Hyperspectral Anomaly Detection via Graphical Connected Point Estimation and Multiple Support Vector Machines

SHANGZHEN SONG, HANLIN QIN, YIXIN YANG, ZHE ZHANG, AND HUIXIN ZHOU, (Member, IEEE)
School of Physics and Optoelectronic Engineering, Xidian University, Xi’an 710071, China
Corresponding authors: Hanlin Qin (hlqin@mail.xidian.edu.cn) and Huixin Zhou (hxzhou@mail.xidian.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61675160, Grant 61705173, and Grant 51801142, in part by the Natural Science Foundation of Shaanxi Province of China under Grant 2018JQ5022, in part by the Fundamental Research Funds for the Central Universities under Grant JB180502, and in part by the 111 Project under Grant B17035.

ABSTRACT Most hyperspectral anomaly detection algorithms are based on various hypothetical models justified by different methods. The closer to the real-world scene distribution a hypothetical model is, the better detection performance usually results, albeit at the expense of increased model complexity. There is also a challenge in most of the detection methods that anomalous components cannot be completely separated from the background due to the difference between hypothetical models and the real-world scene. To address this problem, a novel strategy based on graph theory and the multiple support vector machines technique is proposed in this paper. Firstly, a graphical connected point-based estimation is utilized as a preprocessing procedure to separate potential anomalies from robust backgrounds of the image. Second, without building a hypothetical model, the hyperspectral image is classified into several categories by a clustering method according to the different characteristics of materials. Consequently, for each category, the support vector machine is used to classify anomalous components and background components belonging to specific categories, and the result map of each classification is obtained. Finally, all maps corresponding to various categories are fused by an effective fusion strategy, and the detection result is generated by our method. In the subsequent experiments, we use several state-of-art anomaly detection algorithms in a comparison with our method using both simulated and real-world HSI datasets. The experimental results demonstrate that the proposed method outperforms others.

INDEX TERMS Hyperspectral image, anomaly detection, graphical data description, clustering, machine learning, support vector machine.

I. INTRODUCTION
Hyperspectral images (HSI) captured by hyperspectral sensors can contain a large number of approximately sequential spectral bands, separated by narrow intervals of usually less than 10 nanometers [1], [2]. As a result of the high spatial and spectral resolution of the sensors, an HSI contains abundant feature information of the scene. Target detection (TD) is one of the most popular research fields of HSI that aims to discover the interesting targets and separate them from the complex background [3]. However, most TD algorithms need to obtain the prior information of desired targets and background before detection [4], [5]. The prior knowledge of an HSI cannot be easily obtained in practice because the spectra of surface objects are often affected by atmospheric factors and environmental differences, e.g., the absorption and scattering of air and water vapor, change of illumination intensity, and the limited sensors’ resolution [6]. Therefore, the wide spread practical use of supervised target detection techniques is difficult. Anomaly detection (AD), as an unsupervised technique, has drawn significant attention in recent years. It can detect anomaly targets of a huge area without the need for prior feature information, and has been successfully applied in various domains, such as geology [7], agriculture [8], and public security [9].

In the preceding two decades, a large number of hyperspectral AD algorithms have been proposed. The most widely studied category is based on probability distributions.
Among them the Reed-Xiaoli (RX) is a benchmark due to its good real-time performance and reliability [10]. It assumes that the pixels in hyperspectral images follow a multivariate Gaussian distribution, and the differences between test pixels can be judged by calculating the respective Mahalanobis distance. A number of improved RX algorithms have been proposed as well. For instance, local RX (LRX) [11] restricts the multivariate Gaussian model to a local region of the HSI, and then processes the whole image by moving a dual-sliding window. However, the background distribution in HSI is very complex. Usually it does not obey the hypothesis of the multivariate Gaussian distribution model in practice. To address this problem, Kernel RX (KRX) [12], a typical nonlinear AD method, was proposed. It projects data from the original space to a high-dimensional feature subspace via a kernel function to detect the anomaly targets with abundant feature information. The subspace RX (SSRX) first performs the principal component analysis (PCA) of the original data and obtains several reconstructed bands with the largest energy, also called principal components. Afterwards, the RX method is used to process the principal components to obtain the AD result. Furthermore, researchers have also proposed the representation-based theory which has attracted much attention in recent years. Specifically, the collaborative representation-based detector (CRD) [13] assumes that background pixels can be well approximated by their spatial neighborhoods, and Chen et al. [14] has proposed an approach based on a sparse representation (SRD), which assumes that each test pixel can be represented by a few atoms in the dictionary. Another popular method called low rank representation (LRR) [15]–[17] assumes that the hyperspectral dataset can be represented by a constructed dictionary with various constraints on the coefficient matrix. A series of atoms can be combined to represent the pixels within a small neighborhood.

Regardless of the probability distribution model or the sparse representation model, the drawback is that there will be inevitable errors compared with the actual scene. Moreover, the more complex the improved model is, the better the detection effect, yet the model still cannot solve the fundamental problem. In addition, with the increase in the complexity of the algorithm, the real-time performance will decline. Accordingly, an effective strategy to address these problems is feature extraction. If we obtain sufficient feature information of different categories, it will be quite simple to distinguish them from each other. Among those feature extraction selections, learning-based methods are widely used [18]. There are various of learning based AD algorithms in recently years. Ma et al. proposed a Deep Belief Network (DBN)—based anomaly detector. The high-level features and reconstruction errors are learned via the network which is not affected by background distribution assumption. Zhang and Cheng proposed a stacked autoencoders (AEs)—based adaptive subspace model (SAEASM) to extract the in-depth features of differences using three windows. Xie et al. proposed a spectral constrained adversarial AE (SC−AAE) network which can suppress the background samples while preserving the characteristic of anomalies. The advantage of learning methods is that it can obtain feature information of data by training samples. As long as the sample quality is high and the number is sufficient, the obtained feature information will be more complete. However, most learning methods focus on the improvement of the network itself. They enhance the ability of extracting features by increasing the number of layers of the unsupervised network or making it adversarial. From our perspectives, traditional supervised learning algorithms can also be applied to unsupervised fields, as long as the training sample problem is solved. And compared to unsupervised learning methods, traditional supervised learning methods have advantages of robustness and stability. Support vector machine (SVM), as one of the most popular statistical learning algorithms, has become an attractive and popular tool in pattern recognition and machine learning. It has advantages of having fewer requirements to prior knowledge, being more suitable with small samples size and more robust to noise, and offering a higher learning efficiency [19]–[21]. Therefore, the remarkable advantages of SVM have been fully demonstrated in hyperspectral image processing. It has been widely used in the fields of band selection [22], compression coding [23], spectral classification [24], and spectral unmixing [25].

Based on the limited sample information, SVM seeks the best compromise between the model complexity and learning ability of obtaining the best generalization. Hyperspectral images can be classified by training samples without modeling the HSI dataset. However, due to the complexity and diversity of land materials in the image, if the SVM is used directly to process the whole data by dividing all kinds of materials into a set of unified categories, the error for anomalies and backgrounds will be large, and the anomalous components will contain plentiful background components meanwhile. From this perspective, in this paper we propose a multiple SVM strategy to improve the detection accuracy. First we utilize a graphical connected point estimation (GCPE) as a preprocessing operation to divide the original dataset into two components: potential anomalies and robust background. Graph theory is a branch of mathematics, that mainly describes a graph composed of a number of given points and lines connecting each other [26], [27]. The concept of graph estimation has been illustrated in [28]. After the preprocessing step, a clustering method is used to classify the background into several categories. For each category of the background, a part of data is selected as the training sample. The selected background sample and a part of the potential anomaly are jointly input into the SVM procedure to train the HSI dataset and to obtain the result of classification corresponding to each specific category. Finally, an effective fusion strategy is designed to fuse all the SVM-classified results and generate the final detection result.

To summarize, our proposed algorithm, called the graphical estimation and multiple SVM-based detector (GEsSVM), has the following three main contributions.
1) We employ a graph theory-based preprocessing operation to provide reliable samples for learning methods. Compared with other strategies that need to generate a model to fit a real-world scenario, the graphical construction method only uses the feature information of the data itself for processing. Hence, to some extent, it avoids the errors caused by the incompatibility between the model and the actual scene.

2) A multiple learning method based on SVM is proposed. There are many SVM-based multi-classification methods currently, such as 1-against-all (1-a-a) [29], [30] and 1-against-1 (1-a-1) [30], decision directed acyclic graph (DDAG) SVM [31], and error correcting output codes (ECOC) [32]. However, most of these methods require a large number of training samples of different categories to train the model, which is unrealistic to anomaly detection. In a sense, AD is still essentially a binary classification even if the background components of the image contain many material categories. The ultimate objective is to extract anomalies separated from the background. Accordingly, our proposed method takes advantage of pixels in the original image to perform binary classification in the text of different background categories, which effectively avoids the problem of insufficient samples, and generate a good classification performance.

3) An effective fusion strategy based on activation function is described in order to increase the distinction between the background and anomalies. A series of classification maps generated by multiple SVM-based detector are able to be fused and enhance the contrast meanwhile.

The remainder of this paper is organized as follows. In Section II, we mainly introduce the related work. The methodology of proposed algorithm is illustrated in Section III. Experimental and parameter sensitivity analysis are presented in Section IV. Finally, Section V concludes the paper.

II. RELATED WORKS
A. GRAPHICAL CONNECTED POINT ESTIMATION
For a HSI dataset, we denote it as $X = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^{B \times n}$, $n$ is the number of pixels in a spectral band and $B$ is the number of bands in a dataset. A graphical set is defined as $G = [V, S]$, where $V$ represents a vertex set, and $S \in \mathbb{R}^{n \times n}$ represents a similarity matrix. Each element $S_{ij}$ denotes the similarity between the pixels $x_i$ and $x_j$. To measure the similarity, we choose one of the most popular methods called geodesic distance [33], [34]. In a broad sense, geodesic distance refers to the local shortest path between two points in space. There are two common approaches measuring the distance in graph theory: the Dijkstra’s algorithm [35] and the Bellman-Ford algorithm [36]. Here we utilize Dijkstra’s algorithm to estimate the similarity of two pixels $S_{ij}$ because the time complexity of Dijkstra’s algorithm is much shorter than the other. For an HSI dataset with a large amount of data, the time complexities will vary significantly and affect the performance of the entire detection framework.

As for initializing the similarity matrix $S$, we utilize the $k$-nearest neighbor (KNN) approach [37] to construct the adjacency graph. First, the Euclidean distance between every two vertices $x_i$ and $x_j$ is computed; which is denoted by $E_{ij}$. Next, for a test vertex $x_i$, its neighborhood $X_{knn-i}$ defined by $k$ nearest neighbors is determined. If pixel $x_j$ belongs to $X_{knn-i}$, $S_{ij} = E_{ij}$. Otherwise, $S_{ij}$ is considered to be $\infty$. The usage of KNN specify the number of adjacent points. It avoids setting a fixed threshold by comparing the distance to judge if the point belongs to the adjacency graph, which may cause a “short circuit” or “open circuit” problem. The initialization of $S$ can be specified as (1)

$$S_{ij-initial} = \begin{cases} E_{ij}, & x_j \in X_{knn-i} \\ \infty, & x_j \notin X_{knn-i} \end{cases}$$

After initialization, the elements of $S$ will be updated by Dijkstra’s algorithm. The principle of Dijkstra’s algorithm is an iterative process based on a relaxation rule. The choice of the edge which needs to be relaxed is made using a minimum priority queue [38]. Therefore, the update rule for $S$ is written as (2)

$$S_{ij} = \min(S_{ij}, S_{ik} + S_{kj})$$

According to the minimum priority queue, we select the smaller value by comparing $S_{ij}$ and $S_{ik} + S_{kj}$, and update the similarity matrix $S$ accordingly. The operation is called “relaxation”. Once all elements in $S$ have been relaxed, each one represents the shortest distance between the corresponding two vertices in $X$. The distance between the components of the same or similar categories is very small, and the distance between the components with larger differences is relatively large.

B. PRINCIPLE OF SUPPORT VECTOR MACHINE
Assuming a linearly separable sample set can be represented as $(x_i, y_i), i = 1, 2, \ldots, l, x \in \mathbb{R}^d, y \in \{+1, -1\}$. The decision function in the $d$-dimensional space is $g(x) = wx + b$, and there is a pair $(w,b)$ that satisfies the following formula:

$$\begin{cases} (w \cdot x_i) + b \geq 0, y_i = +1 \\ (w \cdot x_i) + b \leq 0, y_i = -1 \end{cases}$$

(3)

The classification hyperplane equation is $wx + b = 0$. Sample points satisfying $(w \cdot x_i) + b \geq 0$ belong to positive class samples. Others satisfying $(w \cdot x_i) + b \leq 0$ are negative class samples. The interval between hyperplanes $wx + b = 1$ and $wx + b = -1$ is equal to $2\Delta$, where $\Delta$ is represented as

$$\Delta = \|wx + x + b - 1\| = \frac{1}{\|w\|}$$

(4)

The samples are normalized to make all of them satisfy $|g(x)| \geq 1$, while the nearest samples satisfy $|g(x)| = 1$. The problem of finding the optimal classification hyperplane can be transformed into the following constrained optimization
problems:

\[
\min \varphi(w) = \frac{1}{2} \|w\|^2 \\
\text{s.t. } y_i[(w \cdot x_i) + b] - 1 + \xi_i \geq 0 \quad i = 1, 2, \ldots, l
\]  

(5)

where \(\xi_i \geq 0\) is a relaxation term to adjust the degree of separating training samples by the hyperplane.

The Lagrange function is constructed as

\[
L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{l} a_i [y_i((w \cdot x_i) + b) - 1]
\]

(6)

where \(a_i \geq 0\) is the corresponding Lagrange multiplier for each sample point. The partial derivatives \(\frac{\partial}{\partial w}L(w, b, a) = 0\) and \(\frac{\partial}{\partial w}l(w, b, a) = 0\) for \(w\) and \(b\) respectively. After substituted into the Lagrange functions. The result is

\[
L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{l} a_i [y_i((w \cdot x_i) + b) - 1]
\]

\[
= \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j a_i a_j (x_i \cdot x_j)
\]

(7)

Therefore, the Wolfe dual problem of (5) can be expressed as:

\[
\begin{cases}
\max & \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j a_i a_j (x_i \cdot x_j) \\
\text{s.t.} & \sum_{i=1}^{l} a_i y_i = 0, a_i \geq 0, \quad i = 1, 2, \ldots, l
\end{cases}
\]

(8)

By solving the above problems the optimal solution \(a^* = (a_1^*, a_2^*, \ldots, a_l^*)^T\) can be obtained. According to the KT condition, \(a^*\) and \((w^*, b^*)\) satisfy \(a_i^* y_i (w^* \cdot x_i) + b^* - 1 = 0\).

Consequently, \(b^* = y_i - \sum_{j=1}^{l} y_j a_j^* (x_i \cdot x_j)\) can be calculated. The optimal classification function is obtained as

\[
f(x) = \text{sgn} \left\{ w^* \cdot x + b^* \right\} = \text{sgn} \left\{ \sum_{i=1}^{l} a_i^* y_i (x_i \cdot x) + b^* \right\}
\]

(9)

For linearly separable training samples, the hyperplane of the SVM will make a good binary classification, which is shown in Fig. 1(a). However, if all kinds of samples are mixed together, the selection of a linear hyperplane will result in larger errors. Using nonlinear mapping to project the input vectors to a feature space can address this problem effectively, as shown in Fig. 1(b). The classification can be achieved by using an appropriate kernel function \(K(x_i, x_j)\). The corresponding classification function is then written as follows:

\[
f(x) = \text{sgn} \left\{ \sum_{i=1}^{l} a_i^* y_i K(x_i \cdot x) + b^* \right\}
\]

(10)

Several kernel functions are commonly used in SVM. Here we preferentially use the radial basis function (RBF) kernel [39] because it is characterized by strong locality kernel function, and can facilitate nonlinear mapping. Moreover, compared with other kernel functions, the RBF kernel function needs fewer parameters.
represents the data of the hyperspectral image. For ease of understanding, the connectivity size of a neighborhood in the graph is set to two. Then the yellow squares represents the original data. After the initialization of $S$ and updated by Dijkstra’s algorithm, the green squares represent all the elements intensively connected according to the similarity principle. These elements can be regarded as the robust background in the hyperspectral image. In contrast, other regions that are unconnected to the main components are isolated, and are clearly distinguishable from the background components. These regions, represented by red squares, can be regarded as potential anomalies.

**B. PROPOSED MULTIPLE SVM METHOD**

After preprocessing, the potential anomalies are obtained. However, it is comparatively difficult to distinguish the real anomalies from some background components having the similar characteristics. Therefore, it is necessary to propose a refinement of anomalous component extraction. On the other hand, in order to obtain a more accurate classification result by SVM, it is important to select training samples reasonably. Since the number of potential anomalies is more than that of real anomalies, only some of them will be selected as training samples. At the same time, we will also select parts of robust background pixels because of their large amount, which can not only avoid over-fitting, but also improve training efficiency. In this paper, half of the potential anomalies are randomly selected as the training samples of an anomaly, and one-third of robust background pixels are selected as the training samples of the background. A label-assigning function is used to classify the corresponding pixels of the dataset into binary categories.

$$l_i = f(x_i) = \begin{cases} 1 & \text{anomaly} \\ 0 & \text{background} \end{cases} \quad (11)$$

If $x_i$ is classified as an anomaly, the label $l_i$ is assigned to 1. Otherwise, $l_i$ is equal to 0, which means $x_i$ is classified as background. Although the labels of background pixels are all assigned to 0 by SVM, there are various material categories containing in background. The radiation characteristics of some background materials are similar to real anomalies. To this end, in our proposed multiple-SVM strategy, robust backgrounds are firstly classified into several categories by clustering methods because it does not require so large number of samples like multi-classification. One of the most widely used clustering method, $k$-means [40], has a shortcoming that the initial clustering centers are selected randomly. If the initial centers are not selected well, the method may only reach the locally optimal value. To this end, we employ $k$-means++ [41], [42], an improved $k$-means method as our clustering choice. It can make the distance between the initial clustering centers as large as possible. This simple modification will effectively improve the clustering result. After $k$-means++ operation, the clustering result of each category will be regarded as background samples to train the model together with potential anomalies. A total of $K$ models, referred to as SVM model-$i$ ($i = 1,\ldots, K$) are obtained.

After that, the dataset is processed by each SVM model for binary classification. The data will be divided into anomalies and background with label $l_i$. The background component in each SVM model corresponds to a specific material. The process diagram of the multiple-SVM strategy is shown in Fig. 3.

![Process diagram of multiple SVM strategy](image)

**FIGURE 3.** The process diagram of multiple SVM strategy.

It is worth mentioning that in SVM model-$j$, for instance, the pixels are divided into anomalies and the $j$th material. But for other materials $i \neq j$, a few pixels with high brightness may be mistaken for anomalies. When $K$ classification results are generated, the errors of anomalies obtained by each result will be a little different. However, almost all real anomalies can still be successfully distinguished.

**C. FUSION STRATEGY**

After the HSI dataset processed by the multiple-SVM model, we can obtain $K$ different classification results. Then, all classification maps need to be fused into one single map as the final anomaly detection result. The most commonly used method is the simple average fusion [43], which only calculates an average value for the detection result. However, since all the values after averaging are greater than 0 and less than 1, it is realizable to further expand the difference between anomalies with larger values and background points with smaller values. Inspired by the activation function of neural networks, sigmoid function can be considered as a tool for further enhancement of the average result. The formula is written as

$$\text{sig}(x) = \frac{2}{1 + \exp(-a \cdot x)} - 1 \quad (12)$$

The graph of the sigmoid function with various values of parameter $a$ is shown in Fig. 4. It is observed that the function has a better enhancement ability at larger values of $a$. Here we set $a = 5$. After fusion by averaging and the sigmoid function, a simple RX is used to suppress the residual background components. The detailed steps of the proposed GEmSVM method are listed in Algorithm 1. The flowchart is shown in Fig. 5.
IV. EXPERIMENTAL AND PARAMETER SENSITIVITY ANALYSIS

A. HSI DATASET DESCRIPTIONS

In this section, three HSI datasets are utilized to evaluate the proposed GEmSVM method. The first is a synthetic dataset based on a real scene. The anomalous pixels from the outside are mixed with the corresponding background in different proportions and then embedded into the original dataset. It was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over San Diego, CA, USA. The dataset has 224 spectral bands ranging from 370 to 2510 nm and 3.5 m per pixel for spatial resolution. In the experiment, some low SNR bands and others absorbed by atmospheric water vapor (1-6, 33-35, 97, 107-113, 153-166 and 221-224) were removed, 189 effective bands were remained, as shown in Figure 7(a). The anomaly targets were implanted by [45], in which a desired anomaly with spectrum \( t \) is picked outside the chosen scene but in the whole image. A specified abundance fraction \( f \) is used to adjust the mixing degree of \( t \) and a given pixel of the background with spectrum \( b \). The target with spectrum \( z \) can be generated as follows:

\[
z = f \cdot t + (1 - f) \cdot b
\]  

(13)

In the experiment, 16 targets were embedded in the original AVIRIS dataset (four rows and four columns). \( f \) is 0.05, 0.1, 0.2 or 0.4 for different rows and the number of pixels is 1, 2, 3 or 4 for different columns respectively. \( t \) is chosen from the plane in the middle left of the whole scene. The image scene used for simulated data is 100 × 100 of the scene, as shown in Fig. 6(b), and the ground-truth map of anomalies is shown in Fig. 6(c).

The second dataset is also from the San Diego AVIRIS image, which also has 189 spectral bands after removing...
invalid bands. The size of tested image is $100 \times 100$ from up-left of the whole scene. The scene mainly includes three aircrafts, apron, runway, roof of the building and a small amount of vegetation. The aircrafts are anomaly targets to be detected, which consist of 57 pixels totally. The false color image and the ground-truth map are respectively shown in Fig. 7(a) and Fig. 7(b).

The third one is collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) over an urban area that includes a vegetation area, a construction area, and several roads including some vehicles. The spectral resolution of the image is 10 nm, containing 210 bands in total. After removing low SNR bands and others absorbed by atmospheric water vapor (1-4, 76, 87, 101-111, 136-153, and 198-210), 162 bands are remained. The image for experiment is from the up-right of the whole scene. The size of $80 \times 100$ from the whole $307 \times 307$, where vehicles and roofs with 21 pixels are contained as anomalies. The false color image and the ground truth of main backgrounds and anomalies are illustrated in Fig. 8a and Fig. 8b.

**B. DETECTION PERFORMANCE**

In our experiment, the proposed GEmSVM method is firstly preprocessed by GCPE. Fig. 9(a) shows the preprocessing result for the HYDICE dataset represented by a grayscale image. It is observed that many pixels in the image are regarded as potential anomalies and represented in black pixels, which contains all real anomaly components as well as many background pixels. On the other hand, potential anomalies can also be clearly distinguished from robust backgrounds meanwhile. Fig. 9(b) shows a binary image after the multiple-SVM approach, in which most of the background components have been eliminated. There are only real anomalies and very few background pixels remaining. The final detection result is shown in Fig. 9(c).

Several state-of-art anomaly detectors are utilized as comparison for evaluate the performance of the proposed method. These detectors include RX, local RX (LRX), kernel RX (KRX), SSRX, SRD, CRD, and the single-SVM that does not classify the background materials or fuse different maps by a specific fusion strategy. The experiments are performed on a PC with an Intel Core i3-4130 3.4 GHz CPU and 8GB of RAM. All algorithms are implemented in MATLAB R2017a.

For the first synthesized dataset, the results of the proposed algorithm and the comparison methods are shown in Fig. 10. All the parameters of algorithms are set to optimal value. We can see that both CRD and GEmSVM exhibit outstanding performances. However, in Fig. 10(f), the points in the first row, which have the weakest intensity, are not clearly visible. It proves that to a certain extent the background suppression also inhibits the target points. For the GEmSVM method, it ensures all targets are clearly displayed while the background is well suppressed. In addition, for RX and LRX, almost the whole images including all targets are suppressed, leaving only a small amount of background. For SSRX and SRD, if the background is sufficiently suppressed, the target pixels can also be observed relatively clearly except for the weakest row. For single-SVM, all targets are displayed, but there are also many false alarm points mistaken for anomalies. it is clear that GEmSVM has the best performance. However, only by visual inspection is far from obtaining an accurate estimation. We need to make a further quantitative comparison. The receiver operating characteristic (ROC) curves [46],
FIGURE 10. Two-dimensional maps of detection results by different algorithms for the simulated dataset. (a) RX; (b) LRX; (c) KLRX; (d) SSRX; (e) SRD; (f) CRD; (g) Single-SVM; (h) GEmSVM.

FIGURE 11. Two-dimensional maps of detection results by different algorithms for San Diego dataset. (a) RX; (b) LRX; (c) KLRX; (d) SSRX; (e) SRD; (f) CRD; (g) Single-SVM; (h) GEmSVM.

considered as the classic comparison measurement for different detection methods [44], [47], are shown in Fig. 13(a). The curve corresponding to a better performance is located in the top left of the coordinate plane. It can be seen that if the false alarm rate (FAR) is almost 0, the detection rate (DR) of our method is close to 100%. The curve of CRD is very close to GEmSVM, but a little bit lower. Other methods have much lower detection rates than the first two. In addition, we use background-anomaly separation map as another measurement method shown in Fig. 13(b). The red box represents the distributions of anomaly targets, and the green one denotes the distributions of background. It is clear that the gap between the two boxes of GEmSVM is the largest, which means it has the best separation performance for background and anomalies.

For the AVIRIS-San Diego dataset, the color detection results of all algorithms are shown in Fig. 11. We observe that RX, SSRX, and CRD can suppress almost all of backgrounds
FIGURE 12. Two-dimensional maps of detection results by different algorithms for HYDICE dataset. (a) RX; (b) LRX; (c) KLRX; (d) SSRX; (e) SRD; (f) CRD; (g) Single-SVM; (h) GEmSVM.

FIGURE 13. Quantitative analysis of the proposed GEmSVM and compared detectors for the simulated dataset. (a) ROC curves; (b) background-anomaly separation map.

FIGURE 14. Quantitative analysis of the proposed GEmSVM and compared detectors for San Diego dataset. (a) ROC curves; (b) background-anomaly separation map.

Successfully, but at the same time also suppress the desired targets. For KLRX and the single-SVM method, all anomaly pixels can be observed clearly, but there are also many background components as the false alarm. The detection result is affected significantly. For GEmSVM, even though the roof of the building in the bottom-left corner can be seen in the image, it has the best overall performance. Fig. 14(a) shows that the ROC curves of San Diego. None of these compared methods are located in the top-left corner because all methods are affected by more or less numerous false alarm pixels. But the DR metric of the proposed method can still reach 100% if FAR is nearly \(10^{-1}\). SRD and single-SVM can have rather high DR values if the FAR is less than \(10^{-1}\), but if the FAR is less than \(10^{-2}\), their detection rate will decline very quickly. Furthermore, via background-anomaly separation map in Fig. 14(b), the GEmSVM algorithm can separate background and anomaly targets effectively.

For the HYDICE dataset, the scene is more complex and the target size is smaller. Fig. 12 shows the detection results of the eight algorithms. The results of LRX reflect a good ability of background suppression for small targets, but several anomaly targets are suppressed as well. Both CRD and single-SVM algorithms can generate a better performance. However, there are also several false alarm points that cannot be ignored. For RX, although the background is suppressed well, there are also numerous of false alarm points in the right part that result in a significant impact. It can be seen that both SSRX and the proposed method achieve outstanding detection results. SSRX can better suppresses the background, but some target pixels in the lower-left corner of the image are also affected. For GEmSVM, the pixel in the corner can be observed clearly. According to Fig. 15(a), single-SVM and GEmSVM outperform other algorithms. The DR of GEmSVM is higher than single-SVM when FAR is larger than 0.2. From the background-anomaly separation map, as shown in Fig. 15(b), it is observed that GEmSVM can suppress the background more effectively.

In order to further compare the AD methods, we may prefer to turn the ROC curve into a single scalar-valued evaluation criterion. The area under a ROC curve (AUC) is one of the most widely used methods [48]. Generally, the higher the AUC value is, the better the detection performance...
of this method. The AUC value of each method is listed in Table 1, which further shows that the GEmSVM algorithm has the best detection performance among all algorithms in the experiment. The running times of the procedure for the GEmSVM method and all compared detectors are shown in Table 2. Compared with other algorithms, SRD takes the most time, and the running time of GEmSVM is a little longer than single-SVM due to the usage of a $k$-means++ clustering process and the application of SVM classification to each background category. However, because the single-SVM also needs a large number of samples from potential anomalies and robust backgrounds to train the model, the computational cost of GEmSVM does not increase too much to affect its efficiency.

### C. PARAMETER SENSITIVITY ANALYSIS

1) RELEVANT PARAMETER SETTING OF SVM

For SVM model training and learning in our experiment, one of the most important things is to select an appropriate kernel function. When the kernel function is utilized to solve the linear inseparable problem, the strategy usually adopted is to calculate on the low-dimensional feature space to avoid the huge calculation cost of the inner product of the vector in the high-dimensional feature space. Here we select the RBF kernel due to the advantages of not only the realization of the original training data being linearly partitioned in a high-dimensional space, but also no major consumption in computing. Here we set the parameter $\sigma$ of the RBF kernel $\sigma = 0.1$. Another necessary procedure of SVM is the N-fold cross-validation (N-CV). The original data is divided into N groups (usually equally sized). Each subset is treated as a validation set, and the remaining N-1 subsets are used as the training set. The average classification accuracy of the validation set for K models is applied as the performance indicator under this N-CV procedure. N-CV can effectively avoid the occurrence of overfitting and underfitting. In our experiment, N is chosen among {4,5,6,8,10}. As the value of N increases, so does the time required for N-CV. In the experiment, we have observed that the differences among AD performances of methods' results were not very large. But the difference of running time could be up to five times. Therefore, we take N = 5 so that the best detection performance is obtained as well as meet the requirement of timeliness.

### D. OTHER PARAMETER ANALYSIS AND DISCUSSION

There are three other important parameters about the proposed GEmSVM algorithm besides the SVM section: 1) the connectivity size of neighborhood $C$ in Dijkstra’s algorithm, 2) the number of categories $K$ in the $k$-means++ clustering algorithm, and 3) parameter $a$ in the sigmoid function within the fusion part. Parameter sensitivity analyses are performed for each of three datasets, and the other parameters are set to the optimal values when a specific parameter is discussed.

The AUC values of the GEmSVM for values of $C$ changing in the range of [6, 16] is shown by a histogram in Fig. 16. The symbol Inf in the figure indicates that there is more than one infinite value in each row or column of $S$. Thus, any two points in $V$ cannot be connected, and Dijkstra’s algorithm cannot be run successfully. When $C = 8, 9$ and $10$, the AUC values for the simulated dataset are the highest. The running time is relatively short with the increase of $C$, because the number of potential anomalous elements decreases when $C$ is larger. When $C$ is larger than 10, the subsequent AUC values will be affected by the lack of potential anomalous. Overall, we select $C = 10$ to maximize efficiency. Similarly, the best
FIGURE 16. The effects of connectivity size of neighborhood C in the proposed GEmSVM method on three HSI datasets.

detection performance at $C = 12$ and $C = 9$ is obtained for the San Diego and HYDICE datasets, respectively.

Next we evaluate the number of categories $K$ in $k$-means++ and parameter $a$ in the sigmoid function simultaneously. In the experiment $k$ is chosen in the range of {2,3,4,5,6,7,8,9}, and $a$ takes values in {1,2,3,4,5,6,7}, while other parameters remain fixed. Compared with $K$, $a$ has little effect on the algorithm, especially when $k$ is small, and the AUC value is changed very little as $a$ changes. At larger values of $K$, $a$ causes a sudden increase of AUC because SVM can generate more accurate classification results after clustering, so that the performance of the algorithm is significantly improved. The change of the value of $a$ will also have a prominent impact on the performance of the GEmSVM method in the fusion step. As $a$ increases, the AUC gradually stabilizes and thus grows slower. Fig. 17 shows that for the simulated data and the San Diego dataset, the AUC value is the highest when $a = 5$. For the HYDICE dataset, $a = 6$ is the best choice.

On the other hand, for parameter $K$, it can be observed that the number of clusters has a significant impact on the detection performance. If $K$ is small, the result of each dataset is poor. As $K$ increases to a certain value, the algorithm achieves the best detection performance. However, when $K$ continues to increase, the same category of components in the dataset may be forced to be divided into different clusters, which will result in interference with the subsequent SVM procedure and result in a poor detection performance. As shown in Fig. 17, the largest AUC value is obtained for the simulated data if $a = 5$. For the San Diego and HYDICE datasets, we set $a$ to 6. Furthermore, it can be observed that the influence of $a$ or $K$ is not characterized by a monotonic increase, but exhibits a certain degree of jitter. So they reflect a general trend.

V. CONCLUSION

In this paper, a novel hyperspectral anomaly detection algorithm is proposed. It combines the clustering method and one-class SVM to obtain more accurate training results for anomaly detection. To obtain potential anomalies, we utilize the graphical theory to design a graphical connected point estimation as a preprocessing step to distinguish potential anomalies from robust backgrounds. Afterwards, an improved multiple-SVM strategy is designed via binary classification. The $k$-means++ clustering method is utilized to classify robust backgrounds into several categories. For each category, we select some of the elements as training samples, and train with some selected potential abnormal elements using SVM to generate the training model. $K$ training models will be obtained, the number of which are equal to the number of clusters. Next, all data from the original image are processed by the $K$ models to generate $K$ different classification results. A fusion function with an adjustable parameter is used to enlarge the dissimilitude of anomalies and the background. Extensive experiments on both synthetic and real-world datasets demonstrate the superior detection performance of the proposed GEmSVM algorithm.

REFERENCES

[1] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, “Modern trends in hyperspectral image analysis: A review,” IEEE Access, vol. 6, pp. 14118–14129, 2018.
[2] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, “Imaging spectrometry for earth remote sensing,” Science, vol. 228, no. 4704, pp. 1147–1153, 1985.
[3] G. Shaw and D. Manolakis, “Signal processing for hyperspectral image exploitation,” IEEE Signal Process. Mag., vol. 19, no. 1, pp. 12–16, Aug. 2002.
[4] T. Wang, H. Zhang, H. Lin, and X. Jia, “A sparse representation method for a priori target signature optimization in hyperspectral target detection,” IEEE Access, vol. 6, pp. 3408–3424, 2018.
[5] J. Du and Z. Li, “A hyperspectral target detection framework with subraction pixel pair features,” IEEE Access, vol. 6, pp. 45562–45577, 2018.
[6] Y. Yuan, D. Ma, and Q. Wang, “Hyperspectral anomaly detection by graph pixel selection,” IEEE Trans. Cybern., vol. 46, no. 12, pp. 3123–3134, Dec. 2016.
[7] F. A. Kruse, J. W. Boardman, and J. F. Huntington, “Comparison of airborne hyperspectral data and co-1 hyperion for mineral mapping,” IEEE Trans. Geosci. Remote Sens., vol. 41, no. 6, pp. 1388–1400, Jun. 2003.
[8] C. M. Gevaert, J. Suomalainen, J. Tang, and L. Kooistra, “Generation of spectral–temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for precision agriculture applications,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 6, pp. 3140–3146, Jun. 2015.
[9] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, “Automated hyperspectral cueing for civilian search and rescue,” Proc. IEEE, vol. 97, no. 6, pp. 1031–1055, Jun. 2009.
[10] 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, Oct. 1990.
[11] J. Chen and I. Reed, “A detection algorithm for optical targets in clutter,” IEEE Trans. Aerosp. Electron. Syst., vol. AES–23, no. 1, pp. 46–59, Jan. 1987.
[12] 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.
[13] W. Li and Q. Du, “Collaborative representation for hyperspectral anomaly detection,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
[14] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Sparse representation for target detection in hyperspectral imagery,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 3, pp. 629–640, Jun. 2011.
[15] G. Liu, Z. Lin, and Y. Yu, “Robust subspace segmentation by low-rank representation,” in Proc. 27th Int. Conf. Mach. Learn. (ICML), 2010, pp. 663–670.
[16] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, “Anomaly detection in hyper spectral images based on low-rank and sparse representation,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
[17] Y. Yang, J. Zhang, S. Song, and D. Liu, “Hyperspectral anomaly detection via dictionary construction-based low-rank representation and adaptive weighting,” Remote Sens., vol. 11, no. 2, p. 192, Jan. 2019.
[18] S. Song, H. Zhou, H. Li, S. Song, and D. Liu, “Hyperspectral anomaly detection via convolutional neural network and low rank with density-based clustering,” IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens., vol. 12, no. 9, pp. 3637–3649, Sep. 2019.
[19] V. N. Vapnik, “An overview of statistical learning theory,” IEEE Trans. Neural Netw., vol. 10, no. 5, pp. 988–999, Sep. 1999.
[20] J. Guan, R. Lai, and A. Xiong, “Wavelet deep neural network for stripe noise removal,” IEEE Access, vol. 7, pp. 44544–44554, 2019.
[21] P. Xiang, H. Zhou, H. Li, S. Song, T. J. Song, and L. Gu, “Hyperspectral anomaly detection by local joint subspace process and support vector machine,” Int. J. Remote Sens., vol. 41, no. 10, pp. 3798–3819, May 2020.
[22] H. Zhang, H. Bo, and Y. Le, “Kernel function in SVM-RFE based hyperspectral data band selection,” Remote Sens. Techn. Appl., vol. 28, pp. 747–752, Mar. 2013.
[23] Y.-C. Liu, Z.-Y. Chen, J.-T. Fang, and P.-C. Chang, “SVM-based fast intra CU depth decision for HEVC,” in Proc. Data Compress. Conf., Apr. 2015, p. 458.
[24] Y. Wang and H. Duan, “Classification of hyperspectral images by SVM using a composite kernel by employing spectral, spatial and hierarchical structure information,” Remote Sens., vol. 10, no. 3, p. 441, Mar. 2018.
[25] F. A. Mianji and Y. Zhang, “SVM-based Unmixing-to-Classification conversion for hyperspectral abundance quantification,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 11, pp. 4318–4335, Nov. 2011.
[26] B. Scholkopf, J. Platt, and T. Hofmann, “Geometric entropy minimization (GEM) for anomaly detection and localization,” in Proc. Adv. Neural Inf. Process. Syst., vol. 53, pp. 585–592, 2006.
[27] A. Frome, Y. Singer, F. Sha, and J. Malik, “Learning globally-consistent local distance functions for shape-based image retrieval and classification,” in Proc. IEEE 11th Int. Conf. Comput. Vis., Oct. 2007, pp. 1–8.
[28] S. Song, H. Zhou, Y. Yang, K. Qian, J. Du, and P. Xiang, “A graphical estimation and multiple-sparse representation strategy for hyperspectral anomaly detection,” Inf. Phys. Technol., vol. 99, pp. 212–221, Jun. 2019.
[29] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415–425, Mar. 2002.
[30] Y. Tian, Z. Qi, X. Ju, Y. Shi, and X. Liu, “Nonparallel support vector machines for pattern classification,” IEEE Trans. Cybern., vol. 44, no. 7, pp. 1067–1079, Jul. 2014.
[31] N. Cristianini, C. Campbell, and J. Shawe-Taylor, “Dynamically adapting kernels in support vector machines,” in Proc. Conf. Adv. Neural Inf. Proc. Syst. II, 1999, pp. 204–210.
[32] T. Dietterich, “Ensemble methods in machine learning,” in Proc. Int. Workshop Multiple Classifier Syst. Berlin, Germany: Springer, 2000, pp. 1–15.
[33] S. Even, Graph Algorithms. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[34] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, “Exploiting manifold geometry in hyperspectral imagery,” IEEE Trans. Geosci. Remote Sens., vol. 43, no. 3, pp. 441–454, Mar. 2005.
[35] F. Schulz, D. Wagner, and K. Weihe, “Dijkstra’s algorithm on-line: An empirical case study from public railroad transport,” J. Experim. Algorithmics, vol. 5, p. 12, Dec. 2000.
[36] C. Cheng, R. Riley, and S. P. R. Kumar, “A loop-free extended Bellmanford routing protocol without bouncing effect,” in Proc. Symp. Comm. Arch. Protocols, 1989, pp. 224–236.
[37] P. Soucy and G. W. Mineau, “A simple KNN algorithm for text categorization,” in Proc. IEEE Int. Conf. Data Mining, Nov. 2001, pp. 647–648.
[38] T. H. Cormen, Introduction to Algorithms. Cambridge, MA, USA: MIT Press, 2009.
[39] D. Wang, D. S. Yeung, and E. C. C. Tsang, “Structured one-class classification,” IEEE Trans. Syst. Man, Cybern. B, Cybern., vol. 36, no. 6, pp. 1283–1295, Dec. 2006.
[40] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A k-means clustering algorithm,” Appl. Statist., vol. 28, no. 1, pp. 100–108, 1979.
[41] D. Arthur and S. Vassilvitskii, “K-means++: The advantages of careful seeding,” in Proc. 18th Ann. ACM-SIAM Symp. Disc. Alg., 2007, pp. 1027–1035.
[42] X. Li, Q. Lu, Y. Dong, and D. Tao, “SCE: A manifold regularized set-covering method for data partitioning,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 5, pp. 1760–1773, May 2018.
[43] D. Ma, Y. Yuan, and Q. Wang, “Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation,” Remote Sens., vol. 10, no. 5, p. 745, May 2018.
[44] A. Taghipour and H. Ghassemian, “Hyperspectral anomaly detection using attribute profiles,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 7, pp. 1136–1140, Jul. 2017.
[45] L. Zhu and G. Wen, “Hyperspectral anomaly detection via background estimation and adaptive weighted sparse representation,” Remote Sens., vol. 10, no. 2, p. 272, Feb. 2018.
[46] J. Kerekes, “Receiving operating characteristic curve confidence intervals and regions,” IEEE Geosci. Remote Sens. Lett., vol. 5, no. 2, pp. 251–255, Apr. 2008.
[47] G. Manolakis and G. Shaw, “Detection algorithms for hyperspectral imaging applications,” IEEE Signal Process. Mag., vol. 19, no. 1, pp. 29–43, Aug. 2002.
[48] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, Jun. 2006.
YIXIN YANG received the B.S. degree in electronic science and technology from Xidian University, Xi’an, China, in 2014, where she is currently pursuing the Ph.D. degree in optical engineering with the School of Physics and Optoelectronic Engineering.

Her research interests include hyperspectral image processing, target and anomaly detection, low-rank and sparse representation, and machine learning and feature extraction.

ZHE ZHANG received the B.E. degree in electronic science and technology from Xidian University, Xi’an, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Physics and Optoelectronic Engineering. His research interests include target detection and recognition, visual object tracking, and hardware implement.

HUIXIN ZHOU (Member, IEEE) received the Ph.D. degree from Xidian University, Xi’an, Shaanxi, China, in 2004.

He was with Xidian University, in 2004, where he is currently the Deputy Dean of the School of Physics and Optoelectronic Engineering. He has published more than 100 articles. He holds more than 20 authorized patents and eight software copyrights. His current research interests include optoelectronic imaging and real-time image processing, target detection and tracking, high/hyperspectral image processing, and so on. He is a Senior Member with the Photoelectronic Technology Professional Committee of the Chinese Society of Astronautics and the Chinese Optical Society. He is also a member of The Optical Society, USA. He received the first prize for technology innovation from the colleges and universities by the Ministry of Education, China, in 2015.

* * *