Hyperspectral Anomaly Detection by Graph Pixel Selection

Yuan Yuan, *Senior Member, IEEE*, Dandan Ma, and Qi Wang, *Senior Member, IEEE*

***Abstract*—Hyperspectral anomaly detection (AD) is an important problem in remote sensing field. It can make full use of the spectral differences to discover certain potential interesting regions without any target priors. Traditional Mahalanobis-distance-based anomaly detectors assume the background spectrum distribution conforms to a Gaussian distribution. However, this and other similar distributions may not be satisfied for the real hyperspectral images. Moreover, the background statistics are susceptible to contamination of anomaly targets which will lead to a high false-positive rate. To address these intrinsic problems, this paper proposes a novel AD method based on the graph theory. We first construct a vertex- and edge-weighted graph and then utilize a pixel selection process to locate the anomaly targets. Two contributions are claimed in this paper: 1) no background distributions are required which makes the method more adaptive and 2) both the vertex and edge weights are considered which enables a more accurate detection performance and better robustness to noise. Intensive experiments on the simulated and real hyperspectral images demonstrate that the proposed method outperforms other benchmark competitors. In addition, the robustness of the proposed method has been validated by using various window sizes. This experimental result also demonstrates the valuable characteristic of less computational complexity and less parameter tuning for real applications.**

***Index Terms*—Graph theory, hyperspectral anomaly detection, manifold learning.**

Manuscript received May 28, 2015; revised September 6, 2015; accepted October 29, 2015. Date of publication November 20, 2015; date of current version November 15, 2016. This work was supported in part by the National Basic Research Program of China (Youth 973 Program) under Grant 2013CB336500, in part by the State Key Program of National Natural Science of China under Grant 61232010, in part by the National Natural Science Foundation of China under Grant 61172143, Grant 61379094, and Grant 61105012, in part by the Natural Science Foundation Research Project of Shaanxi Province under Grant 2015JM6264, in part by the Fundamental Research Funds for the Central Universities under Grant 3102014JC02020G07 and Grant 3102015BJ(II)JJZ01, and in part by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences. This paper was recommended by Associate Editor X. He. *(Corresponding author: Qi Wang.)*

Y. Yuan is with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China.

D. Ma is with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China.

Q. Wang is with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: crabwq@nwpu.edu.cn).

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

Digital Object Identifier 10.1109/TCYB.2015.2497711

I. Introduction

HYPERSPECTRAL imagery (HSI) processing has been increasingly popular. The captured hyperspectral data cube delivers rich information containing both spatial and high spectral dimensions [1]–[3]. Owing to its reliable and nearly continuous spectra, hyperspectral image can provide discriminative clues to identify the subtle differences of different ground objects [4], which leads to many successful applications [5], [6].

As a greatly important and fundamental issue, anomaly detection (AD) [7]–[9] for HSI has been researched a lot for many years. It can be applied in precision agriculture [10], food and drug testing in public safety [11], rare mineral discovery in geology [12], civilian search and rescue [13], and man-made objects distinction in intelligent defense [14], [15]. In essence, AD is a binary classification problem which classifies the pixel under test as an target or a background [16]. Different from the supervised target detection problem [4], [17]–[19], in which the spectral information of detected target is known as a reference, hyperspectral AD is an unsupervised one without any prior spectral information about the target or background [20], [21]. This technique is more appropriate and consistent with the actual situation, because in most cases the accurate spectrum signatures of many interesting signals or targets cannot be provided as a prior in advance.

Generally, the aim of AD in hyperspectral remote sensing is to locate a target whose distinct spectrum deviates significantly from the surrounding background [22]. Based on this preliminary, a wealth of detection methods have been proposed over the last two decades [23]. Conventional detectors need some rigorous assumptions on the spectrum distribution of background, which is not fully reasonable for the real collected hyperspectral data. Another fact is that because of the low spatial resolution of hyperspectral images, a pixel may cover a large range of area which possesses rich surface materials. As a result, its spectrum signature is typically complex and maybe a mixture of different components of ground objects. Naive assumptions in this case are unable to model the data and may limit their generative ability for a new hyperspectral data.

In order to overcome these limitations existing in traditional methods, this paper proposes a novel scheme based on manifold learning [24], [25] and graph theory [26]. Two main contributions are claimed in this paper and they are summarized as follows.

1) Neither requiring specific assumptions about the background statistics, nor estimating its covariance matrix,

the proposed detector directly constructs a graph discovering the internal relationships of the hyperspectral pixels. The main concept is that a background pixel seems to have close and solid relationships with its surroundings, while an abnormal pixel tends to build distant and fragile relationships with its neighbors. By this strategy, no background distributions should be estimated and it is more adaptive to various kinds of hyperspectral image and more befitting to the practical applications.

2) Not only taking the edges' contributions into consideration, but also considering the vertexes' dedication, we construct a vertex- and edge-weighted graph to guarantee a superior performance of AD. Instead of only observing the relationships between the examined pixel and its k-nearest neighbors like traditional methods, the proposed method explores all the connections of all pairs of pixels. This makes the detector more robust to the interference by noise or specific computation errors, which has been validated by the sensitivity experiments on the window sizes and noise levels.

The remainder of this paper is organized as follows. In Section II, the related works of hyperspectral AD are reviewed. In Section III, the proposed method is described in detail, including the manifold learning reconstruction, the construction of the vertex- and edge-weighted graph, and the abnormal pixel selection process. In Section IV, extensive experiments are conducted on both the simulated and real hyperspectral images to demonstrate the superiority of the proposed method. Finally, the conclusion is drawn in Section V.

II. Related Work

As a pattern recognition problem, AD is used to locate objects that are extraordinary compared with the cluttered background [27]. According to the definition of background scope, existing methods can be roughly divided into two categories, global anomaly detectors (GAD) and local anomaly detectors (LAD) [28]. For the GAD, background is defined with the reference to all the image pixels or a large part of them. And for the LAD, background refers to a small neighborhood of the pixel under test. In this paper, for the sake of organizing the literatures, we adopt another classification standard with the assumption about the background spectrum distribution [29]. There are also two categories, similarly. The first one assumes the whole background conforms to a homogenous spectral distribution or one single type, while the second one assumes the background consists of various classes with different distributions.

1) The well-known Reed Xiaoli (RX) algorithm proposed by Reed and Yu [30], is a typical type of the first kind, which assumes a simple background distribution. This algorithm is deemed to be the benchmark anomaly detector for both multispectral and hyperspectral data. The hypothesis of RX is that the background around a target conforms to the same multivariate normal distribution. It relies on the well-known Mahalanobis distance to measure the difference degree of a spectral vector from its surrounding neighbors. RX usually makes use of a sliding window, in which the observed pixel occuping the center and the rest in the window are used to estimate the background statistics.

However, RX has two intrinsic problems that may result in a poor performance in many applications. For one thing, a Gaussian distribution cannot accurately describe the real hyperspectral image due to various surface materials and complicated process of acquiring signals. For another, the small-sample size is the most well-known challenge. When a high-dimensional background covariance matrix is estimated by a small number of samples, it leads to a badly-conditioned matrix and an unstable result. Based on these considerations, two types of methods are explored to overcome the encountered issues. The first type can be considered as an improved version of RX algorithm. For example, the global RX (GRX) utilizes the whole hyperspectral image to calculate a global covariance matrix. The subspace RX [1] adopts a complement projection operator to suppress the background clutter. The random-selection-based anomaly detector (RSAD) [31] utilizes a random selection process to better compute the background clutter. Besides, the selective kernel principal component analysis RX [32], minimum covariance determinant RX [33], compressive RX [34], regularized-RX (RRX) [35], etc, are also variants of the original RX algorithm. However, the intrinsic problem of simple distribution is still unsolved essentially since the original assumption of RX has not been removed.

Another type relies on the kernel theory, which maps the original data into a higher dimensional feature space through a nonlinear mapping. The greatest strength of kernel methods is that the discrimination between the target and the background can be enhanced in a high-dimensional feature space. Furthermore, kernel methods can map the Gaussian distributed data into nonlinear Gaussian, which has been proved to be beneficial to AD [20], [36]–[38]. For instance, the Kernel-RX (KRX) [36] is a parametric kernel-based algorithm, which can be considered as a nonlinear form of the RX algorithm in a high dimensional feature space. Since KRX still has the issue of anomaly targets' contamination to background estimation in the kernel space, a method named robust nonlinear AD (RNAD) [39] has been proposed recently to solve it. Through utilizing a regression detection strategy to suppress the contamination, RNAD can improve the performance of KRX. Support vector data description (SVDD) [40] is a nonparametric kernel-based anomaly detector, which can directly analyze the support region avoiding the prior assumption. SVDD supposes that the background is enveloped by a minimum enclosing hypersphere in a high-dimensional feature space. Anomalies are identified as those who fall outside this hypersphere. However, the kernel based methods always suffer from the expensive computation burden. Especially for SVDD, its computation is very expensive.

2) For the second kind of complex background distribution, the background is supposed to contain multiple classes with different distributions [41]. The typical processing roadmap is to estimate the information about different classes respectively and then to analyze them synthetically. The cluster-based anomaly detector (CBAD) [42] is a representative example. It first segments the hyperspectral image into different clusters,

and then the original RX is performed in each cluster separately. Another example named multivariate normal inverse Gaussian (MNIG) detector [43] is based on MNIG assumption. It supposes each class conforms to a MNIG distribution. Then for a pixel under test, its negative log-likelihood is estimated with respect to the specific class distribution it belongs to. Since the performance of AD heavily relies on the clustering, it is important to accurately estimate the cluster number. When the number of classes is underestimated, the difference between classes can be decreased which leads to a poor detection performance. And when the number of classes is overestimated, anomaly targets may be classified into one cluster. Consequently, the target cannot be detected [16].

III. Hyperspectral Anomaly Detection

In this paper, we utilize a pixel selection process based on the graph theory [44], [45] to realize AD. The proposed method mainly relies on the philosophy that a pixel will be considered as an anomaly when it is easily picked out from the graph with a high probability. This is equal to say the relations between the abnormal pixels and the background pixels in the graph are more vulnerable than the other relations. There are mainly three steps for the proposed method: 1) manifold learning reconstruction [46], [47]; 2) vertex- and edge-weighted graph construction; and 3) anomaly pixel selection.

First, a manifold feature learning technique named locally linear embedding (LLE) is applied. With the aid of reconstruction error of the examined pixel by its nearest neighbors, the vertex weight can be estimated. After that, we establish a vertex-and edge-weighted graph. The graph takes all the relationships among vertexes into consideration and the weight between two arbitrary vertexes is calculated in the Euclidean space. With the obtained graph, a pixel selection process is carried out to locate the abnormal target. The pixel with higher probability being picked out from the whole graph is more inclined to be a target. Fig. 1 shows the overall flowchart of the proposed method.

A. Manifold Learning Reconstruction

This part will introduce the manifold learning technique known as LLE [48] for the later definition of vertex weight. Manifold learning is based on the philosophy that the data embedded in a high-dimensional vector space can be mapped into a low-dimensional manifold space. During this mapping, it is able to reserve the coherent data structure corresponding to that in the original high-dimensional feature space. Considering that the LLE has been widely used to analyze a variety of problems in hyperspectral image processing field, especially for hyperspectal AD [23], [49], [50], the proposed method adopts LLE to effectively describe the relationships between hyperspectral pixels.

The LLE algorithm is successful in discovering the underlying manifold structure by a linear reconstruction of local neighbors. Suppose the D-dimensional data point $X_i \in R^D$ corresponds to a pixel in a hyperspectral image I. Therefore, given a hyperspectral patch $X \in R^{N\times D}$ containing N pixels, each pixel X_i can be expressed by its local neighbors with a linear combination. When the K nearest neighbors are considered, the reconstruction error ε_i can be measured by the following formula:

$$\varepsilon_i = \left\| X_i - \sum_{j=1}^{K} w_j^i X_j \right\|^2 \tag{1}$$

where w_j^i denotes the contribution of the jth neighbor of X_i. When taking all the pixels into consideration, the extended version of the reconstruction error can be obtained

$$\varepsilon = \sum_i \left\| X_i - \sum_{j=1}^{K} w_j^i X_j \right\|^2. \tag{2}$$

Through minimizing the cost function, the weights w_j^i are estimated finally. Two constrains are imposed on w_j^i in this procedure. First, the pixels participating in the reconstruction of X_i must belong to their local neighborhood scope; otherwise zero will be enforced on w_j^i strictly. Second, for an observed pixel, all the weights contributing to its reconstruction should be summed to one: $\sum_j w_j^i = 1$. The aforementioned two constrains can bring with numerous benefits to the minimization problem.

Through a linear mapping involving translation, rotation, and rescaling, the original data in the high-dimensional feature space are mapped into a global coordinate system on the low-dimensional manifold. In the course of mapping, the reconstruction weights reflecting the intrinsic geometric properties of data set are designed to keep invariant. That means the mapped data Y_i in its embedded low-dimensional manifold coordinates, should also be reconstructed by the same weights w_j^i from its neighbors. Therefore, the cost function in the embedded manifold can be denoted as following:

$$\varepsilon(Y) = \sum_i \left\| Y_i - \sum_{j=1}^{K} w_j^i Y_j \right\|^2 \tag{3}$$

where $Y_i \in R^d$ is the mapped low-dimensional vector ($d \ll D$). By the minimizing the above equation with the fixed weights, Y_i is finally obtained through solving a sparse eigenvector problem.

In summary, through the LLE technique, each pixel in the hyperspectral image can be expressed by its neighbors with a reconstruction error. The larger the error is, the less similar it is to its neighbors. Benefiting from this virtue, we use these reconstruction error obtained from (3) to stand for the weight of vertex in the graph.

B. Vertex- and Edge-Weighted Graph Construction

Traditional methods rely on the assumptions about the spectral distribution of background statistics. Nevertheless, these assumptions are usually hard to satisfy due to the complexity of real hyperspectral image in practice, which can lead to the failure of AD. There have been graph-based methods to tackle these limitations, but the results are still far from satisfying. For example, Basener *et al.* [51] proposed a topology-based AD algorithm (abbreviated as GTAD for

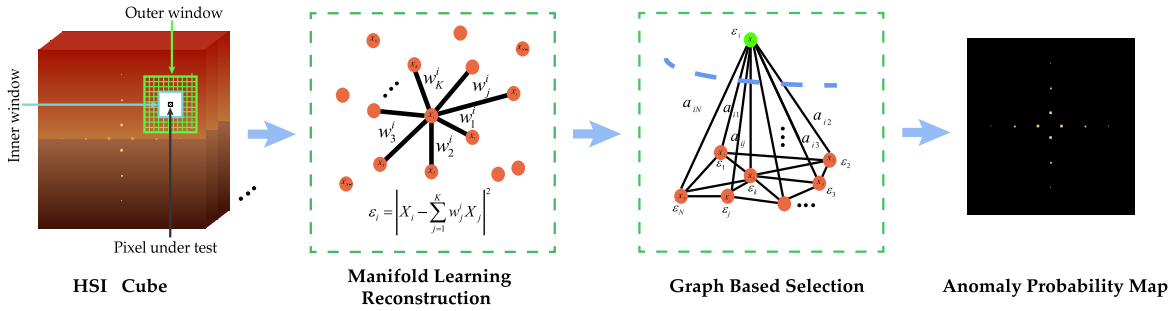

Fig. 1. Flowchart of the proposed method.

short in this paper) in dimensionally large data sets. This paper builds a graph first and its edges connect those close pairs of points in spectrum level. An observed pixel is identified as background when the number of its nearest neighbors reaches a certain percentage for a given distance radius. Recently, Messinger and Albano [52] (named as ADuT in this paper) constructed a spectral graph in high-dimensional space. The edge of the graph is generated between two pixels (named as vertexes) if they satisfy a similarity criteria. According to the nearest neighbor approach, the degree and weight can be calculated. Based on these two values, an anomaly can be detected through computing weighted vertex volume (WVV). In addition, there are also some literatures based on constructing a graph to carry out hyperspectral target detection. Ziemann and Messinger [50] proposed a method applying the graph model into the LLE technique in order to adaptively determine the number of the nearest neighbors. Munoz *et al.* [53] adopted topological AD to obtain the model of background, which can be seen as a preprocessing operation.

In this paper, we propose a novel method for AD in hyperspectral images via constructing a vertex- and edge-weighted graph from a new view. The philosophy behind our method is that the anomaly pixels tend to be picked out more easily than the background pixels in the established graph. This is because an anomaly pixel usually deviates from the background, and its distinctiveness makes its connections with other background pixels vulnerable.

We define the vertex- and edge-weighted graph as $G = (V, E, \varepsilon, A)$. Here $V = \{X_1, X_2, ..., X_N\}$ is the vertex set of size N, each of which corresponds to a pixel in the hyperspectral image. $E \subseteq V \times V$ is the edge set and ε represents the vertex weight set obtained by using the LLE algorithm discussed in the previous part. $A = \{a_{ij}\}$ denotes the $N \times N$ symmetric weight matrix, where

$$a_{ij} = \sqrt{\sum_{l=1}^{D} (X_{il} - X_{jl})^2}. \tag{4}$$

The element a_{ij} represents the similarity of each pair of hyperspectral pixels (vertexes), which is computed by Euclidean distance. The less the similarity is, the larger the distance is. So far, we have finished constructing the vertex- and edge-weighted graph.

C. Abnormal Pixels Selection

With the graph constructed from a hyperspectral image patch, the anomaly pixels can be identified. The assumption is that an anomaly vertex usually has fragile relationships with others and tends to be more easily selected from the graph. Before detailed introduction of the pixel selection process, an affinity matrix is firstly defined

$$\hat{A} = \Pi A \Pi \tag{5}$$

where $\Pi = \text{diag}(\varepsilon)$. In this paper, we expect vertex's contributions and edge's contribution can further affect each other, which makes the difference between anomaly and background more significant. Therefore, we formulate the (5) to achieve this purpose. The effect of (5) is equal to the fact that each edge weight is multiplied by two reconstruction errors corresponding to the two vertexes connected by the edge. As an anomaly usually tends to have a larger edge weight and a higher reconstruction error with respect to the background, (5) can make the large difference more salient and the small difference more not manifest. Consequently, the obtained affinity matrix is capable of considering the relationships of edges and vertexes. It has properties of non-negativity, symmetry, and monotonicity. To be specific, for any two pixels in a hyperspectral image, $\hat{A}_{ij} \geq 0, \hat{A}_{ij} = \hat{A}_{ji}$. Moreover, with respect to ε_i, ε_j and a_{ij}, $\hat{A}_{ji}$ will be monotonically increasing. These properties are exactly consistent with the quality of our AD issue. For this problem, an anomaly tends to have higher reconstruction errors and associated edge distances, corresponding to a $\hat{A}_{ij}$ with pretty larger value. Therefore, the affinity matrix reflects the intrinsic properties of targets and background.

With the obtained affinity matrix $\hat{A}$, we then introduce the abnormal pixel selection process, which seems like the subset selection [54]. Let vector $P \in R^N$ denote the abnormal probability and each element p_i is the probability of an pixel being selected from the graph. Consequently, the problem of selecting an abnormal pixel from N data points can be formulated as

$$\begin{aligned} \max \quad & \frac{1}{2} P^T \hat{A} P \\ \text{s.t.} \quad & P \in \Delta \end{aligned} \tag{6}$$

where $\Delta = \{P \geq 0, \mathbf{1}^T P = 1\}$.

This is a quadratic programming problem and we refer to the replicator dynamics [54] to find the optimal solutions. Give an initialization of $P(0)$, through the following iteration which

is a discrete-time version of first-order replicator equation, the corresponding local solution can be efficiently computed

$$p_i(t+1) = p_i(t)\frac{\left(\hat{A}P(t)\right)_i}{P(t)^T\hat{A}P(t)}, i = 1, ..., N. \quad (7)$$

The simplex Δ is invariant under these dynamics [55], [56], which means that $P(t+1)$ will automatically satisfy the constraint $\mathbf{1}^T P = 1$ when $P(t)$ satisfies it. In other words, every trajectory initiating in simplex Δ will maintain in Δ for the future subsequent times [55], [56]. Since $\hat{A}$ is symmetric and non-negative, the objective function in (7) will strictly increase. When converged, its asymptotically stable points are equal to the local solutions. Therefore, we will obtain the anomaly probability of each pixel.

The above description covers the main steps of the proposed method. For the hyperspectral AD community, anomaly is often detected by sliding a window, which is also adopted by our method. Inspired by Banerjee *et al.* in [40], a hollow-window is designed in order to make the anomaly more significant. For an examined pixel, there are two surrounding windows including the inner and outer ones. The inner window can be treated as a guard window, which has the virtue of preventing the bad influence caused by other anomaly pixels. This is because when similar anomalies exist in a nearby region, the examined anomaly pixel will have a smaller reconstruction error, leading to a missed detection. But with a proper inner window, the pixels within it will be excluded and the region between the inner and outer windows is taken as the referential background. This has a higher probability to suppress the occurrence of missed detection. In this paper, s_{out} and s_{in} denote the sizes of outer window and inner window, respectively.

Our method adopts this sliding window strategy to examine the whole image gradually, which is called LGAD for short. For each sliding window N represents the total number of the graph's vertexes including the examined central pixel and the pixels between inner window and outer window. In practical processing, the indexes (or coordinates) of N pixels are known. Therefore, when the output $N \times 1$ probability vector P is obtained, the corresponding anomaly probability of the examined pixel can be determined based on its index (or coordinate). Then through sliding the window to traverse the whole image, the anomaly probability of each pixel can be obtained finally, which compose the anomaly probability map of the whole image. For a further explanation, anomaly probability maps with range [0, 1] are utilized to visualize the detection results, in which a larger value represents a higher anomaly probability.

In summary, once the vertex set is obtained, the LLE algorithm is operated firstly on the vertex data set to compute the reconstruction errors, which are regarded as the vertex weights. Then, the edge weights are calculated between any pairs of pixels in order to obtain an affinity matrix. After that, the vertex- and edge-weighted graph can be obtained. Finally, an anomaly pixel selection is formulated as a quadratic programming problem to fulfill the task. With the above introduction, the proposed LGAD is finally summarized in Algorithm 1.

Algorithm 1 LGAD for Hyperspectral Image Anomaly

Input:

Parameter setting: K, d, I, s_{in}, s_{out}.

Method:

For each sliding window

1: Obtain a patch X of the observed hyperspectral image I based on the given window sizes;
2: Get the reconstruction errors ε by (3);
3: Compute the weight matrix A by (4);
4: Obtain the affinity matrix $\hat{A}$ by (5);
5: Solve the objective function (6) by the replicator dynamics (7).

End

Output:

Anomaly probability map.

IV. Experiments and Analyses

In this section, we conduct several experiments on both simulated and real hyperspectral images to evaluate the effectiveness of the proposed method. First, the employed hyperspectral data sets are introduced. Then we describe the experimental setup consisting of parameter setting, evaluation metrics, and comparison methods. Finally, the experimental results and a detailed analysis are presented.

A. Data Set

We verify the performance of our proposed method on four kinds of hyperspectral images including two simulated data sets and two publicly available real-world data sets. The simulated data sets are used to directly verify the proposed method due to their definite anomaly targets, while the real-world hyperspectral data sets estimate the practical performance of our method.

The first data set is the simulated scene. It has 200×200 pixels with 105 bands and is characterized by the complex background distribution and pure anomaly pixel, which means the background consists of the combination of several material spectra while each anomaly target is a single materia spectrum. The synthetic procedure is described in detail as follows. Five kinds of material spectra including lawn grass, dry long grass, blackbrush leaf, sage brush, and tumblewe are selected from the U.S. Geological Survey vegetation spectral library.[1] For the background, the top 100 lines are compounded of dry long grass and lawn grass, and the bottom 100 lines are constituted by blackbrush leaf and sage brush. In this process of synthesis, the mixture percentage of each material is randomly generated. For the targets, pure spectra of sage brush, dry long grass, and tumblewe are used, respectively, to denote the definite anomaly pixels corresponding to three different kinds of size. They have a size of 1×1, 2×2, and 3×3. A total of 12 anomaly targets are embedded sparsely in the background. Fig. 2(a-1) and (a-2) shows the simulated scene and specific locations of the anomaly targets.

[1] http://speclab.cr.usgs.gov/spectral-lib.html

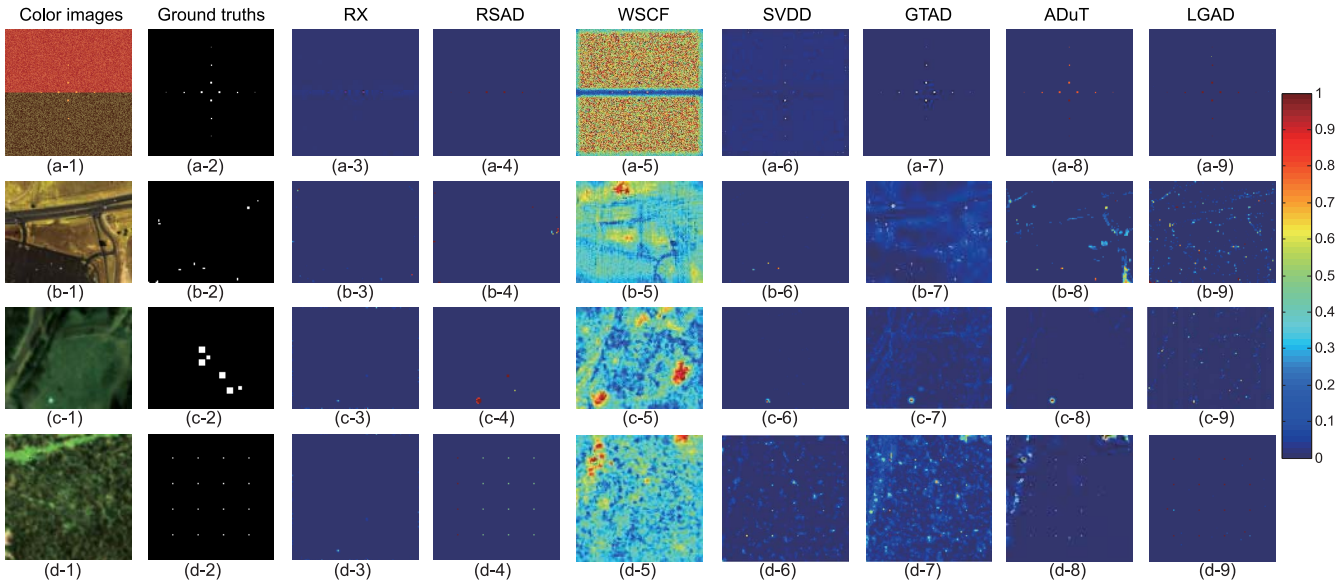

Fig. 2. False color images of the HSIs, the ground truth masks and the visualization of the detection results. The first column shows the false color pictures of the simulated scene, the HYDICE urban data set, the self test data set, and the synthetic subpixel scene, respectively. The second column illustrates their corresponding ground truths. The detection results of six methods including RX, RSAD, WSCF, SVDD, GTAD, ADuT, and LGAD are presented orderly from the third column to the eighth column.

The second data set is the HYDICE urban data set. It is a real-world hyperspectral image downloaded from the Website of U.S. Army Engineer Research and Development Center.[2] This reflectance spectrum data is collected by HYDICE on an airborne platform, covering an urban scene with spectral resolution of 10 nm, spectral range of 400–2500 nm, and spatial resolution of 1 m. The original hyperspectral image has a size of $307 \times 307 \times 210$. According to the band removal rule [23], we remain 160 bands finally by removing these bands 1–4, 76, 87, 101–112, 136–153, and 197–210, which have low-signal-to-noise ratio (SNR) and water vapor absorption. Because the ground truth for the whole scene is difficult to determine, in this paper, we crop a sub-image of 80×100 pixels from the upper right of the whole image, which contains several cars and roofs regarded as anomalies. The corresponding ground truth of the sub-image is defined consulting the work [23], [49]. The sub-image and its ground truth are shown in Fig. 2(b-1) and (b-2), respectively.

The third data set is the self test data set. It is also a real-world data set collected by an airborne HyMap sensor around the small town of Cooke City, Montana, USA. This HSI can be downloaded online at http://dirs.cis.rit.edu/blindtest/. It has 126 spectral bands with the reflectance spectral range covering from 0.45 to 2.5 micrometers, and the spatial resolution is about 3 m. In the image, there are several real targets located in the large grass background. And target locations are also provided directly. Considering the high computational costs of AD methods when dealing with a large size of hyperspectral image, we also crop a sub-image to conduct the experiments. The sub-image has a size of $80 \times 100 \times 126$ containing six targets. The color image of our sub-image and the corresponding ground truth are shown in Fig. 2(c-1) and (c-2).

The fourth data set is the synthetic subpixel scene. Considering that the former three HSIs have only pixel-level anomaly targets, we simulate this image to demonstrate the detection performance on subpixel targets. In order to resemble the actual situation, we crop a grass region of $100 \times 100 \times 126$ from the self test data set as background and add the synthesized targets. For the subpixel abnormal targets, we synthesize them based on the following equation: $T_{\text{sub}} = B \times (1 - p) + T \times p$, where T is the original pure target spectrum, B is the selected background spectrum, and p is the mixture percentage of anomaly target. In implementation, 100 pixels are randomly selected from the background region and their average spectrum are treated as the background spectrum B. Two definite targets F_1 (red cotton target) and F_2 (yellow nylon target) provided by the data set are set as the anomaly spectra T. We implant 16 anomaly targets of 1×1 into the background and they are arranged by columns, the p of which corresponds to 0.8, 0.5, 0.5, and 0.3. Since the spectrum of each target is impure, it leads to a subpixel AD. The illustration is shown in Fig. 2(d-1) and (d-2).

B. Experimental Details

Before presenting and analyzing the experimental results in detail, let us introduce three key points including the evaluation criteria, the competitors and parameter setup in the following part.

1) Evaluation Criteria: In order to analyze and compare the performance of anomaly detectors fairly, a valid evaluation criteria is very important.

In this paper, three popular criteria are adopted. The first one is the receiver operating characteristic (ROC) curve, which

[2]http://www.erdc.usace.army.mil/Media/FactSheets/FactSheetsArticleView/tabid/9254/Article/476681/hypercube

can be regarded as a classic comparison measurement. The curve describes the relationship between the target detection rate and the false alarm rate. When a discrimination threshold is determined, a set of values of these two rates will be computed. Therefore, a curve can be plotted with various threshold settings. In addition, the area under the curve (AUC) derives from ROC, which is an integration of the ROC. Besides, another evaluation metric called Bhattacharyya distance (BD) proposed by Yuan *et al.* [29] is used as an auxiliary in this paper. This metric reflects the detectors' capabilities of separating targets from background.

2) Competitors: To verify the effectiveness of the proposed method, we employ a number of benchmark methods to compare.

In this paper, local RX, SVDD, whitening and spatial correlation filtering (WSCF) [57], RSAD, GTAD, and ADuT serve as the competitors. We select these detectors taking a global view of popularity, recency, and variety. RX containing local and global versions is the most conventional method, which is always taken as the inevitable competitor. SVDD is a very popular method with the top citation and SVDD has derived many variants. WSCF applies a 2-D spatial filtering to realize a fast AD. RSAD is a very recent method which uses an idea of random sampling. GTAD and ADuT are two representative methods based on graph model. Owing to the virtues of these six popular benchmark methods, the comparison is more convinced which can objectively reveal the superiority of the proposed method. Besides, the single-based anomaly detector (SBAD) [22] is used to compare with the proposed method on the performance of subpixel detection.

3) Parameter Setup: In this part we will elaborate some critical parameters involved in the experiments.

First, two main parameters lie in the process of LLE, which are the number of nearest neighbors K and the dimension of low manifold space d. In our implementation, we set $K = 20, d = 10$ based on the performance on a benchmark dataset, and the detailed parameter selection will be discussed in Section IV. Then for the graph construction, we need to ascertain the sizes of windows. Since all the comparative methods except for GTAD and ADuT make use of a sliding hollow window, we use the same size for all the detectors. Considering different hypespectral images have different targets, proper sizes are pretty necessary. In this paper, we fix the sizes of outer window and inner window to be $s_{\text{out}} = 17$ and $s_{\text{in}} = 7$. As for SVDD, we adopt an approximate minimax technique introduced in [40] to determine its kernel parameter. In addition, the number of nearest neighbors in GTAD is set as 20 according to [51] and the size of its separated tiles is 20×20. As for ATuD, the involved parameters are consistent with the original work [52].

C. Comparison Results

In this section, the performance of our proposed algorithm is evaluated by comparing with six benchmark competitors. The experimental results on different HSIs will be analyzed thoroughly and detailedly in a sequential manner.

1) Simulated Scene: The visualization of the detection results on this simulated data is shown in Fig. 2(a-2)–(a-9). It can be obviously found that our LGAD is capable of detecting all the abnormal targets with a high intensity. RX and RSAD can perform well on some of the anomalies which seem to be more manifest than others. The performance of WSCF is poor, which nearly fails to distinguish anomalies from background. Though SVDD can also succeed in detecting the overall targets, the discrimination of its detected targets is not significant compared with the background. The performance of GTAD and ADuT is excellent for their detected targets are salient. But their intensities are relatively lower than LGAD, which demonstrates the superior target identification capability of the proposed method. As shown in Fig. 3(a) and (e), LGAD demonstrates its quantitative superiority than RX, RSAD, WSCF, and SVDD on account of the highest ROC curve and the largest AUC and BD values. Similarly, both GTAD and ADuT also have the same good performance on this simulated image. These phenomena verify the effectiveness of graph-based methods. At the same time, our method's performance proves that the constructed vertex- and edge-weighted graph enhances the distinctiveness of targets from the background. Moreover, without an assumption on the background statistics, LGAD possesses better adaptability to a complex background. It cannot only highlight more anomaly pixels, but also has a remarkable ability to suppress the background.

2) HYDICE Urban Data Set: Figs. 2(b-2)–(b-8) and 3(b) and (f) illustrate the detection results on this data. From Fig. 2(b-2) to 2(b-8), the visual detection results are illustrated in sequence corresponding to RX, RSAD, WSCF, SVDD, GTAD, ATuD, and our LGAD. It can be seen that WSCF hardly works on this image either, which mainly highlights the boundaries and fails to identify the abnormal targets. The global RSAD method tends to make the rarely appeared pixels stand out easily. However, this is insufficient to recognize all the anomalies. As for other competitors, RX and SVDD both have similar omissions but in different places. GTAD, ATuD, and LGAD all achieve quite nice results. However, the LGAD is even more outstanding than GTAD and ATuD, because it almost can detect all the anomaly targets with a very high anomalous probability. However, it is clear that an accurate estimation cannot be obtained by visual inspection. We need to make a further quantitative comparison. According to the comparison results plotted in Fig. 3(b) and (f), it can be obviously observed that the proposed method obtains a highest AUC values and its ROC curve stays over the other curves when the false positive rate slightly increases, which demonstrate the superior performance of LGAD. As for the BD indicator, our value is only lower than GTAD. Overall, the detection performance of is satisfying.

3) Self Test Data Set: The comparison results are presented in Figs. 2(c-2)–(c-8) and 3(c) and (g). Analyzing carefully from these figures, we can see that the proposed LGAD is significantly superior to the other competitors for its good effectiveness to recognize anomaly target. As shown in Fig. 2(c-3)–(c-8), RX, RSAD, SVDD, GTAD, and ATuD all have many misses on this image, and WSCF detects many

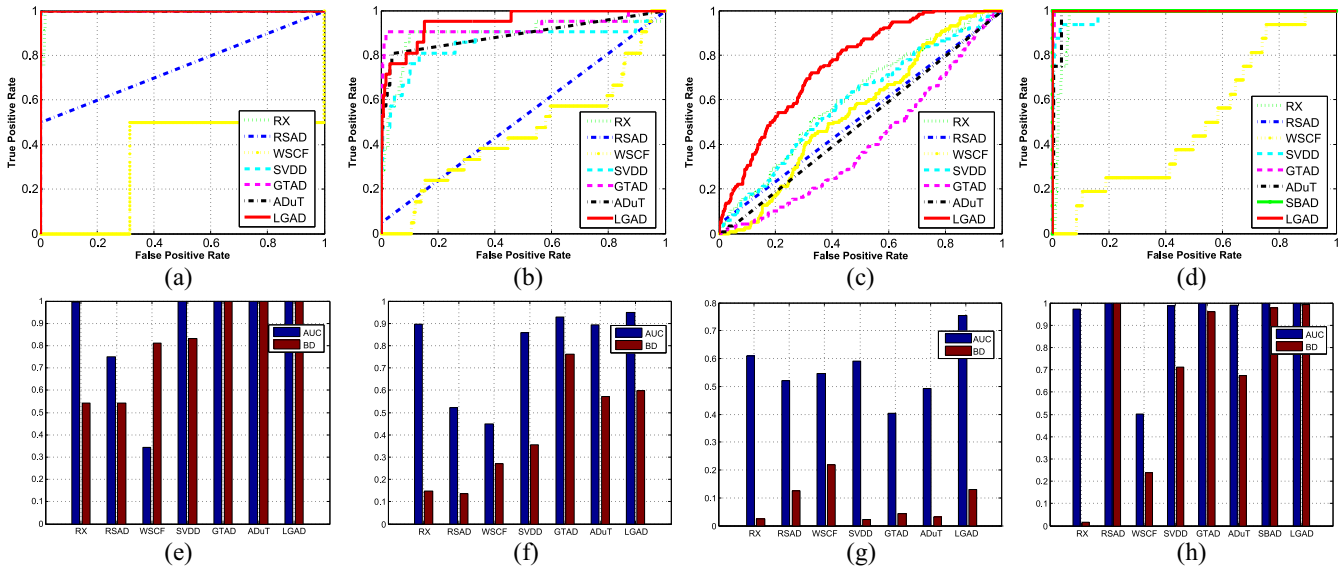

Fig. 3. Quantitative comparison of the AD results by different algorithms. The first row presents the ROC cures and the second row shows the AUC and BD bars. From left to right, the columns respectively correspond to the comparative results on the simulated scene, the HYDICE urban data set, the self test data set, and the synthetic subpixel scene.

textures which is unfortunately ineffective to AD. As observed from the ROC curves shown in Fig. 3(c), our method completely defeat all the other competitors, because its curve keeps over other curves. This fully confirms its good ability to detect anomaly. Fig. 3(g) illustrates the performance under AUC and BD. The AUC value of LGAD is larger than all the other competitors' values. The case of BD is slightly different. Although LGADs ability to suppress the background is not the best, it has achieved a larger values comparing with most of the benchmark methods. The excellent detection ability makes LGAD more effective and useful in practice.

4) Synthetic Subpixel Scene: This data is simulated to verify the detection ability of the proposed method to the subpixel anomaly targets. Figs. 2(d-3)–(d-8) and 3(d) and (h) illustrate the detection results on this scene. Since the SBAD is only implemented on this subpixel dataset, we have not shown its visual result in Fig. 2 for the convenience of layout, but its quantitative results are shown in Fig. 3(d) and (h). In Fig. 2(d-3)–(d-8), it can be found the anomaly probability maps of RSAD and LGAD are close to the ground truth mask, who assign high anomalous values to the subpixel targets. However, LGAD is a note above RSAD, because all kinds of subpixel targets detected by LGAD are all salient while RSAD does not. SVDD, GTAD, and ADuT yield visually similar results, but they can badly suppress the background. RX seems not to assign high values to the targets, which reveals its weak ability to distinguish targets from background. The performance of WSCF is quite poor as discussed before. Fig. 3(d) and (h) clearly shows that both RSAD and LGAD are successful, simultaneously on the values of AUC and BD. Besides, the BD value of the propose method is just higher than the SBAD method which is a specific subpixel anomaly detector. In summary, the excellent accuracy proves the effectiveness of the proposed LGAD to detect the subpixel anomalies. The reason is that LGAD simultaneously takes advantage of the vertex and edge weights. These two weights estimate the differences between targets and background from different aspects, which can strengthen the distinctiveness of anomalies. Therefore, the proposed method has the capability to detect these subpixel anomalies.

D. Discussion

In this part, several problems related to the proposed method are discussed. The first one is about the robustness to noise. The second one analyzes the sensitivity of the proposed method to the window sizes.The third one is about the selection of the parameters. In the end the computational complexity is discussed.

1) Robustness to Noise: The robustness to noise is a crucial and necessary aspect for the proposed method in the practical conditions. In order to evaluate this property, the spectral correlated noise [58] is added to each testing image, keeping the SNR ranging from 60 to 20 dB with an interval of 10 dB. In addition, the shape parameter η is set to 0.18 according to [29]. Therefore, we can consequently compare the performance of different detectors under distinct noise levels.

Fig. 4 shows the quantitative comparison. It is remarkable that our LGAD can cope with all the significant levels of noise. Moreover, for the AUC values, LGAD really achieves the best performance on all circumstances. As for the BD indicator, although LGAD cannot always keep its superiority on all the HSIs, it still ranks the best or the second in most time. In order to make a more manifest illustration, the amounts of relative changes before and after adding noise are compared. Table I presents the averaged relative changes over the five kinds of noise levels for different datasets. In Table I, HSI1, HSI2, HSI3, and HSI4, respectively, represent the four employed HSIs. It can be seen clearly that GTAD, ADuT, and LGAD are absolutely predominant compared to other benchmarks RX,

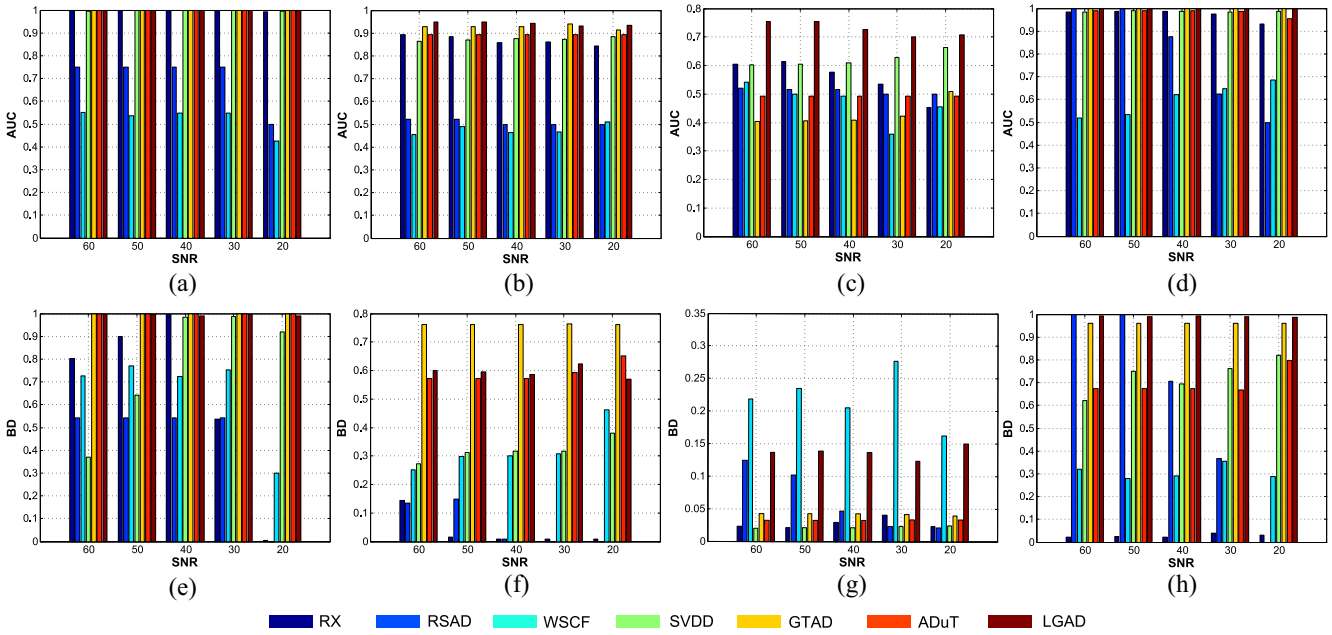

Fig. 4. Quantitative comparison of the detection results on the noisy images with different levels of spectral correlated noises. From left to right, the columns respectively correspond to the comparative results on the simulated scene, the HYDICE urban data set, the self test data set, and the synthetic subpixel scene.

TABLE I
AMOUNTS OF RELATIVE CHANGES ON THE PERFORMANCES BEFORE AND AFTER ADDING NOISE. FOR A CLEARER AND FAIRER COMPARISON, RED ONE IS THE BEST RESULT, GREEN ONE REPRESENTS THE SECOND RESULT FOR THE EACH DATA

		RX	RSAD	WSCF	SVDD	GTAD	ADuT	**LGAD**
AUC	HSI 1	0.0036	0.0500	0.1789	**0.0014**	**0.0000**	**0.0000**	**0.0000**
	HSI 2	0.0268	0.0139	0.0277	0.0139	**0.0055**	**0.0003**	0.0081
	HSI 3	0.0544	**0.0099**	0.0776	0.0308	0.0253	**0.0000**	0.0260
	HSI 4	0.0175	0.2000	0.0994	**0.0019**	**0.0000**	0.0069	**0.0000**
BD	HSI 1	0.3236	0.1082	0.1561	0.2090	**0.0000**	**0.0000**	**0.0064**
	HSI 2	0.1141	0.0829	0.0618	0.0464	**0.0006**	**0.0202**	**0.0143**
	HSI 3	0.0057	0.0613	0.0287	0.0015	**0.0008**	**0.0004**	0.0100
	HSI 4	0.0110	0.3854	0.0683	0.0609	**0.0006**	**0.0259**	**0.0028**

RSAD, WSCF, and SVDD, which strongly demonstrates the good robustness to noises of the graph based methods. Overall, LGAD is insensitive to noises for its smaller relative changes on different datasets, which is satisfying in spite of not outperforming GTAD and ADuT in each case. But we think this is not important because the difference between LGAD and GTAD, ADuT is not large and the relative changes of LGAD are also very close to zero.

2) Sensitivity to the Sizes of Windows: For a method using a window strategy, its performance is closely related to its size. However, it is difficult to find a perfect size to satisfy all the hyperspectral images. Therefore, a capable method should have a good tolerance to the window size, which means it can has a decent performance even not with the most appropriate window.

In order to demonstrate the adaptability to different window sizes, we compare the proposed method with the representative of competitors. Since the computational complexity of SVDD is unpractically expensive, its sensitivity experiment on window sizes is not included. For the other three methods of RX, RSAD, and WSCF, considering both their performance and recency, RSAD is chosen as the representative to compare with the proposed LGAD. This is because RSAD makes an effort to obtain a proper subset of background, which is able to reduce the contamination of anomaly targets to background statistics. With this knowledge, we can presume RSAD should have better robustness to the changes of window sizes than RX and WSCF. Therefore, in this paper only comparing LGAD with RSAD is feasible and persuasive.

The comparative experiments are conducted on the four hyperspectral images. 18 different window sizes $s_{\text{out}} \times s_{\text{in}}$ (9×3, 13×3, 13×7, 13×9, 17×3, 17×7, 17×9, 17×13, 21×3, 21×7, 21×9, 21×13, 21×17, 25×3, 25×7, 25×9, 25×13, and 25×17) are tested, respectively, while other parameters are fixed. Fig. 5 shows the statistics of AUC and BD values with box graphs. The black dot inside each box represents the median value, the upper and lower edges of the box are 25th and 75th percentiles, the whiskers signify the most extreme values, and the symbol "+" marks the outliers.

In both cases, the median value of LGAD for each HSI is significantly higher than that of RSAD. This means that the general performance of LGAD is much better than the other.

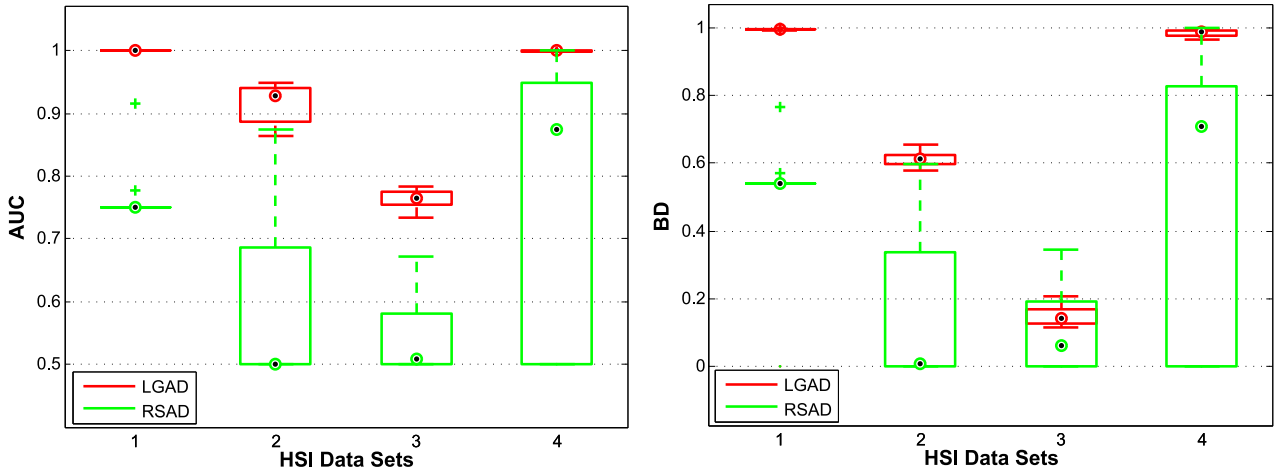

Fig. 5. Performance of LGAD and RSAD on the employed HSIs under 18 different sizes of windows. The left figure shows the values of AUC, and the right one illustrates the values of BD. The four data set respectively corresponds to the simulated scene, the HYDICE urban data set, the self test data set, and the synthetic subpixel scene.

For the first data set, RSAD has more outliers, which indicates its instability. For the other three data sets, RSAD has larger boxes, which implies the larger variance of its performance. From these figures, we can see clearly that for these frequently-used window sizes on different HSIs, the performance of our method do not have a violent vibration and is satisfactory in general. Therefore, the proposed method is robust to the window sizes. This characteristic makes our method more applicable in the real situation. In our implementation, comprehensively considering both the possible scales of anomalies and the appropriate parameter choices for all the competitors, we use a fixed window size 17×7 in all the experiments.

3) Parameter Selection: After discussing the effects of different window sizes, we need to analyze another two parameters in the LLE procedure, which are the number of nearest neighbors K and the dimension of low manifold space d. The additional experiment is performed with the HYDICE urban data set to analyze the effects of these two parameters. K and d is jointly considered. The values of parameter K are 7, 9, 15, 20, 25, 30, and 40, respectively, and ten different d values are also adopted including 4, 6, 8, 10, 12, 15, 20, 30, 40, and 50, while other parameters are fixed. Considering the AUC value is a popular and universal indicator, we discuss these two parameters' effects mainly by analyzing the changes of AUC values.

The results under different parameter combinations are illustrated in Fig. 6. Each curve represents the AUC values with the increase of dimension d under a given K. Generally speaking, it can be found that when d changes from 8 to 20, the AUC values are relatively stable for a fixed K, while the detection performance begins to deteriorate sharply when d reaches 20. Meanwhile, when K is given a larger value changing from 20 to 40, these curves get close to each other and have strong similarity. Moreover, when $K = 20$, the corresponding AUC curve almost achieves the highest performance, especially $d = 10$ at the same time. Therefore, it can be concluded that the proposed method performs well when $K = 20$ and $d = 10$.

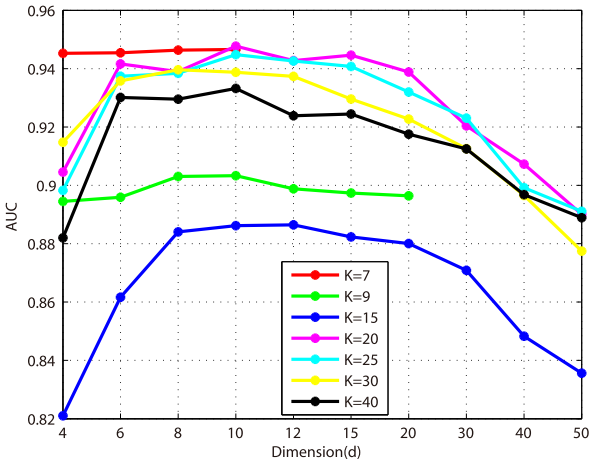

Fig. 6. Performance of the proposed method under different values of K and d on the HYDICE urban data set.

TABLE II
COMPARISON OF RUNNING TIMES (S)

	HSI 1	HSI 2	HSI 3	HSI 4
RX	93.34	72.00	58.56	48.81
RSAD	632.81	10.50	7.06	7.68
WSCF	3.21	1.00	0.99	0.73
GTAD	10.29	0.93	0.81	0.62
ADuT	3927.51	10.57	14.81	9.24
SVDD	83553.40	15672.51	20906.53	15668.66
LGAD	10256.36	2180.89	2846.60	2204.61

4) Computational Complexity: All the detectors are implemented on a machine with Intel Core i3-2130 3.4-GHz CPU and 16-GB RAM in the MATLAB platform. The time cost of each method is compared in Table II. The main time consumption for the proposed method comes from two aspects: 1) LLE procedure for the reconstruction errors and 2) abnormal pixels selection for anomaly probability. Since the LGAD adopts a sliding hollow window strategy, we will analyze the

time consumption in each sliding. Therefore, in LLE processing, we need $\mathcal{O}(DN^2 + DNK^3 + dN^2)$ [59] computation time to obtain the reconstruction errors. As for the abnormal pixels selection, replicator dynamics procedure costs the main computational time. For the input N pixels (vertexes), the total number of edges in the vertex- and edge-weighted graph is $\mathcal{O}(N(N-1)/2)$. Suppose the average number of iterations for the replicator equation is t, then the time complexity involved in the abnormal pixels selection is $\mathcal{O}(tN(N-1)/2)$. Therefore, the proposed method needs $\mathcal{O}(DN^2+DNK^3+dN^2+tN(N-1)/2)$. The SVDD implemented in our experiment is also a local method, whose time complexity is $\mathcal{O}(N^3)$ according to [60]. Generally, N is much larger than K and d. Therefore, compared with SVDD, our method is faster. Moreover, this superiority will be manifest with the increase of N. Although our method's efficiency is not dominant compared with the other five competitors except SVDD, its performance outperforms them on all the datasets. We believe that the efficiency of the proposed method can be drastically improved through code optimization or GPU acceleration in the future.

V. Conclusion

AD for hyperspectral image is very conventional and important. But its performance is far from satisfying. In this paper, we propose a novel graph based method to detect anomaly targets. Without any assumptions on the distribution of background statistics, our method is more adaptive to different kinds of real-world hyperspectral images. It can discover the intrinsic relationships among pixels via constructing a vertex- and edge-weighted graph. In order to verify the effectiveness of the proposed method, extensive experiments have been conducted on both simulated and real-world hyperspectral images. Six benchmark methods representing the state-of-the-art are also used as competitors. The results demonstrate the proposed method is more superior than the competitors. The proposed method not only takes advantage of effectiveness, but also has good robustness to noise and adaptability to window sizes. These desirable characteristics make the proposed method more applicable and efficient in the real situations.

In the future, we aim to explore the spatial relationships between pixels to refine the constructed graph. We also plan to investigate the weight definition between graph vertexes. We think these two aspects can lead to a more accurate detection result.

References

[1] D. W. J. Stein *et al.*, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[2] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.

[3] C.-I. Chang, H. Ren, and S.-S. Chiang, "Real-time processing algorithms for target detection and classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 760–768, Apr. 2001.

[4] L. Zhang, D. Tao, X. Huang, B. Du, and L. Zhang, "Hyperspectral remote sensing image subpixel target detection based on supervised metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4955–4965, Aug. 2014.

[5] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[6] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Compression of hyperspectral remote sensing images by tensor approach," *Neurocomputing*, vol. 147, pp. 358–363, Jan. 2015.

[7] S. T. Sarasamma and Q. A. Zhu, "Min-max hyperellipsoidal clustering for anomaly detection in network security," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 4, pp. 887–901, Aug. 2006.

[8] Y. Yuan, J. Fang, and Q. Wang, "Online anomaly detection in crowd scenes via structure analysis," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 548–561, Mar. 2015.

[9] S. T. Sarasamma, Q. A. Zhu, and J. Huff, "Hierarchical Kohonenen net for anomaly detection in network security," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 2, pp. 302–312, Apr. 2005.

[10] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 337–352, 2004.

[11] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging—An emerging process analytical tool for food quality and safety control," *Trends Food Sci. Technol.*, vol. 18, no. 12, pp. 590–598, 2007.

[12] F. A. Kruse, J. W. Boardman, and J. F. Huntington, "Comparison of airborne hyperspectral data and EO-1 hyperion for mineral mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1388–1400, Jun. 2003.

[13] 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.

[14] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Linc. Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[15] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, Jun. 2014.

[16] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.

[17] S. Matteoli, N. Acito, M. Diani, and G. Corsini, "Local approach to orthogonal subspace-based target detection in hyperspectral images," in *Proc. 1st Workshop Hyperspect. Image Signal Process. Evol. Remote Sens.*, Grenoble, France, 2009, pp. 1–4.

[18] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

[19] S. Matteoli, N. Acito, M. Diani, and G. Corsini, "An automatic approach to adaptive local background estimation and suppression in hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 790–800, Feb. 2011.

[20] H. Kwon and N. M. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 178–194, Feb. 2006.

[21] Q. Du and H. Ren, "Real-time constrained linear discriminant analysis to target detection and classification in hyperspectral imagery," *Pattern Recognit.*, vol. 36, no. 1, pp. 1–12, 2003.

[22] S. Khazai, A. Safari, B. Mojaradi, and S. Homayouni, "An approach for subpixel anomaly detection in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 769–778, Apr. 2013.

[23] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.

[24] Y. Deng, Y. Li, Y. Qian, X. Ji, and Q. Dai, "Visual words assignment via information-theoretic manifold embedding," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1924–1937, Oct. 2014.

[25] Y.-M. Zhang, K. Huang, X. Hou, and C.-L. Liu, "Learning locality preserving graph from data," *IEEE Trans. Cybern.*, vol. 44, no. 11, pp. 2088–2098, Nov. 2014.

[26] W. Zhao, Z. Guan, and Z. Liu, "Ranking on heterogeneous manifolds for tag recommendation in social tagging services," *Neurocomputing*, vol. 148, pp. 521–534, Jan. 2015.

[27] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.

[28] Z. Yuan, H. Sun, K. Ji, Z. Li, and H. Zou, "Local sparsity divergence for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1697–1701, Oct. 2014.

[29] Y. Yuan, Q. Wang, and G. Zhu, "Fast hyperspectral anomaly detection via high-order 2-D crossing filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 620–630, Feb. 2015.
[30] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
[31] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.
[32] Y. Gu, Y. Liu, and Y. Zhang, "A selective KPCA algorithm based on high-order statistics for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 43–47, Jan. 2008.
[33] S. Matteoli, M. Diani, and G. Corsini, "Hyperspectral anomaly detection with kurtosis-driven local covariance matrix corruption mitigation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 532–536, May 2011.
[34] J. E. Fowler and Q. Du, "Anomaly detection and reconstruction from random projections," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 184–195, Jan. 2012.
[35] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," in *Proc. SPIE Def. Security Symp.*, Orlando, FL, USA, 2008, Art. ID 696604.
[36] 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.
[37] P. Gurram, H. Kwon, and T. Han, "Sparse kernel-based hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 943–947, Sep. 2012.
[38] T. Wang, B. Du, and L. Zhang, "A kernel-based target-constrained interference-minimized filter for hyperspectral sub-pixel target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 626–637, Apr. 2013.
[39] R. Zhao, B. Du, and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1227–1234, Apr. 2014.
[40] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.
[41] Y. Yang, Z. Ma, Y. Yang, F. Nie, and H. T. Shen, "Multitask spectral clustering by exploring intertask correlation," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1083–1094, May 2015.
[42] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005.
[43] S. P. Catterall, "Anomaly detection based on the statistics of hyperspectral imagery," in *Proc. 10th SPIE Conf. Imagery Spectrosc.*, Denver, CO, USA, 2004, pp. 171–178.
[44] Z. Guan, L. Zhang, J. Peng, and J. Fan, "Multi-view concept learning for data representation," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3016–3028, Nov. 2015.
[45] Z. Guan, J. Wu, Q. Zhang, A. Singh, and X. Yan, "Assessing and ranking structural correlations in graphs," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Athens, Greece, 2011, pp. 937–948.
[46] X. Lu, Y. Yuan, and P. Yan, "Alternatively constrained dictionary learning for image superresolution," *IEEE Trans. Cybern.*, vol. 44, no. 3, pp. 366–377, Mar. 2014.
[47] H. Qiao, P. Zhang, D. Wang, and B. Zhang, "An explicit nonlinear mapping for manifold learning," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 51–63, Feb. 2013.
[48] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
[49] L. Ma, M. Crawford, and J. Tian, "Anomaly detection for hyperspectral images based on robust locally linear embedding," *J. Infrared Millim. Terahertz Waves*, vol. 31, no. 6, pp. 753–762, 2010.
[50] A. K. Ziemann and D. W. Messinger, "Hyperspectral target detection using graph theory models and manifold geometry via an adaptive implementation of locally linear embedding," in *Proc. 20th SPIE Algorithms Technol. Multispect. Hyperspect. Ultraspect. Imagery*, Baltimore, MD, USA, 2014, Art. ID 90880B.
[51] B. Basener, E. J. Ientilucci, and D. W. Messinger, "Anomaly detection using topology," in *Proc. Def. Security Symp.*, Orlando, FL, USA, 2007, Art. ID 65650J.
[52] D. W. Messinger and J. Albano, "A graph theoretic approach to anomaly detection in hyperspectral imagery," in *Proc. IEEE Workshop Hyperspect. Image Signal Process. Evol. Remote Sens.*, Lisbon, Portugal, 2011, pp. 1–4.
[53] L. P. D. Munoz, D. W. Messinger, and A. K. Ziemann, "Target detection using the background model from the topological anomaly detection algorithm," in *Proc. 19th Algorithms Technol. Multispect. Hyperspect. Ultraspect. Imagery*, Baltimore, MA, USA, 2013, Art. ID 87430M.
[54] X. Liu, J. He, B. Lang, and S.-F. Chang, "Hash bit selection: A unified solution for selection problems in hashing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, 2013, pp. 1570–1577.
[55] M. Pavan and M. Pelillo, "A new graph-theoretic approach to clustering and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Madison, WI, USA, 2003, pp. 145–152.
[56] M. Pavan and M. Pelillo, "Dominant sets and pairwise clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 167–172, Jan. 2007.
[57] J.-M. Gaucel, M. Guillaume, and S. Bourennane, "Whitening spatial correlation filtering for hyperspectral anomaly detection," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 5. Philadelphia, PA, USA, 2013, pp. 333–336.
[58] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[59] L. K. Saul and S. T. Roweis. (2000). *An Introduction to Locally Linear Embedding*. [Online]. Available: http://www.cs.toronto.edu/~roweis/lle/publications.html
[60] A. Banerjee, P. Burlina, and R. Meth, "Fast hyperspectral anomaly detection via SVDD," in *Proc. IEEE Int. Conf. Image Process.*, vol. 4. San Antonio, TX, USA, 2007, pp. 101–104.

Yuan Yuan (M'05–SM'09) is currently a Full Professor with the Chinese Academy of Sciences, Beijing, China. She has authored or co-authored over 150 papers, including about 100 in reputable journals, such as the IEEE TRANSACTIONS and *Pattern Recognition*, as well as conference papers in CVPR, BMVC, ICIP, and ICASSP. Her current research interests include visual information processing and image/video content analysis.

Dandan Ma received the B.E. degree in electronic information science and technology from Xidian University, Xi'an, China, in 2013. She is currently pursuing the master's degree with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an. Her current research interests include computer vision and machine learning.

Qi Wang (M'15–SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently an Associate Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His current research interests include computer vision and pattern recognition.