SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

September 2022, Vol. 65 192103:1–192103:13
https://doi.org/10.1007/s11432-021-3310-x

Ensemble of half-space trees for hyperspectral anomaly detection

Ju HUANG[1,2] & Xuelong LI[3,4,1*]

[1]*Shaanxi Key Laboratory of Ocean Optics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an* 710119*, China;*
[2]*University of Chinese Academy of Sciences, Beijing* 100049*, China;*
[3]*School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an* 710072*, China;*
[4]*Key Laboratory of Intelligent Interaction and Applications (Northwestern Polytechnical University), Ministry of Industry and Information Technology, Xi'an* 710072*, China*

Received 23 March 2021/Revised 2 June 2021/Accepted 30 June 2021/Published online 29 August 2022

Abstract Most methods for hyperspectral anomaly detection (HAD) construct profiles of background pixels and identify pixels unconformable to the profiles as anomalies. Recently, isolation forest-based algorithms were introduced into HAD, which identifies anomalies from the background without background modeling. The path length is used as a metric to estimate the anomaly degree of a pixel, but it is not flexible and straightforward. This paper introduces the half-space tree (HS-tree) method from the theory of mass estimation into HAD and proposes a metric involving mass information and tree depth to measure the anomaly degree for each pixel. More specifically, the proposed HS-tree-based detection method consists of three main steps. First, the key spectral-spatial features are extracted using the principal component analysis and the extended morphological attribute profile methods. Then, the ensemble of HS-trees are trained using different randomly selected subsamples from the feature map. Finally, each instance in the feature map traverses through each HS-tree and the anomaly scores are computed as the final detection map. Compared with conventional methods, the experimental results on four real hyperspectral datasets demonstrate the competitiveness of our method in terms of accuracy and efficiency.

Keywords hyperspectral image, anomaly detection, extended morphological attribute profile, mass estimation, half-space tree

Citation Huang J, Li X L. Ensemble of half-space trees for hyperspectral anomaly detection. Sci China Inf Sci, 2022, 65(9): 192103, https://doi.org/10.1007/s11432-021-3310-x

1 Introduction

Hyperspectral images (HSIs) captured by imaging spectrometer contain hundreds or even thousands of narrow spectral bands ranging from the visible to the infrared spectrum [1, 2]. Due to their extremely high spectral resolution, HSIs provide approximately continuous spectral curves. HSIs viewed as a three-dimensional data cube contain both spatial and spectral information. Hence, HSIs can provide discriminative information for identifying physicochemical characteristics of different land covers, leading to several applications such as unmixing [3], classification [4], anomaly detection [5]. Among them, hyperspectral anomaly detection (HAD) is widespread for several applications in military and other fields such as military reconnaissance [6, 7], geological survey [8], and agriculture evaluation [9]. The HAD is considered as a special type of target detection that detects rare and unique objects with different spectral features from those of their surroundings without any prior spectral information on the target and the background.

Several HAD methods have been developed in the past three decades. Most existing HAD approaches can be roughly divided into three categories: statistical model-based, representation-based, and deep

* Corresponding author (email: li@nwpu.edu.cn)

learning-based. The statistical model-based was proposed first and is widely applicable. A typical example of the statistical model-based approach is the Reed-Xiaoli (RX) algorithm [10], which assumes that the background distribution conforms to a multivariate Gaussian distribution and anomaly pixels are recognized by computing the Mahalanobis distance between the test pixels and statistical characteristics of the background. However, the multivariate Gaussian distribution model is insufficient for real HSIs, and the background modeling process may be contaminated by anomaly pixels. Hence, some improved versions of the RX detector have been proposed from different perspectives. A common example is the background purification strategy [11–13], which simplifies background distribution to meet the assumption. Taitano et al. [11] proposed a locally adaptable iterative RX (LAIRX) method that purifies the background using RX detector iteratively until the detection results become unchanged. The kernel-based technique is another important improvement approach [14–16]. Kwon et al. [14] introduced the kernel-based technique and proposed the kernel-RX detector, which enlarges the difference between anomalies and the background by mapping the input into a high-dimensional feature space. To reduce the computational load of KRX, the cluster KRX [16] was developed. It classifies background pixels into clusters and obtains the detection index using the singular value decomposition. Since the HSIs contain different substances, researchers have developed the Gaussian mixture model-based HAD methods. The representative work is the cluster-based anomaly detector [17], which classifies the HSI into some classes and then identifies anomaly objects in each class using the RX detector.

Another category is the representation-based approach. Here, background pixels can be expressed approximately as a set of typical spectra or bases, whereas anomalies cannot. The background joint sparse representation [18] method assumes that background pixels are expressed by a few atoms in a dictionary. For the collaborative representation based detection method [19], it assumes that background pixels are approximately expressible by their spatial neighbors. In [20–22], anomalies and the background are modeled by sparse and low-rank representations, respectively. By considering the mixed characteristic of HSI, the spectral unmixing technique obtains the abundance map instead of original pixels for anomaly detection [23, 24]. Besides, the tensor decomposition-based [25] methods have been recently researched.

Moreover, deep learning-based HAD has drawn increasing attention. In [26], the autoencoder is adopted to learn feature representation, and the reconstruction errors between the input and the recovered images are used to identify anomalies. Moreover, the deep belief network [27] and transferred learning [28] methods have been incorporated into HAD.

These three major categories of methods focus on background modeling, which builds profiles of background pixels, and identifies pixels not satisfying the profiles as anomalies. These methods are optimized to model the background accurately, but not optimized to detect anomalies. Hence, the detection probability can be low in some scenes. Recently, the isolation forest (iForest)-based technique [29] that directly differentiates anomalies without background modeling was introduced into HAD. The main idea is that anomalies are more susceptible to isolation than background pixels in a tree structure. Therefore, they possess shorter average path lengths [29] over a collection of isolation trees. In 2019, Li et al. [30] first applied iForest for HAD and further proposed the kernel iForest-based detector [31] in 2020. In [32], the extracted multiple features instead of original pixels were used to construct isolation trees. In [33], an isolation-based discriminative forest model is proposed. The metric is the path length for all iForest-based detectors. However, it is a global metric, which is unable to rank local anomalies. Moreover, path length has many definitions, depending on the construct used to define the length [34].

How can we define an appropriate and straightforward measure to rank anomaly? We provide the answer in this paper by introducing a metric called the anomaly score, which involves the mass information [34] and the depth of tree node. Furthermore, the half-space tree (HS-tree), a mass estimation method, is adopted to compute the anomaly score. Its key advantages include: (1) it ranks in a data cloud directly with simple counting and ensemble approach, leading to low computation load; (2) the metric derived from HS-tree is straightforward and feasible, which can be used to rank global and local anomalies.

More specifically, a novel HS-tree-based hyperspectral anomaly detector (HSTD) is proposed in this paper. It constructs an ensemble of HS-trees and computes the anomaly score with those trees. The proposed detection method comprises three main steps. First, since HSIs contain a mass of noisy and irrelevant bands, we reduce redundancies and extract key spectral and spatial features of the HSIs. The principal component analysis (PCA) and the extended morphological attribute profile (EMAP) methods are successively applied to extract key spectral-spatial features. Second, the obtained feature map is used as the input to train an ensemble of HS trees. Finally, each instance in the feature map goes through

each HS-tree from the root node to a leaf node, and the anomaly scores of each instance are computed to form the final detection map. Major contributions of our paper are summarized as follows.

(1) The HSI contains noisy and irrelevant bands and HS-tree performance is sensitive to these issues. Thus, the key spectral-spatial features of the HSI are extracted using PCA and EMAP methods. The feature map, not the original HSI, is used as the training set to build the HS-tree.

(2) So far, this is the first time that the HS-tree method is employed for hyperspectral remote sensing applications. To evaluate the anomaly degree of a pixel, a metric, called anomaly score, is introduced in our paper. Furthermore, to take advantage of local information, an improved metric, called the relative anomaly score, is brought forward, which measures local anomaly degree by considering the local distribution of data. The pixels with smaller scores are anomalies.

(3) Compared with other methods, our proposed method is concise, highly efficient, and parameter insensitive. We verified these on real datasets and obtained competitive anomaly detection results.

The rest of this paper is organized as follows: mass estimation, HS-tree, and the metric used to estimate the degree of the anomaly are briefly introduced in Section 2. The proposed method is presented in Section 3. The experiments and analysis performed on four real datasets are presented in Section 4. Finally, conclusion is given in Section 5.

2 Mass estimation and half-space trees

The number of instances in a region is defined as the mass. Two regions have the same mass if their numbers of instances are equal, regardless of the properties of regions. The mass for an instance $\boldsymbol{x}$ is estimated as the average of masses of all regions covering $\boldsymbol{x}$. The masses for all instances form the mass distribution of a dataset, which offers a ranking from central to fringe instances. The instances with a smaller mass are at the fringe and are anomalies. To build plenty of regions and estimate the mass for each instance, we employ HS-tree, a practical mass estimation method. The HS-tree, as shown in Figure 1, is an iterative half-space division operation, yielding several regions covering the instances in the dataset. The mass of each region is obtained when the splitting operation is completed. Multiple HS-trees (i.e., the ensemble of HS-trees) generate different regions covering the instances. By definition, the mass for $\boldsymbol{x}$ is the average of masses on all regions generated by the ensemble of HS-trees. The introductory insight to this methodology is plain and intuitive. For theoretical deduction and proofs, refer to [34]. Here, we employ HS-tree in HAD and review its implementation.

The HS-tree is a special binary tree, which randomly selects a feature every time and recursively divides the dataset into two equal-sized nodes at the mid-value of the selected feature. This half-space split process continues until the predefined size limit of the leaf node s_l or the depth limit of the tree h_l is reached. When all the training instances go through the tree structure, the size of the leaf node, referred to as mass, and the depth of the current leaf node h will be unchanged. These two results, as the output of the tree model, are used to define the anomaly score that quantitatively estimates the degree of the anomaly of an instance.

A simple example is shown in Figure 1, where the HS-tree and anomaly score can be observed. A data space with many 2-feature data is divided into two equal-sized subspaces recursively until the abovementioned predefined conditions are met. Here, the size limit of the leaf node s_l is 3 and the depth limit of the tree h_l is 4. This split process produces 10 regions, marked with different colors. It can be observed that the instance $\boldsymbol{x}_a$ is isolated through three splits, and the size of the region covering $\boldsymbol{x}_a$ is 1. This implies that the mass of the region is 1 for the HS-tree, and we mark it as mass($\boldsymbol{x}_a$) for the sake of description. Besides, the number of splits is the key information, which influences the mass($\boldsymbol{x}_a$) value. Hence, the anomaly degree of $\boldsymbol{x}_a$ can be simultaneously associated with the number of splits and mass($\boldsymbol{x}_a$), as shown in Figure 1(a). With respect to the tree structure shown in Figure 1(b), the anomaly degree of $\boldsymbol{x}_a$ is related to the mass and the depth of the current leaf node h (for this case, the values are 1 and 3, respectively). The anomaly degree of $\boldsymbol{x}$ for one HS-tree can be defined as

$$s(\boldsymbol{x}) = \text{mass}(\boldsymbol{x}) \times 2^h, \tag{1}$$

where $s(\boldsymbol{x})$ denotes the anomaly score of $\boldsymbol{x}$ for a particular tree, mass($\boldsymbol{x}$) is the mass of a leaf node (or region) containing $\boldsymbol{x}$, and h is the current depth of the leaf node or the number of splits. From this definition, $s(\boldsymbol{x}_a) = 1 \times 2^3 = 8$ and $s(\boldsymbol{x}_b) = 4 \times 2^4 = 64$. Evidently, $\boldsymbol{x}_a$ is likely an anomaly and $\boldsymbol{x}_b$ is likely a normal instance. The instance with a smaller anomaly score, which is the average value computed

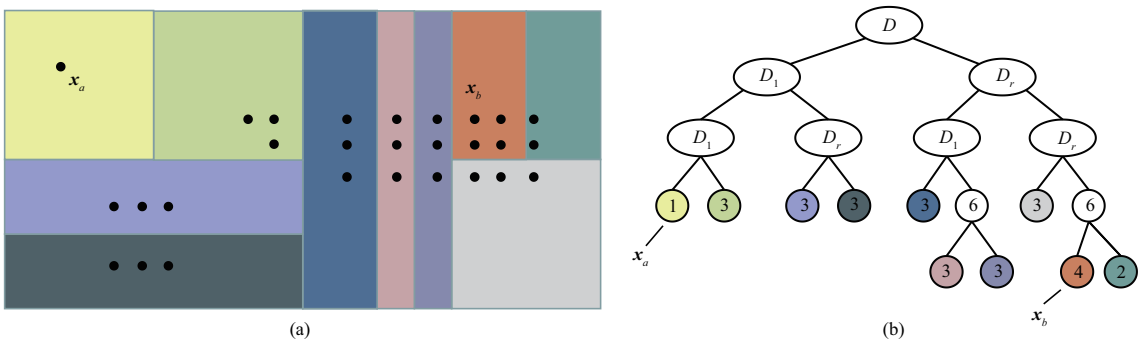

Figure 1 (Color online) Example of dataset partitioned by a simple HS-tree. (a) A space division process, involving a data space being recursively split into two equal-sized subspaces. This process produced 10 regions. The instance $\boldsymbol{x}_a$ is isolated through three splits, the region size where the $\boldsymbol{x}_a$ lies in is 1, and that for where the $\boldsymbol{x}_b$ lies in is 4. (b) The HS-tree construction process corresponding to (a). For this example, mass($\boldsymbol{x}_a$) = 1 and mass($\boldsymbol{x}_b$) = 4.

using many randomly constructed HS-trees, is more likely an anomaly. The conclusion conforms to the definition of mass for one instance. Hence, many HS-trees (i.e., the ensemble of HS-trees) will be built from the above description. The final anomaly score $S(\boldsymbol{x})$ is defined as

$$S(\boldsymbol{x}) = \frac{1}{c}\sum_{i=1}^{c} s_i(\boldsymbol{x}), \tag{2}$$

where c is the number of HS-trees, $s_i(\boldsymbol{x})$ is the anomaly degree derived from an HS-tree. We note that the HS-tree is built from a set of training instances, randomly selected from the dataset.

In a sense, the implementation of the HS-tree is similar to that of iForest. However, the difference is that the HS-tree defines a unique form of tree growth and explores the information on the tree nodes. The HS-tree can be considered as an extension of iForest, and the path length can be considered as a proxy to the anomaly score [34]. When the mass of each node is fixed at 1, the anomaly score degenerates to the path length. So far, mass estimation and HS-tree have not been introduced in any remote sensing application. Our paper is the first attempt to apply this method to HAD.

3 Proposed method

The framework of the proposed algorithm is displayed in Figure 2. There are three main steps: (1) extracting a spectral-spatial feature map of HSI; (2) using the feature map to train an ensemble of HS-trees; (3) Detecting anomalies with the HS-trees.

3.1 Extracting spectral-spatial features of HSI

The HSIs are high dimensional image and contain noisy and irrelevant bands, and the HS-tree method is inapplicable for extremely high dimensional data [35]. Thus, we reduce spectral dimension and extract discriminant features of the HSI [36, 37]. The PCA [38] is utilized to reduce the dimension of the HSI, which constructs a linear transformation to preserve principal spectral features. The first ς principal components (PCs) denoted as $\{\mathrm{PC}_1, \mathrm{PC}_2, \ldots, \mathrm{PC}_\varsigma\}$ are the extracted spectral features. For a comprehensive description of the HSI, the spatial structural information is to be extracted. The EMAP can model the spatial structural characteristics of the HSI according to different attributes setting [39]. This methodology has been applied for spatial feature extraction in many hyperspectral remote sensing tasks [39, 40]. Based on the obtained PCs, the EMAP are computed as follows:

$$\mathrm{EMAP} = \{\mathrm{AP}(\mathrm{PC}_1), \mathrm{AP}(\mathrm{PC}_2), \ldots, \mathrm{AP}(\mathrm{PC}_\varsigma)\}, \tag{3}$$

where AP is a multi-level decomposition of an input image based on the attribute filter [41]. The attribute filter performs a morphological transform on the connected components (i.e., regions of connected pixels which are equal intensity spatially) according to a given criterion (e.g., area, volume). Specifically, the attribute filter gets the attribute value on a connected component according to the criterion T, and compares the value against a threshold λ to complete the image processing. If the value of a connected

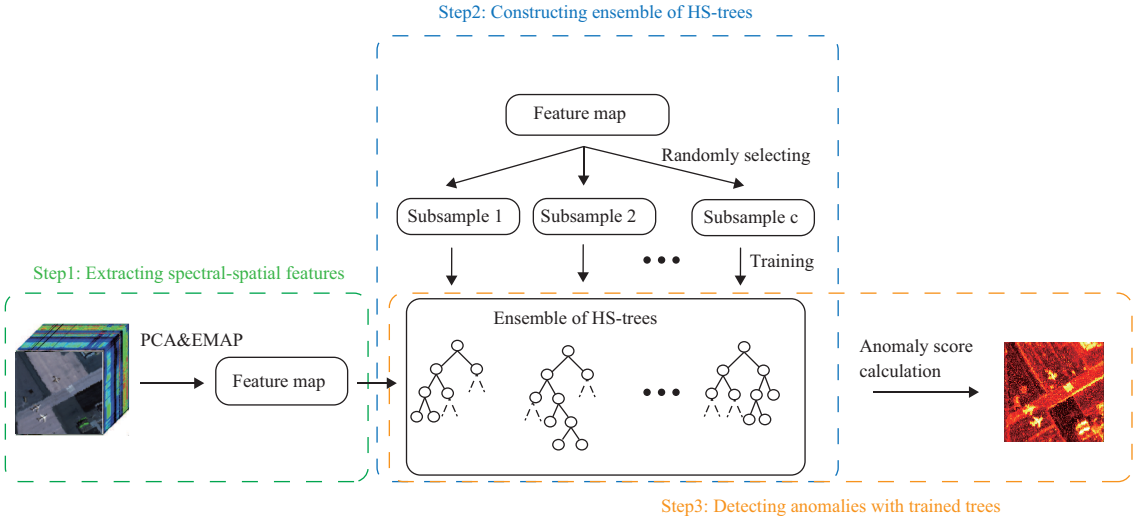

Figure 2 (Color online) Framework of the proposed algorithm.

component C_i meets $T(C_i) > \lambda$, C_i becomes unchanged, otherwise it will be incorporated into an adjacent region. If the connected component is incorporated into the region with lower intensity, the process is called attribute thinning operation, otherwise it is called attribute thickening operation. In this paper, the area attribute a is adopted as the criterion and the corresponding threshold sequence is set to $\{\lambda_1, \lambda_2, \lambda_3\}$. Let ϕ_i and γ_i represent the attribute thickening and thinning operations [42], respectively. The AP for each PC is a combination of a series of attribute thickening and thinning operations:

$$\mathrm{AP}(\mathrm{PC}_i) = \left\{\phi^a_{\lambda_3}(\mathrm{PC}_i), \phi^a_{\lambda_2}(\mathrm{PC}_i), \phi^a_{\lambda_1}(\mathrm{PC}_i), \mathrm{PC}_i, \gamma^a_{\lambda_3}(\mathrm{PC}_i), \gamma^a_{\lambda_2}(\mathrm{PC}_i), \gamma^a_{\lambda_1}(\mathrm{PC}_i)\right\}. \tag{4}$$

Since Eq. (4) contains both the spectral information and the spatial information, the EMAP composed of APs is used as the low-dimension feature map $D \in \mathbb{R}^{N\times 7\varsigma}$. Additionally, each row vector $\boldsymbol{x}$ in D denotes an instance in this paper.

3.2 Constructing ensemble of HS-trees

Based on the feature map D, the HS-tree is trained. We randomly select M instances from D to obtain a subsample $\mathcal{D}$, which is used as a training set. The $\mathcal{D}$, as the root node, will be divided into two equal-sized nodes at a randomly selected feature q, and the split point mid_q is the mid-value of the feature q. These two nodes are called the left node $\mathcal{D}_l$ and the right node $\mathcal{D}_r$, respectively. The training instances in $\mathcal{D}$ are filtered through one of the two nodes depending on which side of the split point the training instances lie in. This node building process is repeated until the size limit s_l or the depth limit h_l is reached. The leaf nodes, resembling the data buckets, keep countering and increasing as the training instances fall into them. When all the instances in $\mathcal{D}$ go through the tree structure, the size and depth of all the leaf nodes will be recorded. These two results are used to calculate anomaly degree in detection phase. Figure 1(b) is an example of the construction process of the HS-tree. The HS-tree construction process is repeated c times using different subsamples to obtain an ensemble of HS-trees. In our paper, $h_l = \log_2 M$ is set by default, while other parameters are set according to the parameter analysis in Subsection 4.4.

3.3 Detecting anomalies with trained trees

In this step, each instance in D goes through each HS-tree from the root node to a leaf node. The anomaly score of each instance is computed as (1). Let $\{T_1, T_2, \ldots, T_c\}$ be the HS-trees using different subsamples, and $s_i(\boldsymbol{x}_t)$ denote the anomaly score of a test instance $\boldsymbol{x}_t \in D$ which goes through the tree T_i. It is worth noting that an instance traversing through different HS-trees may have different anomaly scores. Hence, the anomaly score of each instance is an average of anomaly scores deriving from different HS-trees. The anomaly score $S(\boldsymbol{x}_t)$ is computed according to (2):

$$S(\boldsymbol{x}_t) = \frac{1}{c}\sum_{i=1}^{c} s_i(\boldsymbol{x}_t). \tag{5}$$

Although the anomaly score $S(\boldsymbol{x}_t)$ may already be used to identify anomalies, it is inappropriate for practical application. The value range of $S(\boldsymbol{x}_t)$ is very large and is not normalized to $(0, 1]$. Furthermore, the anomaly score $S(\boldsymbol{x}_t)$ dose not consider how isolated an instance is from its local neighborhood, and local anomalies may be masked by the background clusters of low density. Thus, we borrow the concept of relative density and further bring forward an improved metric called the relative anomaly score $r_i(\boldsymbol{x}_t)$ to measure the anomaly degree of an instance. It is a proportion of anomaly scores as follows:

$$r_i(\boldsymbol{x}_t) = \frac{\widehat{s}_i(\boldsymbol{x}_t)}{s_i(\boldsymbol{x}_t) \times M}, \tag{6}$$

where $\widehat{s}_i(\boldsymbol{x}_t)$ is the anomaly score of the immediate father node of the leaf node where the $\boldsymbol{x}_t$ falls into. M is a normalization term which is the size of the subsample $|\mathcal{D}|$. The $r_i(\cdot)$ is normalized to $(0, 1]$. Similarly, the final anomaly score is an average of relative anomaly scores over c HS-trees as follows:

$$R(\boldsymbol{x}_t) = \frac{1}{c}\sum_{i=1}^{c} r_i(\boldsymbol{x}_t). \tag{7}$$

When each instance in D traverses through all trees, the relative anomaly score $R(\boldsymbol{x}_t)$ is calculated to obtain the final detection map.

4 Experiments and discussion

4.1 Datasets

Four HSI datasets collected from different scenes are utilized to assess our method performance. These datasets have different characteristics, including the data acquisition sensors, the materials in the background, and the shape and type of anomaly objects. The pseudo-color images and the ground truth maps which mark the locations of anomaly pixels are shown and processed in Subsection 4.3.

The first dataset is the San Diego dataset, which was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) from the San Diego airport area, CA, USA. The original size is 400×400, and a part with the size of 100×100 in the top-left corner is chosen as the test image. The spectral resolution is 10 nm and the spatial resolution is 3.5 m/pixel. It has 189 spectral bands ranging from 370 to 2510 nm, after removing water absorption, low SNR, and poor-quality bands. For this dataset, three airplanes are taken as the anomaly objects, and the background mainly contains hangars, airstrips, and meadows.

The second dataset is the HYDICE dataset, which was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) over an urban area, CA, USA. The original size is 307×307, and a part with the size of 80×100 is selected as the test image. It has 175 spectral bands ranging from 400 to 2500 nm, after removing water absorption, low SNR, and poor-quality bands. For this dataset, the vehicles and buildings are taken as the anomaly objects, and the background mainly includes parking lots, pools, roads and soils.

The third dataset is the Texas Coast dataset, which was collected by the AVIRIS sensor over an urban area in the Texas Coast, TX, USA. The size of selected test image is 100×100. It has 207 spectral bands ranging from 400 to 1350 nm, and the spatial resolution is 17.2 m/pixel. For this dataset, the houses are taken as the anomaly objects and the background comprises highways, meadows, and vegetation. This image is contaminated by the strip noises, therefore it is a challenge to recognize all anomalies accurately.

The fourth dataset is the Pavia dataset, which was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) over a city center in Pavia, Italy. The size of selected part is 108×120. It has 102 spectral bands ranging from 430 to 860 nm, and its spatial resolution is 1.3 m/pixel. The background mainly includes a bridge and water, and the vehicles are generally taken as the anomaly objects.

4.2 Experimental setup

To assess the performance, the proposed HSTD method is in comparison with eight popular algorithms. They are the RX detector [10], local RX (LRX) detector [10], cluster kernel RX detector (CKRX) [16], collaborative representation-based detector (CRD) [19], background joint sparse representation (BJSR) [18],

low-rank and sparse matrix decomposition (LRaSMD) [20], and kernel isolation forest-based detector (KIFD) [31]. Among them, the RX, LRX and CKRX are statistical modeling-based algorithms. The CRD, BJSR and LRaSMD are representation-based methods. They employ collaborative representation, sparse representation, and low-rank decomposition techniques, respectively. The KIFD is an up-to-date detection algorithm based on iForest.

Two popular criteria are utilized to evaluate the detection performance quantitatively. One is the receiver operating characteristic (ROC) curve which describes the relationship between the detection probability (DP) and the false alarm rate (FAR). The DP and FAR are defined as

$$\mathrm{DP} = \frac{P_{\mathrm{D}}}{P_{\mathrm{T}}}, \quad \mathrm{FAR} = \frac{P_{\mathrm{F}}}{P}, \tag{8}$$

where P_{D} denotes the number of pixels which are identified correctly, P_{T} denotes the total number of anomaly pixels which are given in the ground truth, P_{F} denotes the number of pixels which are misidentified as anomalies, and P denotes the total number of pixels in the image. Generally, the x-coordinate is the FAR, and the y-coordinate is the DP. If the curve of a method is closer to the top-left corner, it indicates the method has a better detection performance. The another is the area under the ROC curve (AUC). A better detector has a larger AUC value.

Additionally, some critical parameters involved in the comparison experiments will be set. For the HSTD method, there are four main parameters: the number of principal components ς, the size of subsample M, the size limit of leaf node s_l, and the number of HS-trees c. According to the parameter analysis in Subsection 4.4, the ς is set to 3 for the San Diego and Pavia datasets and is set to 6 for the HYDICE and Texas Coast datasets. Since our HSTD method is insensitive to the changes of parameter M, s_l, and c, we set them to 5% of the total pixels, 2, 30, respectively. For the competitors, the optimal parameters are set according to the corresponding papers. For the CKRX detector, the width of the Gaussian radial basis function kernel is set to 40, and the number of clusters for four datasets is set to 150, 70, 100 and 70, respectively. As we all know, the detection performance of the LRX, CRD and BJSR methods is closely related with the window size (the inner window W_{in} and the outer window W_{out}). Hence, the W_{in} is changed from 3 to 19 and the W_{out} is changed from 5 to 35, in order to select the optimal window sizes. For the San Diego dataset, the $(W_{\mathrm{in}}, W_{\mathrm{out}})$ is set to $(17, 35)$, $(15, 31)$ and $(15, 31)$ for the LRX, BJSR, and CRD methods, respectively. For the HYDICE dataset, the $(W_{\mathrm{in}}, W_{\mathrm{out}})$ is set to $(3, 5)$ for the LRX, BJSR and CRD methods. For the Texas Coast dataset, the $(W_{\mathrm{in}}, W_{\mathrm{out}})$ is set to $(3, 7)$, $(3, 21)$ and $(3, 21)$ for the LRX, BJSR, CRD methods. For the Pavia dataset, the $(W_{\mathrm{in}}, W_{\mathrm{out}})$ is set to $(3, 7)$, $(3, 21)$ and $(3, 21)$ for the LRX, BJSR, and CRD methods, respectively. For the LRaSMD method, the rank value r is set to 5, and the cardinality k is set to 0.004 for four datasets. For the KIFD method, the parameters ζ, q and M are set to 300, 1000 and $3\% \times N$ (N is number of pixels in the HSI) for four datasets, respectively.

4.3 Detection performance

For the San Diego dataset, the detection maps are shown in Figure 3. For the HSTD and KIFD methods, the outlines and shapes of three airplanes are clear, although some background pixels are misidentified as anomalies. The RX, CKRX and LRaSMD methods can locate the anomalies, but the shape information is not clear. The LRX and CRD methods obtain low detection probability, and a small number of anomalies are identified correctly. The BJSR method cannot suppress the influence of the background effectively, and therefore there are more pixels misidentified than other methods. The ROC curves and the AUC values are displayed in Figure 4. The AUC value of the HSTD method is 0.986, which is the second highest in all competitors.

For the HYDICE dataset, the detection maps are displayed in Figure 5. Our proposed HSTD method can detect all anomaly pixels. The RX, LRX, and CRD methods can restrain the background effectively, but few pixels can be identified. The BJSR and CKRX methods are good at detecting anomalies, but perform poorly in suppressing the background. The ROC curves and the AUC values are shown in Figure 6. The AUC value of the HSTD method is 0.993, which is slightly smaller than the KIFD method.

For the Texas Coast dataset, the detection maps are shown in Figure 7. Obviously, all anomaly targets are detected by the HSTD method. But the effect of background suppression is not satisfactory. Compared with the KIFD method based on iForest, our HSTD method based on HS-tree detects nearly all of the anomaly objects clearly. The ROC curves and the AUC values are shown in Figure 8. When

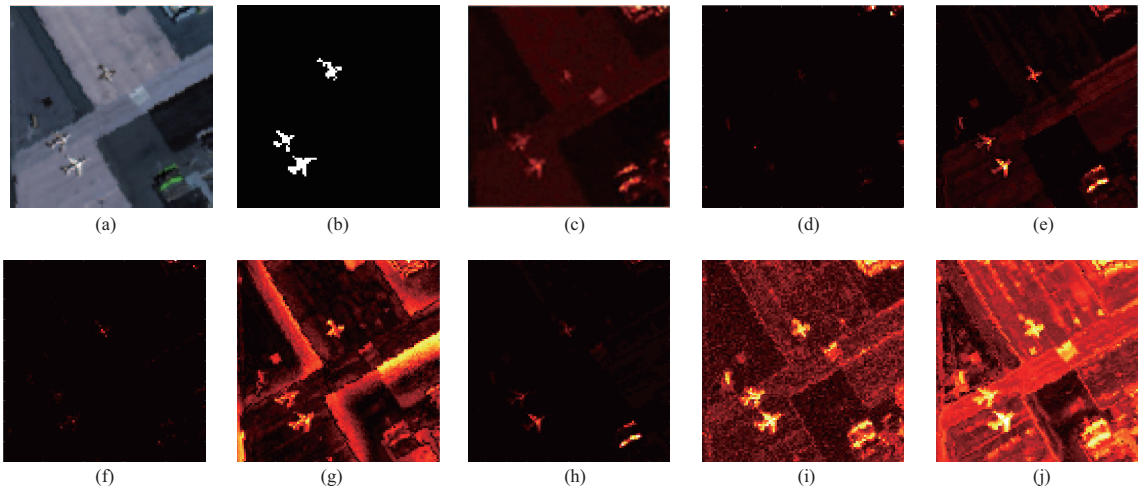

Figure 3 (Color online) Detection maps obtained by different methods for the San Diego dataset. (a) Pseudo-color image; (b) ground truth; (c) RX; (d) LRX; (e) CKRX; (f) CRD; (g) BJSR; (h) LRaSMD; (i) KIFD; (j) HSTD.

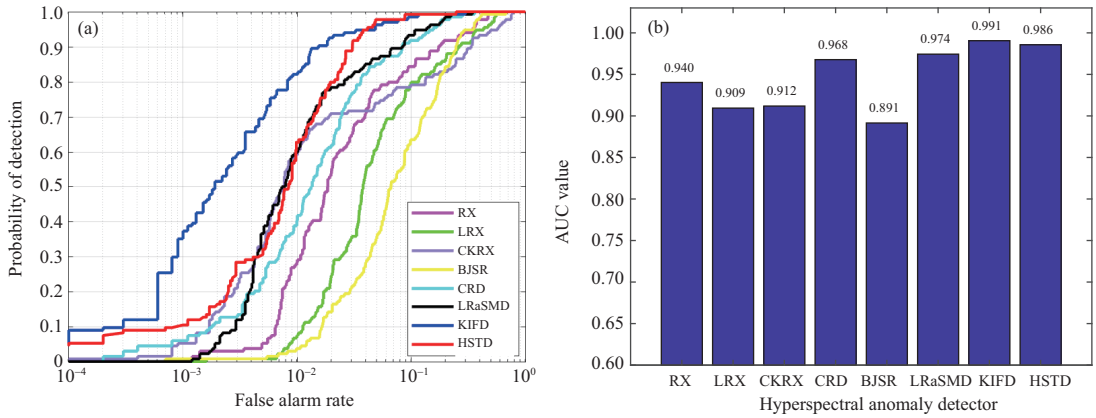

Figure 4 (Color online) Detection accuracy evaluation for the San Diego dataset. (a) ROC curves; (b) AUC values.

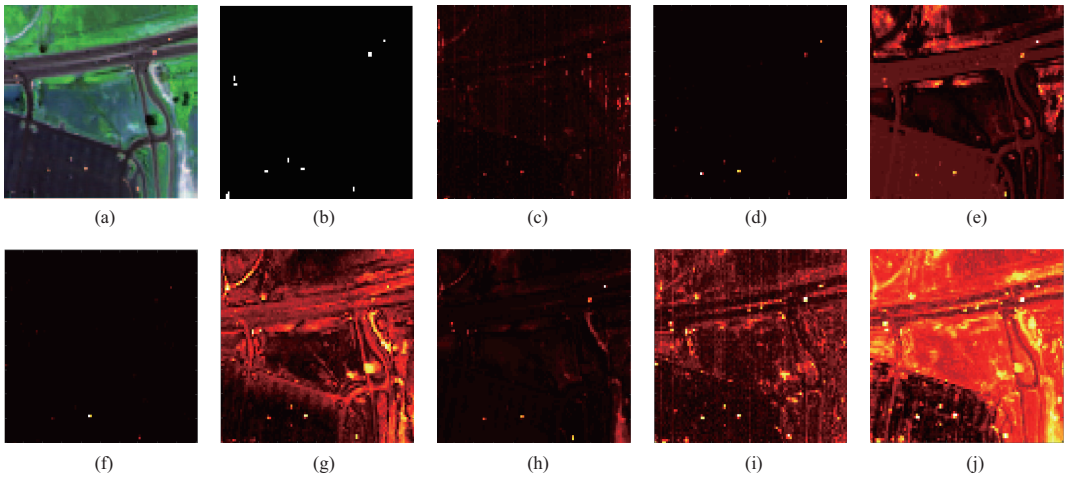

Figure 5 (Color online) Detection maps obtained by different methods for the HYDICE dataset. (a) Pseudo-color image; (b) ground truth; (c) RX; (d) LRX; (e) CKRX; (f) CRD; (g) BJSR; (h) LRaSMD; (i) KIFD; (j) HSTD.

the false alarm rate is set to 10^{-2}, the detection probability is close to 1. The AUC value of the HSTD method is 0.999, which is the highest in all algorithms.

For the Pavia dataset, the detection maps are displayed in Figure 9. we can find that the proposed HSTD method can accurately locate and identify anomaly objects, although some background pixels,

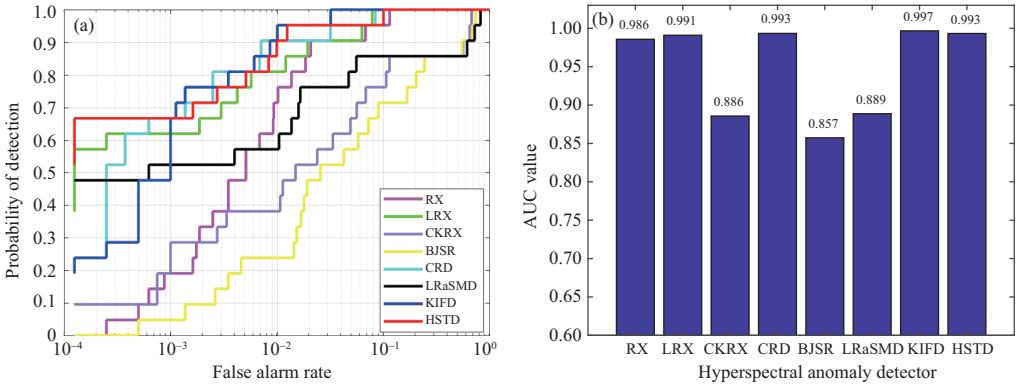

Figure 6 (Color online) Detection accuracy evaluation for the HYDICE dataset. (a) ROC curves; (b) AUC values.

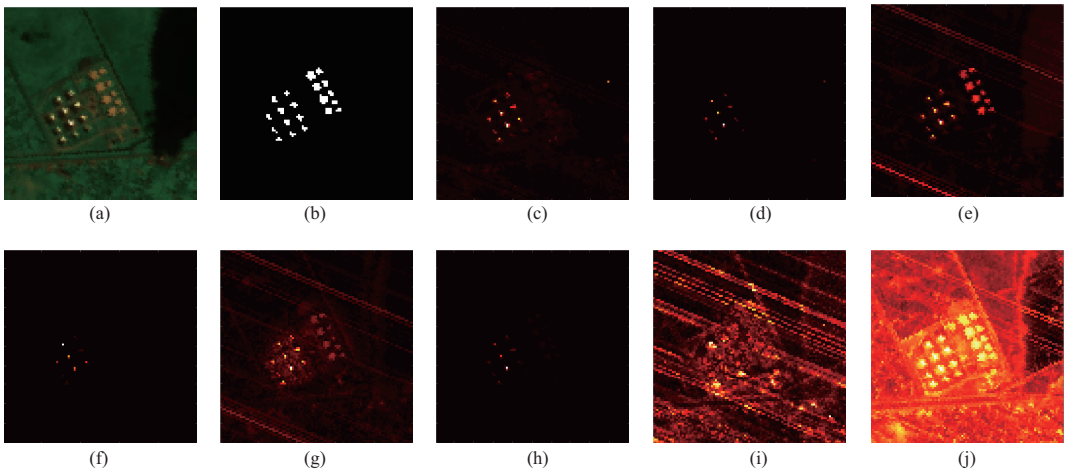

Figure 7 (Color online) Detection maps obtained by different methods for the Texas Coast dataset. (a) Pseudo-color image; (b) ground truth; (c) RX; (d) LRX; (e) CKRX; (f) CRD; (g) BJSR; (h) LRaSMD; (i) KIFD; (j) HSTD.

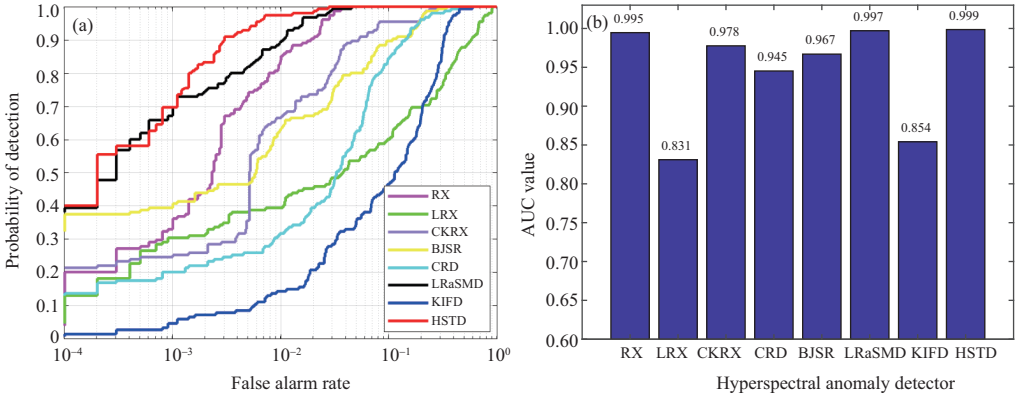

Figure 8 (Color online) Detection accuracy evaluation for the Texas Coast dataset. (a) ROC curves; (b) AUC values.

especially the edge of the bridge, are falsely detected as anomalies. The ROC curves and the AUC values are shown in Figure 10. The AUC value of the HSTD method is 0.993.

After the experiments described above, it can be found that the tree-based methods, such as the KIFD and HSTD, are adept at identifying the anomalies, but perform poorly in suppressing the background. The main reason is as follows: the tree-based methods aims to split anomalies from background directly without background profiling, and thus the background information is not explored adequately, leading to high false alarm rate.

In addition, the time complexities are discussed. The HS-tree involves no distance calculation and

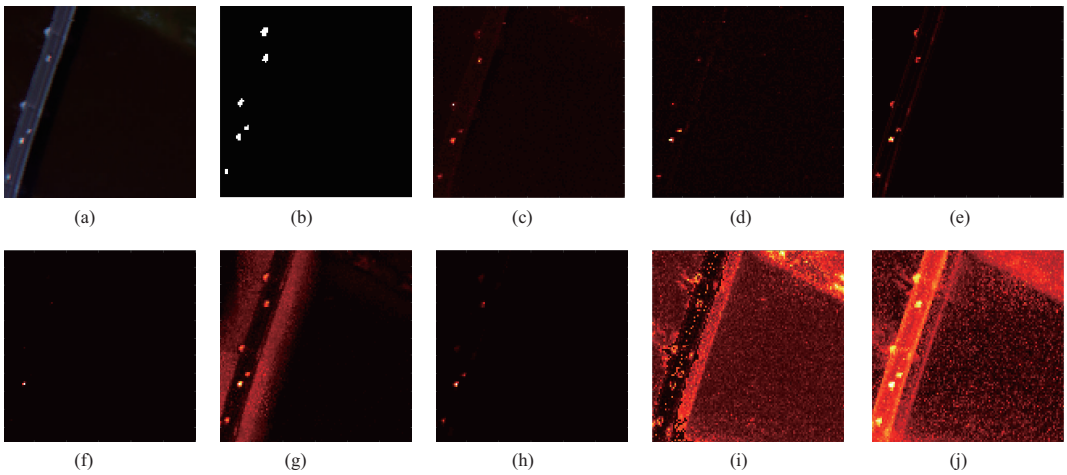

Figure 9 (Color online) Detection maps obtained by different methods for the Pavia dataset. (a) Pseudo-color image; (b) ground truth; (c) RX; (d) LRX; (e) CKRX; (f) CRD; (g) BJSR; (h) LRaSMD; (i) KIFD; (j) HSTD.

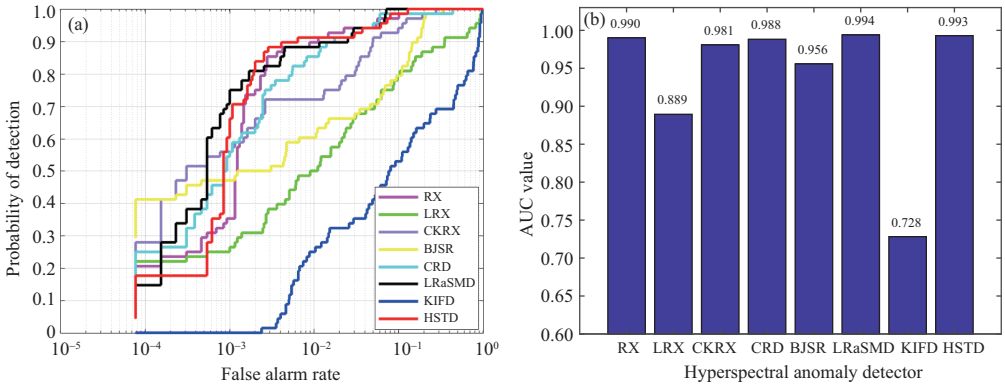

Figure 10 (Color online) Detection accuracy evaluation for the Pavia dataset. (a) ROC curves; (b) AUC values.

Table 1 Computing time (in second) of different methods

Method	San Diego	HYDICE	Texas Coast	Pavia
RX	0.10	0.09	0.11	0.12
LRX	37.25	26.70	42.48	18.66
CKRX	4.32	1.92	0.93	2.31
CRD	3553.78	2194.31	3680.26	2694.65
BJSR	5.33	3.88	5.62	5.94
LRaSMD	31.10	20.20	37.43	13.93
KIFD	85.44	58.90	59.10	60.05
HSTD	1.03	0.86	0.95	1.04

is built using only a small subsample. Hence, the HS-tree is built very quickly. The construction of an ensemble of c HS-trees has time complexity $O\left(h_l M c\right)$. The detection phase has time complexity $O\left(h_l N c\right)$, where N is the number of pixels in the image. Hence, the time complexity of the proposed method is $O\left(c h_l\left(M+N\right)\right)$. The computing time experiments are carried out on a machine with a 2.9-GHz CPU and a 16-G memory. The computing time of different methods is shown in Table 1. It can be seen that our method has less time cost than other methods except the RX detector.

4.4 Parameter analysis and discussion

The parameter analysis of the HSTD is discussed in this subsection. There are four main parameters: the number of PCs ς, the size of subsample M, the size limit of leaf node s_l, and the number of HS-trees c. Here, we use the AUC value to assess the effects of the parameters changes. The effects of the parameter ς on four datasets are shown in Figure 11(a). For the San Diego dataset, the AUC value decreases slightly

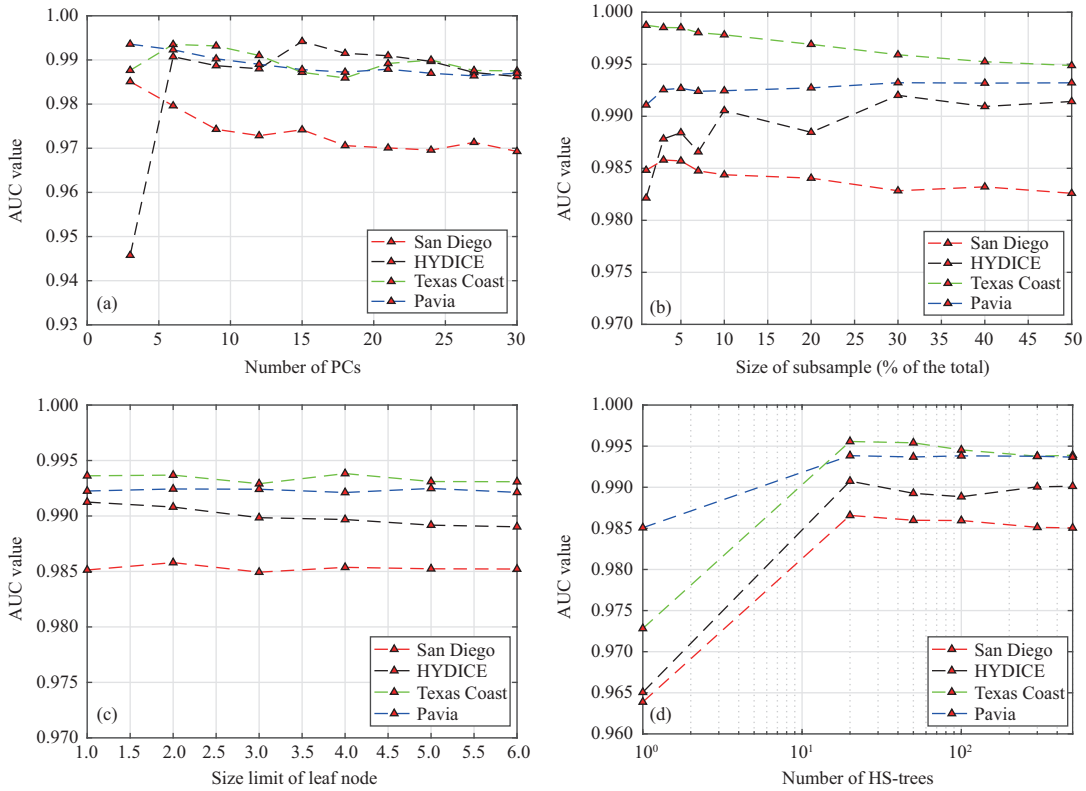

Figure 11 (Color online) Effect of the four parameters on the detection performance. (a) The number of PCs ς; (b) the size of subsample M; (c) the size limit of leaf node s_l; (d) the number of HS-trees c.

as the ς increases. For the Texas Coast and Pavia datasets, the AUC value changes slightly. For the HYDICE dataset, the AUC value increases quickly at first, then reaches the peak at 6 and then decreases slightly. The influences of parameter M on four datasets are shown in Figure 11(b). The parameter M is set to 1%, 3%, 5%, 7%, 10%, 20%, 30%, 50% of the total pixels. It can be observed that the trends of AUC values are irregular. For the HYDICE and Pavia datasets, the AUC values become larger as the M increases. As to the San Diego and Texas Coast datasets, the AUC values become smaller slightly. As shown in Figure 11(c), the AUC value is hardly influenced by the changes of the size limit s_l. The effects of the parameter c are shown in Figure 11(d). The AUC values on four datasets increase at first, then reach the peak at 30 and then keep stable. The $c = 30$ is optimal parameter already, and further increasing the c is unnecessary, which may increase time complexities. To summarize, the proposed method is insensitive to the most parameters, although the ς has a little influence on the performance.

5 Conclusion

In this paper, we introduce the HS-tree method from the theory of mass estimation into HAD and propose a measure involving mass information and tree depth to assess the anomaly degree for each pixel. First, our proposed method extracts the spatial-spectral features from the HSI using PCA and EMAP methods. Then, a collection of HS trees is trained using different subsamples randomly selected from the feature map. Finally, all instances in the feature map go through these trained HS trees and anomaly scores are computed to form the final anomaly detection map. Our method is the first to apply HS-tree to HAD. Furthermore, to take advantage of global and local information, we present a relative anomaly measure, which assesses the anomaly degree locally. The experimental results show the competitiveness of the proposed method in terms of accuracy and efficiency.

Since the HSTD algorithm splits anomalies from the background directly without background profiling, the background information is not explored, leading to poor results of background suppression. Hence, adding background information into our anomaly measure is our future focus.

Acknowledgements This work was supported by the Key Research Program of Frontier Sciences, Chinese Academy of Sciences

(Grant No. QYZDY-SSW-JSC044) and National Natural Science Foundation of China (Grant No. 61871470).

References

1 Gu Y F, Liu T Z, Gao G M, et al. Multimodal hyperspectral remote sensing: an overview and perspective. Sci China Inf Sci, 2021, 64: 121301

2 Chen J, Ling Z C, Qiao L, et al. Mineralogy of Chang'e-4 landing site: preliminary results of visible and near-infrared imaging spectrometer. Sci China Inf Sci, 2020, 63: 140903

3 Pu H Y, Wang B, Zhang L M. Simplex geometry-based abundance estimation algorithm for hyperspectral unmixing (in Chinese). Sci Sin Inform, 2012, 42: 1019–1033

4 Wang Q, He X, Li X. Locality and structure regularized low rank representation for hyperspectral image classification. IEEE Trans Geosci Remote Sens, 2019, 57: 911–923

5 Hou Z F, Wei L, Tao R, et al. Collaborative representation with background purification and saliency weight for hyperspectral anomaly detection. Sci China Inf Sci, 2022, 65: 112305

6 Eismann M T, Stocker A D, Nasrabadi N M. Automated hyperspectral cueing for civilian search and rescue. Proc IEEE, 2009, 97: 1031–1055

7 Li X, Chen M, Nie F, et al. A multiview-based parameter free framework for group detection. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017. 4147–4153

8 Kruse F A, Boardman J W, Huntington J F. Comparison of airborne hyperspectral data and EO-1 hyperion for mineral mapping. IEEE Trans Geosci Remote Sens, 2003, 41: 1388–1400

9 Gao Z, Shao Y, Xuan G, et al. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. Artif Intell Agr, 2020, 4: 31–38

10 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

11 Taitano Y P, Geier B A, Bauer K W. A locally adaptable iterative RX detector. EURASIP J Adv Signal Process, 2010, 2010: 341908

12 Du B, Zhang L. Random-selection-based anomaly detector for hyperspectral imagery. IEEE Trans Geosci Remote Sens, 2011, 49: 1578–1589

13 Gao L, Guo Q, Plaza A, et al. Probabilistic anomaly detector for remotely sensed hyperspectral data. J Appl Remote Sens, 2014, 8: 083538

14 Kwon H, Nasrabadi N M. Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery. IEEE Trans Geosci Remote Sens, 2005, 43: 388–397

15 Zhao R, Du B, Zhang L. A robust nonlinear hyperspectral anomaly detection approach. IEEE J Sel Top Appl Earth Observations Remote Sens, 2014, 7: 1227–1234

16 Zhou J, Kwan C, Ayhan B, et al. A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images. IEEE Trans Geosci Remote Sens, 2016, 54: 6497–6504

17 Carlotto M J. A cluster-based approach for detecting man-made objects and changes in imagery. IEEE Trans Geosci Remote Sens, 2005, 43: 374–387

18 Li J, Zhang H, Zhang L, et al. Hyperspectral anomaly detection by the use of background joint sparse representation. IEEE J Sel Top Appl Earth Observations Remote Sens, 2015, 8: 2523–2533

19 Li W, Du Q. Collaborative representation for hyperspectral anomaly detection. IEEE Trans Geosci Remote Sens, 2015, 53: 1463–1474

20 Sun W, Liu C, Li J, et al. Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery. J Appl Remote Sens, 2014, 8: 083641

21 Xu Y, Wu Z, Li J, et al. Anomaly detection in hyperspectral images based on low-rank and sparse representation. IEEE Trans Geosci Remote Sens, 2016, 54: 1990–2000

22 Wang W, Li S, Qi H, et al. Identify anomaly component by sparsity and low rank. In: Proceedings of the 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2015. 1–4

23 Qu Y, Guo R, Wang W, et al. Anomaly detection in hyperspectral images through spectral unmixing and low rank decomposition. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016. 1855–1858

24 Qu Y, Wang W, Guo R, et al. Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition. IEEE Trans Geosci Remote Sens, 2018, 56: 4391–4405

25 Zhang X, Wen G, Dai W. A tensor decomposition-based anomaly detection algorithm for hyperspectral image. IEEE Trans Geosci Remote Sens, 2016, 54: 5801–5820

26 Bati E, ćalışkan A, Koz A, et al. Hyperspectral anomaly detection method based on auto-encoder. In: Proceedings of SPIE, 2015. 220–226

27 Ma N, Peng Y, Wang S, et al. An unsupervised deep hyperspectral anomaly detector. Sensors, 2018, 18: 693

28 Li W, Wu G, Du Q. Transferred deep learning for anomaly detection in hyperspectral imagery. IEEE Geosci Remote Sens Lett, 2017, 14: 597–601

29 Liu F, Ting K, Zhou Z. Isolation forest. In: Proceedings of the 8th IEEE International Conference on Data Mining, 2008. 413–422

30 Zhang K, Kang X, Li S. Isolation forest for anomaly detection in hyperspectral images. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium, 2019. 437–440

31 Li S, Zhang K, Duan P, et al. Hyperspectral anomaly detection with kernel isolation forest. IEEE Trans Geosci Remote Sens, 2020, 58: 319–329

32 Wang R, Nie F, Wang Z, et al. Multiple features and isolation forest-based fast anomaly detector for hyperspectral imagery. IEEE Trans Geosci Remote Sens, 2020, 58: 6664–6676

33 Chang S, Du B, Zhang L. A subspace selection-based discriminative forest method for hyperspectral anomaly detection. IEEE Trans Geosci Remote Sens, 2020, 58: 4033–4046

34 Ting K M, Zhou G T, Liu F T, et al. Mass estimation. Mach Learn, 2013, 90: 127–160

35 Bandaragoda T R, Ting K M, Albrecht D, et al. Isolation-based anomaly detection using nearest-neighbor ensembles. Comput Intell, 2018, 34: 968–998

36 Li X L, Chen M L, Nie F P, et al. Locality adaptive discriminant analysis. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017. 2201–2207

37 Li X L, Zhao B. Video distillation (in Chinese). Sci Sin Inform, 2021, 51: 695–734

38 Agarwal A, El-Ghazawi T, El-Askary H, et al. Efficient hierarchical-PCA dimension reduction for hyperspectral imagery. In: Proceedings of 2007 IEEE International Symposium on Signal Processing and Information Technology, 2007. 353–356

39 Kang X, Zhang X, Li S, et al. Hyperspectral anomaly detection with attribute and edge-preserving filters. IEEE Trans Geosci Remote Sens, 2017, 55: 5600–5611

40 Mura M D, Villa A, Benediktsson J A, et al. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. IEEE Geosci Remote Sens Lett, 2011, 8: 542–546

41 Mura M D, Benediktsson J A, Waske B, et al. Morphological attribute profiles for the analysis of very high resolution images. IEEE Trans Geosci Remote Sens, 2010, 48: 3747–3762

42 Breen E J, Jones R. Attribute openings, thinnings, and granulometries. Comput Vision Image Underst, 1996, 64: 377–389