better than the other algorithms.

Third, box plots are exploited to further evaluate the detection performance of proposed KIFD method via separability maps, as shown in Fig. 11. There are two boxes for each detector. The red boxes and green boxes illustrate the distributions of anomaly class and background class, respectively. The position of the boxes reflects the separability between anomaly class and background class. In other words, a larger separability distance means the corresponding detector achieves a superior detection performance. By observing Fig. 11, the following observations can be found. First, the separation gaps obtained by the proposed KIFD method are quite bigger than the RX, LRX and AED methods on all experimental data sets. Second, the separation gaps obtained by KIFD method are slightly larger than the IFD method on the experimental data sets. Third, the AED method achieves same separation gaps of the KIFD method on the San Diego and Segundo data sets. However, the KIFD method achieves superior background-anomaly separation than the AED method on the HYDICE and

TABLE III  
RUNNING TIME (SECONDS) OF THE ALGORITHMS FOR THE EXPERIMENTAL DATA SETS

Time (s)	KIFD	IFD	RX [11]	LRX [11]	AED [18]	CRD [22]
San Diego	85.44	42.86	0.19	136.48	0.29	2128.32
HYDICE	58.57	28.92	0.09	18.40	0.21	64.00
Segundo	206.94	151.69	0.96	1229.67	1.73	9567.85
Grand Isle	378.09	300.16	1.92	1042.89	3.35	1665.18

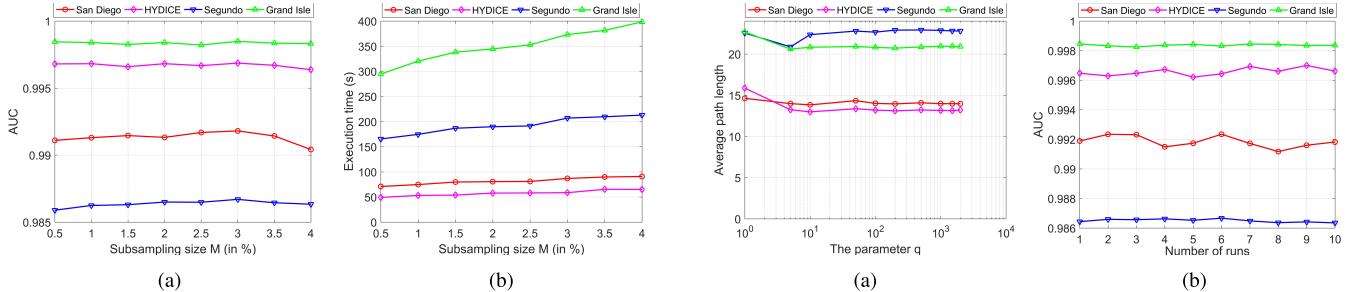


Fig. 12. (a) Influence of the parameter  $M$  on the detection performance of the proposed KIFD on each data set. (b) Influence of the parameter  $M$  on the execution time of the proposed KIFD on each data set.

Grand Isle data sets. Therefore, it can be concluded that the proposed KIFD method obtains a better overall background-anomaly separation than the other algorithms.

Finally, the time-complexity of constructing an iForest is  $O(qM^2)$ , where  $q$  is the number of tree,  $M$  is the subsampling size of tree. The time-complexity of evaluating anomaly scores is  $O(qMN)$ , where  $N$  is the number of pixels in the HSI. Therefore, the complexity of the proposed method is  $O(qM(M + N))$ . Also, the computation cost of the aforementioned methods is compared on a computer with 2.8-GHz CPU and 8-GB memory. We measure the computing time of the compared detection methods on the experimental data sets. All methods are implemented in MATLAB and the detailed results are presented in Table III. As shown in this table, it is important to notice that the KIFD is computationally more expensive than the IFD as expected, while is much more efficient than the LRX and CRD methods.

### E. Discussion

In this section, we undertook the experiments to show how the randomness of proposed KIFD method affects to results. Fig. 12 presents the influence of different subsampling sizes  $M$  to the detection performance and the computing time on each data set. As shown in Fig. 12(a), it can be seen that the AUC values increase at first, then decrease slightly, and finally reach their peaks at 3 for the San Diego and Segundo data sets. For the HYDICE and Grand Isle data sets, we can see that the proposed KIFD obtains a steady detection performance when  $M$  is less than 3. However, when  $M$  continues to increase, the AUC values of proposed KIFD start to decrease. Therefore, in this article, **we select  $M = 3\% \times N$  ( $N$  is the number of pixels in the HSI) as the default parameter of the proposed method**. As shown in Fig. 12(b), execution time of the proposed KIFD method achieves nearly linear growth when subsampling size  $M$  increases. In this article,

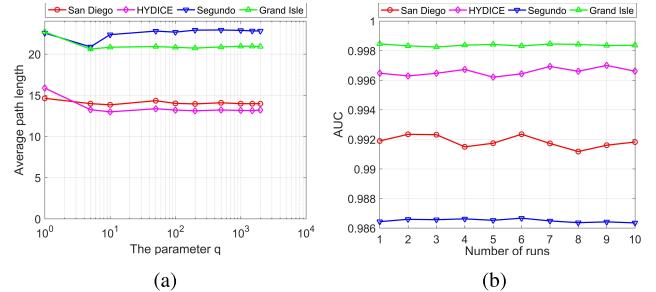


Fig. 13. (a) Influence of the parameter  $q$  to the average path length of the proposed KIFD on each data set. (b) Influence of the randomness selection step on the proposed KIFD method.

anomaly score of each hyperspectral pixel is evaluated by the average path length over  $q$  trees. The average path length would be more desirable when  $q$  is large enough. Fig. 13(a) presents the effect of  $q$  to the average path length of the proposed KIFD on each data set. Here, we can observe that the average path lengths have a trend to drop first and then rise when  $q$  varies from 1 to 1000. When  $q$  continues to increase, the average path lengths converge. Therefore, in this article, **we select  $q = 1000$  as the default parameter of the proposed method**. Finally, Fig. 13(b) presents the influence of this randomness selection step on the proposed KIFD method. It can be clearly observed that the AUC values of the proposed method vary in a small range in ten repeated experiments. For example, the AUC values vary within the range between 0.9912 and 0.9923 on the San Diego data set. Therefore, it can be concluded that the random selection operation actually has a little impact on the detection performance of the proposed method.

### V. CONCLUSION

In this article, we have proposed a novel iForest-based anomaly detection algorithm as well as its kernel version for HSIs. In the proposed IFD method, a global iForest is first performed on original HSI to obtain an initial anomaly scores map. Then, the initial score map is refined with a local iForest to obtain the final detection map. The advantage of the proposed framework is that it can make full use of the global and local information in a HSI. Additionally, to better separate the anomaly and background, the original hyperspectral data is projected into a kernel feature space by using kernel method. Experiments on four real hyperspectral data sets show that the KIFD can outperform the IFD and other state-of-the-art methods in terms of both objective and subjective evaluations. The iForest method is capable of detecting anomalies with high efficiency. In the future, how to implement the proposed

method to detect the oil spill by optical remote sensing will be the focus.

## REFERENCES

- [1] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.
- [2] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, Jul. 2014.
- [3] K. Kotwal and S. Chaudhuri, "Visualization of hyperspectral images using bilateral filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2308–2316, May 2010.
- [4] X. Kang, P. Duan, S. Li, and J. A. Benediktsson, "Decolorization-based hyperspectral image visualization," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4346–4360, Aug. 2018.
- [5] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [6] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.
- [7] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.
- [8] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.
- [9] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.
- [10] Y. Wang *et al.*, "A posteriori hyperspectral anomaly detection for unlabeled classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3091–3106, Jun. 2018.
- [11] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [12] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.
- [13] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "A locally adaptive background density estimator: An evolution for RX-based anomaly detectors," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 323–327, Jan. 2014.
- [14] A. P. Schaum, "Hyperspectral anomaly detection beyond RX," *Proc. SPIE*, vol. 6565, May 2007, Art. no. 656502.
- [15] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [16] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005.
- [17] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.
- [18] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [19] S. Li, K. Zhang, Q. Hao, P. Duan, and X. Kang, "Hyperspectral anomaly detection with multiscale attribute and edge-preserving filters," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1605–1609, Oct. 2018.
- [20] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.
- [21] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.
- [22] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [23] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [24] R. Zhao, B. Du, and L. Zhang, "Hyperspectral anomaly detection via a sparsity score estimation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3208–3222, Jun. 2017.
- [25] Y. Qu *et al.*, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.
- [26] F. Li, X. Zhang, L. Zhang, D. Jiang, and Y. Zhang, "Exploiting structured sparsity for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4050–4064, Jul. 2018.
- [27] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
- [28] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and Fisher discriminative convolutional neural networks for object detection," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 265–278, Jan. 2019.
- [29] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 597–601, May 2017.
- [30] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discovery from Data*, vol. 6, no. 1, pp. 1–39, Mar. 2012.
- [31] E. M. Knorr and R. T. Ng, "Algorithms for mining distance-based outliers in large datasets," in *Proc. VLDB*, 1998, pp. 392–403.
- [32] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 93–104.
- [33] C. Chen *et al.*, "iBOAT: Isolation-based online anomalous trajectory detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 806–818, Jun. 2013.
- [34] Z.-G. Ding, D.-J. Du, and M.-R. Fei, "An isolation principle based distributed anomaly detection method in wireless sensor networks," *Int. J. Automat. Comput.*, vol. 12, no. 4, pp. 402–412, Aug. 2015.
- [35] D. Zhang, N. Li, Z.-H. Zhou, C. Chen, L. Sun, and S. Li, "iBAT: Detecting anomalous taxi trajectories from GPS traces," in *Proc. 13th Int. Conf. Ubiquitous Comput.*, 2011, pp. 99–108.
- [36] B. Schölkopf *et al.*, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2758–2765, Nov. 1997.
- [37] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K. R. Müller, "Fisher discriminant analysis with kernels," in *Proc. IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process.*, Madison, WI, USA, Aug. 1999, pp. 41–48.
- [38] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
- [39] P. Duan, X. Kang, S. Li, and P. Ghamisi, "Noise-robust hyperspectral image classification via multi-scale total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1948–1962, Jun. 2019.
- [40] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.
- [41] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, nos. 285–296, pp. 23–27, 1975.
- [42] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [43] D. Manolakis and G. S. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.



**Shutao Li** (M'07–SM'15–F'19) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University, where he is currently a Full Professor. In 2001, he was a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong. From 2002 to 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, London, U.K. In 2005, he joined the Department of Computer Science, The Hong Kong University of Science and Technology, as a Visiting Professor. He has authored or coauthored more than 200 refereed papers. His research interests include image processing, pattern recognition, and artificial intelligence.

Dr. Li is a member of the Editorial Boards of *Information Fusion* and *Sensing and Imaging*. He was a recipient of two Second-Grade State Scientific and Technological Progress Awards of China in 2004 and 2006. He is currently an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.



**Kunzhong Zhang** received the B.Sc. degree from Guangxi University, Nanning, China, in 2017. He is currently pursuing the M.S. degree in electrical engineering with Hunan University, Changsha, China.

His research interests include hyperspectral image anomaly detection.



**Puhong Duan** (S'17) received the B.Sc. degree from Suzhou University, Suzhou, China, in 2014, and the M.S. degree from the Hefei University of Technology, Hefei, China, in 2017. He is currently pursuing the Ph.D. degree with the Laboratory of Vision and Image Processing, Hunan University, Changsha, China.

His research interests include image fusion and hyperspectral image visualization and classification.



**Xudong Kang** (S'13–M'15–SM'17) received the B.Sc. degree from Northeastern University, Shenyang, China, in 2007, and the Ph.D. degree from Hunan University, Changsha, China, in 2015.

In 2015, he joined the College of Electrical Engineering, Hunan University. His research interests include hyperspectral feature extraction, image classification, image fusion, and anomaly detection.

Dr. Kang received the Second Prize in the Student Paper Competition in the IEEE Geoscience and Remote Sensing Society (IGARSS) 2014. In IGARSS 2017, he was selected as the Best Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2016. He serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.