

Fast Hyperspectral Anomaly Detection via High-Order 2-D Crossing Filter

Yuan Yuan, *Senior Member, IEEE*, Qi Wang, and Guokang Zhu

Abstract—Anomaly detection has been an important topic in hyperspectral image analysis. This technique is sometimes more preferable than the supervised target detection because it requires no *a priori* information for the interested materials. Many efforts have been made in this topic; however, they usually suffer from excessive time cost and a high false-positive rate. There are two major problems that lead to such a predicament. First, the construction of the background model and affinity estimation are often overcomplicated. Second, most of these methods have to impose a stringent assumption on the spectrum distribution of background; however, these assumptions cannot hold for all practical situations. Based on this consideration, this paper proposes a novel method allowing for fast yet accurate pixel-level hyperspectral anomaly detection. We claim the following main contributions: 1) construct a high-order 2-D crossing approach to find the regions of rapid change in the spectrum, which runs without any *a priori* assumption; and 2) design a low-complexity discrimination framework for fast hyperspectral anomaly detection, which can be implemented by a series of filtering operators with linear time cost. Experiments on three different hyperspectral images containing several pixel-level anomalies demonstrate the superiority of the proposed detector compared with the benchmark methods.

Index Terms—Anomaly detection, high order, hyperspectral image, remote sensing, 2-D crossing.

I. INTRODUCTION

HYPERSPECTRAL imaging systems have the ability to collect digital images with very densely sampled or nearly continuous radiance spectra for each pixel in the scene. The captured rich information about the spectral signatures can help identify minor differences between various materials. This characteristic enables hyperspectral images to be beneficial to a wide range of applications. For example, they have already been successfully applied to environmental monitoring [1], pro-

duction quality inspection [2], medical imaging [3], biological analysis [4], etc.

For the hyperspectral-image-based applications, target detection has been one of the most interesting and fundamental tasks [5]. Based on the requirements of *a priori* spectral information for the interested materials, existing works can be divided into two categories: supervised and unsupervised. The former works only when the objects of interest and the corresponding spectral signatures are given in advance. Typically, it detects targets by selecting all the pixels with spectral characteristics that are highly correlated to the referenced ones [6], [7]. However, from time to time, the interested objects and the accurate spectral calibrations are difficult to specify in advance. With regard to such a situation, unsupervised target detection, i.e., anomaly detection, is a more preferred and pertinent technique. Most anomaly detectors rely on no *a priori* information in addition to the fact that the anomalies are rare with respect to the background [8], [9].

Detecting anomalies in a hyperspectral image refers to distinguishing anomalous or prominent pixels, which are with distinct spectral signatures significantly deviating from their neighborhoods [10], [11]. To be specific, an anomaly can be considered as an outlier embodied by the background observations. Distinguishing these outliers is important in hyperspectral image analysis since they often represent unusual occurrences crucial for further investigation [8]. In general, traditional methods for anomaly detection involve knowledge extraction for background description and affinity function construction to measure the deviation of the examined data from the extracted knowledge. Although many efforts have been made to the two functional components, the progress is still far from satisfying. Existing methods usually suffer from excessive time cost and a high *false-positive rate* (FPR) [12]. This predicament is mainly due to the overcomplicated model construction and affinity estimation, as well as the stringent assumptions imposed on the spectrum distribution of the background.

Motivated by the need for overcoming the limitations of existing works, in this paper, we propose a high-order 2-D crossing-based method for fast yet accurate hyperspectral anomaly detection. The proposed method is termed *two-dimensional crossing-based anomaly detector* (2DCAD), which takes advantage of the following aspects.

- Decrease the FPR for hyperspectral anomaly detection without any *a priori* assumption. The proposed method is inspired by the excellence of zero-crossing analysis, which is carried out based on monitoring the instantaneous points at which the signal passes through a preset “zero”

Manuscript received July 26, 2013; revised December 26, 2013, March 11, 2014, and April 23, 2014; accepted May 7, 2014. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707104, by the National Natural Science Foundation of China under Grant 61172143, Grant 61105012, and Grant 61379094, and by the Fundamental Research Funds for the Central Universities under Grant 3102014JC02020G07.

Y. Yuan and G. Zhu are with the Center for OPTICAL IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: yuanyu@opt.ac.cn; zhuguokang@opt.ac.cn).

Q. Wang is with the Center for OPTICAL IMagery Analysis and Learning (OPTIMAL), 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/TGRS.2014.2326654

value [13]. It is a low-complexity approach without any *a priori* assumption on signature distribution but is originally dedicated to time-series processing. Therefore, in this paper, this approach is reformulated to be pertinent for multichannel spatial signal processing and now aims at seeking out regions of rapid change in the spectrum. Based on this formulation, a conceptually clear method is obtained for hyperspectral anomaly detection.

- Effectuate fast discrimination of anomalies without statistic model construction for the background, as well as the affinity estimation for the test pixel. In the proposed anomaly detector, different orders of 2-D crossing analysis are integrated to find the most anomalous pixels. This is achieved by employing a series of “center-surround” operators for difference generation and the *one-bit transform* (1BT) [14], [15] for 2-D crossing count calculation. All the involved operators are formulated within a filtering framework. This contributes to a conceptually simple yet practical detector and facilitates an efficient implementation with linear complexity.

The rest of this paper is organized as follows. Section II reviews the works on the topic of hyperspectral anomaly detection. Section III describes each component of 2DCAD in detail. Section IV presents the extensive experiments to prove the superiority of the proposed method in pixel-level hyperspectral anomaly detection, and the conclusion follows in Section V.

II. RELATED WORK

Generally, anomaly detectors tackle anomalies as outliers deviating from the knowledge extracted from their surroundings. According to the implicit assumptions imposed on background knowledge, efforts toward hyperspectral anomaly detection can be roughly divided into two categories. One category is based on the assumption that all the spectral signatures in the background are homogeneous in some aspects. In this view, the background is considered as one single type and described with a unified model, such as normal distribution, Gaussian mixture, and Student’s mixture. The other category assumes that the spectral signatures in the background are fitted with different class distributions. These distributions are modeled, respectively, and the deviation with respect to each one is calculated for further judgment.

Among the works that belong to the first category, which is probably the most popular, is the *Reed–Xiaoli* (RX) detector proposed by Reed and Yu [16]. RX is recognized to be the benchmark method in many multispectral/hyperspectral detection applications [10], [12], [17]. In this method, the nonstationary multivariate Gaussian model is assumed to characterize the local background pixels around the target. After estimating the mean vector and covariance matrix on the basis of the local neighborhood, the Mahalanobi distance between each examined pixel and the statistic model is calculated and compared with a threshold for the final discrimination.

Despite its popularity, RX is found to be with a high FPR in many applications [12], [17], [18]. There are two main

problems with RX that lead to its poor practicality. The first problem is that the local multivariate Gaussian model cannot always provide an adequate representation for background, particularly when there are multiple materials and textures [12], [18]. The other problem is that the affinity function of RX involves estimation and inversion of a high-dimensional covariance matrix, frequently under a small sample size [12], [19]. These operations are highly complex but badly conditioned and unstable. There are indeed many efforts made to overcome these two limitations, which provide a lot of variants for RX, such as selective KPCA RX [20], subspace RX [21], kernel RX [22], minimum covariance determinant RX [23], *random-selection-based anomaly detector* (RSAD) [10], and compressive RX [8]. However, as the basic assumption has never been revised, to the best of our knowledge, there is still no refreshing progress in practice.

Recently, many other methods assuming the homogeneous background have announced promising results with better FPR than RX series. These methods differ from RX series essentially in the description of background spectra. For example, Banerjee *et al.* [12] introduce a concept of *support vector data description* (SVDD) for hyperspectral anomaly detection. SVDD is developed by Tax and Duin in [24] and Tax *et al.* in [25], which models a distribution by learning the support as the minimal enclosing hypersphere enveloping the data in specified feature space. Benefiting from this technique, the *a priori* distribution that RX depends on could be removed, and an anomaly could be intuitively tackled as an outlier excluded by the support. There also exist many variants of the SVDD detector, such as parameter-adaptive kernel SVDD [18], Gaussian kernel SVDD [19], and sparse kernel SVDD [26]. However, as SVDD is not an expeditious technique, the SVDD-based anomaly detectors are computationally too expensive even for a medium-sized hyperspectral image.

In addition to those previously mentioned, kurtosis is considered as a useful high-order moment to characterize a distribution and has also been employed for anomaly detection. Du and Kopriva [11] report that the small targets contribute to the tail of the distribution in a hyperspectral image. They propose to unequally constrain the kurtosis and project the spectral signatures with a set of projectors yielding maximum constrained kurtoses. Then, the anomalies are successively discriminated in the gradually projected feature space. This method is considered to be quick but is found to fail when the background cannot be perfectly characterized as a Gaussian distribution. Correlation analysis is another strategy that can support fast outlier identification. Inspired by this strategy, Gaucel *et al.* [27] develop an anomaly detector based on *whitening and spatial correlation filtering* (WSCF). They start with the whitening transformation for the hyperspectral pixels and proceed to filtering the whitened image to obtain the statistics of the correlations between each test pixel and its surrounding neighbors. Finally, a thresholding strategy is employed for the discrimination. Although this method is indeed very efficient, it lacks effectiveness and robustness in practice. More recently, *Parzen windowing* (PW) has also been investigated in hyperspectral anomaly detection. PW performs *probability density function* estimation in a window through a kernel function,

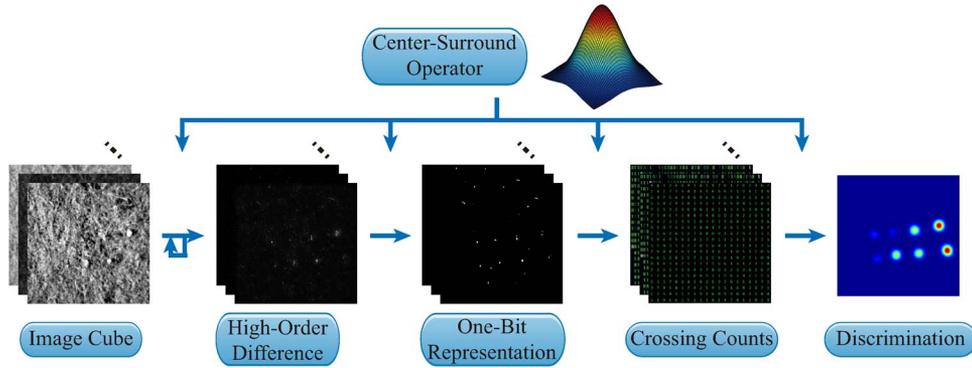


Fig. 1. Flowchart of the proposed high-order 2DCAD.

without a specific assumption of the background distribution. Based on such an approach, in the works of Veracini *et al.* [28] and Matteoli *et al.* [17], the anomalies within a scene are directly detected through the likelihood ratio test (LRT) by a threshold procedure. Unfortunately, suffering from complex bandwidth selection, the PW-based anomaly detectors are computationally intensive as well.

Different from the aforementioned methods, the second category includes methods that assume the multiclass background. These methods estimate for each background class a description and then calculate and integrate the deviations of all the pixels with respect to these descriptions for the final discrimination. As for this category, one of the most typical samples is the *multivariate normal inverse Gaussian* (MNIG) detector [29]. In the MNIG detector, each background class is assumed to be an independent MNIG distribution. After estimating all the distributions on the basis of the classes, the LRT rule is employed to judge whether a pixel is an anomaly. Another example is the *cluster-based anomaly detector* (CBAD) [30]. This method is performed after the entire scene has been segmented into different clusters, and the anomalies are detected in each cluster individually. Obviously, methods of the second category specifically require *a priori* information about the background classes and heavily depend on another technique to make reliable classification. Therefore, these methods may be limited more stringent than those belonging to the first category.

III. 2DCAD

This section details the proposed 2DCAD for hyperspectral anomaly detection. The overall flowchart is summarized in Fig. 1. In this method, the most crucial component is the low-complexity-based high-order 2-D crossing analysis for high-dimensional hyperspectral images, without any *a priori* assumption on the signal distribution. There are mainly four steps for this procedure. First, the high-order difference signals are constructed from the input hyperspectral image cube. A subsequent 1-bit representation is then applied to generate a series of effective bit maps of different orders. After that, the 2-D crossing count is calculated for each bit map, which can be considered as a statistics of anomaly or not. Finally, a higher-level difference operator is employed to get a more

concise expression, and the discrimination function is designed to make the final decision. All these operations are efficiently implemented using a spatial weighted center-surround filtering framework. The following will describe the theory behind high-order 2-D crossing analysis and the corresponding discrimination function for hyperspectral anomaly detection in detail.

A. High-Order 2-D Crossing Analysis

Hyperspectral anomaly detection aims at distinguishing the pixels with the anomalous spectral signatures significantly deviating from their surrounding neighborhoods. This problem can be converted to locating the points where there is an inconsistency between two consecutive pixels, if a clip operator can be formulated here to dichotomize the signals. A similar idea has a long history in time-series processing referring to zero-crossing analysis, for which the clip operation could be directly a hard segmentation using a preset “zero” value. However, the same strategy will fail in hyperspectral anomaly detection, because it is difficult to define the “zero” value for the raw multichannel spatial signals. To cope with this problem, we formulate a high-order 2-D crossing analysis on the basis of the high-order difference operation specific for hyperspectral imagery.

1) *High-Order Difference Construction*: Assume a hyperspectral image composed of pixels $\mathbf{H} = \{\mathbf{h}_i\}$. The first-order difference $\mathbf{L}_i^{(1)}$ is constructed through a spatial weighted center-surround operator, i.e.,

$$\mathbf{L}_i^{(1)} = \sum_{j \in \mathbb{N}_i} \|\mathbf{h}_i - \mathbf{h}_j\|^2 w_{ij} \quad (1)$$

$$w_{ij} = \frac{1}{z_i} \exp\left(-\frac{1}{2\sigma^2} \|p_i - p_j\|^2\right). \quad (2)$$

Here, \mathbb{N}_i denotes N local neighbors around the location p_i . w_{ij} is a spatial weighting term that controls the contributions of the neighboring pixel at location p_j over the influence radius. $\sigma = \log(N)$ is employed to control the strength of w_{ij} , and z_i is a normalization factor ensuring $\sum_{j \in \mathbb{N}_i} w_{ij} = 1$.

In practice, the spatial weighting term is chosen as a Gaussian blurring kernel, which can be evaluated very efficiently [31]. Specifically, for a fast implementation of this

center-surround operation, (1) is first decomposed by factoring out the quadratic error function, i.e.,

$$\mathbf{L}_i^{(1)} = \mathbf{h}_i^2 - 2\mathbf{h}_i \sum_{j \in \mathbb{N}_i} \mathbf{h}_j^T w_{ij} + \sum_{j \in \mathbb{N}_i} \mathbf{h}_j^2 w_{ij}. \quad (3)$$

Then, a permutohedral lattice embedding approach [32], which can yield a linear time approximation for high-dimensional filtering, is employed to evaluate $\sum_{j \in \mathbb{N}_i} \mathbf{h}_j^T w_{ij}$ and $\sum_{j \in \mathbb{N}_i} \mathbf{h}_j^2 w_{ij}$. Permutohedral lattice embedding exploits the band-limiting effects of Gaussian smoothing and performs the fast evaluation by extracting a sparse number of samples on simplices of a high-dimensional permutohedral lattice structure. By using this approach, a first-order difference could be constructed in linear time $O(N)$.

Once the first-order differences have been obtained, there is a direct way to construct the higher k th-order differences, in a recursive manner. More specifically, there is recursively a weighted center-surround operator implemented as

$$\mathbf{L}_i^{(k)} = \sum_{j \in \mathbb{N}_i} \left\| \mathbf{L}_i^{(k-1)} - \mathbf{L}_j^{(k-1)} \right\|^2 w_{ij}. \quad (4)$$

It is manifest that, by employing the same strategy as (1), each higher-order difference can be evaluated in $O(N)$ time.

2) *High-Order 2-D Crossing Count*: After obtaining the difference signals, the high-order 2-D crossing method also requires a *clip operator* to build the binary signals, as well as a *change monitor* utilized to locate the points where there are rapid changes compared with their neighbors.

In regard of the *clip operator*, the proposed method is implemented by making use of single-bit-depth representation. This type of representation has been generally successful in discriminating patterns with different values [33]. In this paper, we use a 1BT filter to transform the difference signals into either 0 or 1. The bit map of the k th-order differences is accomplished in the form of

$$\mathbf{B}_i^{(k)} = \begin{cases} 1, & \text{if } \mathbf{L}_i^{(k)} \geq F(\mathbf{L}_i^{(k)}) \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Here, $F(\mathbf{L}_i^{(k)})$ denotes the filtered result of $\mathbf{L}_i^{(k)}$ obtained by a multibandpass filter. This kind of 1BT has achieved great success in both visible video [14], [33] and hyperspectral image [15] processing. Similar to the works of Natarajan *et al.* in [14] and Demir *et al.* [15], a 17×17 multibandpass-filter kernel is employed, which is defined as

$$f(i, j) = \begin{cases} 1/25, & \text{if } i, j \in [0, 4, 8, 12, 16] \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

After applying the 1BT filter, the original hyperspectral image becomes a binary map. Then, as mentioned shortly before, there is a need for the *change monitor* to calculate the high-order crossing counts. Different from the case of time-series analysis where there is only one direction forward time relation, for the examined pixel in the 2-D 1BT map, the *change monitor* has to count all the transitions (i.e., changes from 1 to 0, and *vice versa*) of the pixel from its neighbors in a surrounded 2-D

region with the consideration of their spatial distribution. Therefore, in this paper, the *change monitor* is formulated by

$$\mathbf{D}_i^{(k)} = \sum_{j \in \mathbb{N}_i} \left(\mathbf{B}_i^{(k)} \oplus \mathbf{B}_j^{(k)} \right) w_{ij} \quad (7)$$

where $\mathbf{D}_i^{(k)}$ denotes the k th-order crossing counts, i is the examined position, j is the spatial index in the center-surround neighborhood structure, and \oplus denotes the Boolean exclusive-or (XOR) operation. The spatial weighting term w_{ij} is imposed to emphasize the contributions of neighbors closer to position i . Beneficial from the Boolean simplicity, (7) also has very low complexity and is thus particularly suitable for fast computation.

B. High-Order 2-D Crossing for Anomaly Detection

As for the ultimate purpose, performing the high-order 2-D crossing analysis is dedicated to distinguishing whether a test position presents an anomalous signal different from its neighbors. It is found that different orders of 2-D crossing counts unequally contribute to the discrimination of anomaly. This phenomenon is illustrated in Fig. 2(a). We can see clearly that not all the statistics of crossing counts is suitable for separating the two kinds of points. Sometimes, their distributions are mixed together. Inspired by an excellent work of magnetic signal processing [34], we further employ a difference operator on the calculated high-order crossing counts to partition the two distributions, as well as to capture the internal relations among different orders. Then, the final discrimination function is consequently formulated on the basis of the 2-D crossing count differences.

1) *Difference Operator for High-Order 2-D Crossing Counts*: Given the high-order 2-D crossing counts, the employed difference operator is defined by

$$\Delta_i^{(k)} = \begin{cases} \mathbf{D}_i^{(k)}, & k = 1 \\ \left| \mathbf{D}_i^{(k)} - \mathbf{D}_i^{(k-1)} \right|, & \text{otherwise.} \end{cases} \quad (8)$$

This operator is concise and intuitive in implementation and, at the same time, can take into account the relationships between two adjacent orders. It is found that $\Delta_i^{(k)}$ is more effective than $\mathbf{D}_i^{(k)}$. Fig. 2(b) presents the statistics of $\Delta_i^{(k)}$. Compared with Fig. 2(a), it is obvious that $\Delta_i^{(k)}$ is more discriminative for the two kinds of points. This conclusion is consistent with that in [34], which indicates that an effective detector can be constructed on the basis of $\Delta_i^{(k)}$.

2) *Outline of Anomaly Detection Procedure*: After the given procedure, the final decision will be made for the anomaly or not. Similar to [12] and [16], 2DCAD uses the hollow-window information to design the discrimination function. The sliding window consists of two regions. As shown in Fig. 3, there are both inner and outer windows surrounding the test pixel. The inner area serves as a guard band, whereas the outer area constitutes the background for the test pixel. Note that the dimensions s_{in} and s_{out} of the sliding hollow window are selected according to the spatial resolution of the scene and the expected size of the anomaly targets. Specifically, the inner area

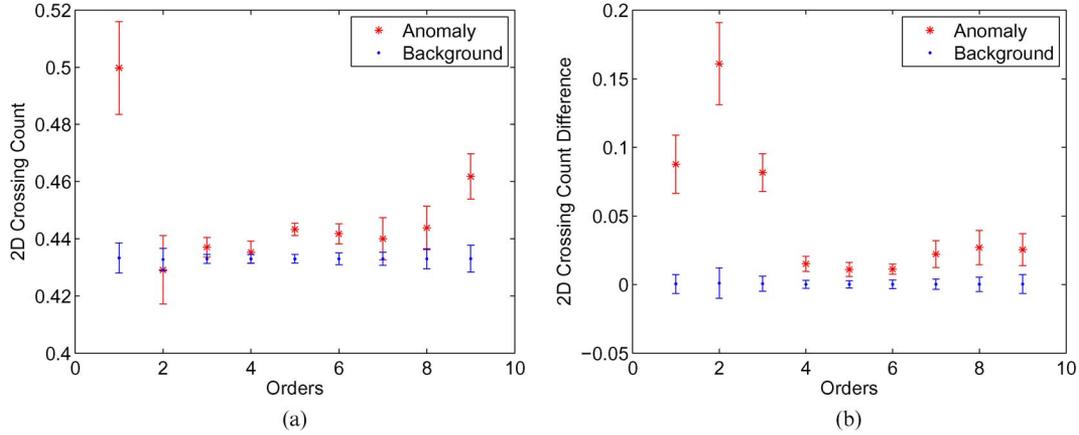


Fig. 2. Mean values and standard deviations for (a) normalized 2-D crossing counts and (b) normalized 2-D crossing differences. The statistics are calculated in a real-world hyperspectral image composed of 180×180 pixels with 88 anomalies.

should accommodate the expected typical or largest anomaly targets in the scene, and the outer region should be large enough for background estimation while preventing excessive contamination by anomalies. The discrimination function is formulated as

$$\varphi = \sum_{k=1}^K \frac{\left\| \Delta_{\text{inner}}^{(k)} - \Delta_{\text{outer}}^{(k)} \right\|^2}{\Delta_{\text{outer}}^{(k)}} \quad (9)$$

where there are K orders of 2-D crossing count difference taken into the discrimination. $\Delta_{\text{inner}}^{(k)}$ and $\Delta_{\text{outer}}^{(k)}$ denote the averaged value of the k th-order 2-D crossing count difference over the inner and outer regions, respectively.

For better understanding, the overall procedure is briefly summarized in Algorithm 1. For each examined pixel, the proposed detector takes $O(KN)$ time for each of the first three steps as well as the final discrimination and $O(K)$ time for the fourth step. Suppose there are a total of I pixels in a hyperspectral image, 2DCAD runs in complexity $O(I \times (4KN + K))$ involving only basic arithmetic. As k and N are typically not very large in practice, the proposed method is very efficient.

Algorithm 1 2DCAD for Hyperspectral Anomaly Detection

Input:

The observed hyperspectral data, \mathbf{H} ;
 The maximum order, K ;
 The sizes of the inner and outer regions, s_{in} and s_{out} .

Processing Steps:

- 1: Recursively generate $\mathbf{L}^{(k)}$ by (3) and (4);
- 2: Conduct the clip operator to obtain $\mathbf{B}^{(k)}$ by (5);
- 3: Employ the change monitor to calculate $\mathbf{D}^{(k)}$ by (7);
- 4: Use the difference operator to generate $\Delta^{(k)}$ by (8);
- 5: Make discrimination φ by (9).

Output:

Discrimination matrix, φ .

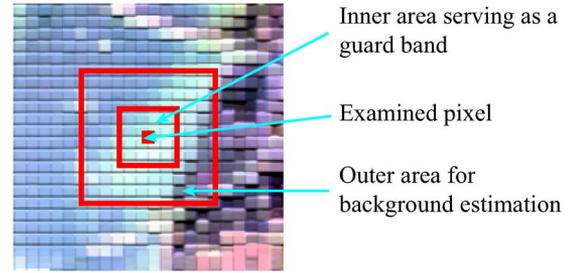


Fig. 3. Example of a sliding hollow window.

IV. EXPERIMENTS

The following section starts with the introduction of the employed hyperspectral data sets and then describes the experimental setup concerning the evaluation metrics and the benchmark detectors. In the end, the comparison results are presented and analyzed in detail.

A. Data Sets

To evaluate the performance of the proposed 2DCAD under different types of hyperspectral images, a simulated data set and two publicly available real-world data sets gathered by famous remote imaging systems are employed. The details of these hyperspectral data sets are introduced as follows.

- 1) *Simulated scene*. This data set was synthesized with 200×200 pixels comprising the spectra of lawn grass, dry long grass, blackbrush leaf, sage brush, and tumblewe. The corresponding spectral signatures are selected from the U.S. Geological Survey vegetation spectral library.¹ All these spectra are previously calibrated to the bands consistent with AVIRIS data and are down-sampled with a dimension of 105. Similar to [10], the synthetic scene has two background regions. The top 100 lines consist of lawn grass and dry long grass, with the mixture proportion of lawn grass regularly varying from 100% to 50.5%. For the remaining 100 lines, the background is composed of blackbrush leaf and sage brush, and the percentage of blackbrush leaf varies from

¹<http://speclab.cr.usgs.gov/spectral-lib.html>.

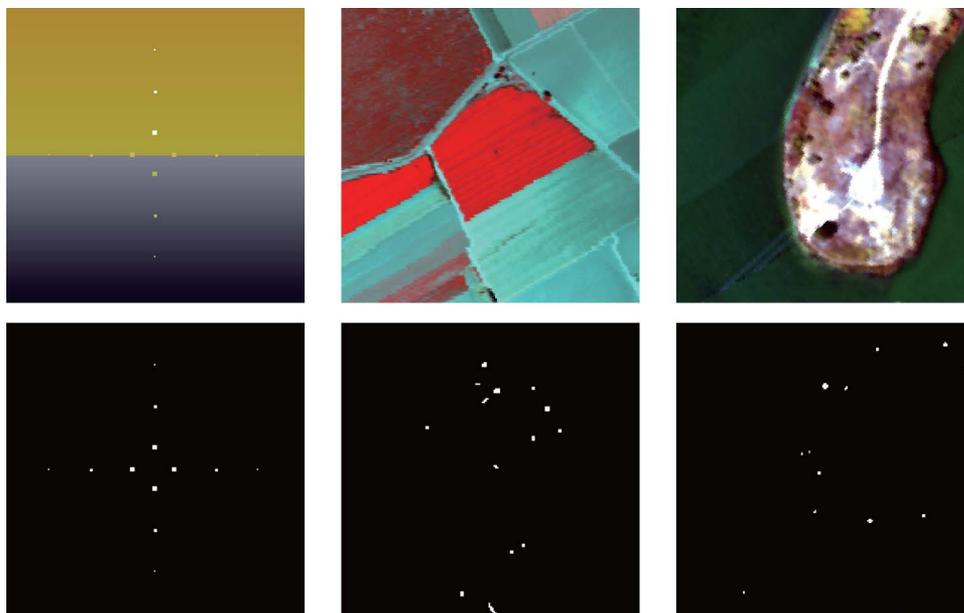


Fig. 4. Visualization of the scenes and the ground truth masks. The first row shows the color representations of the *Simulated scene*, the *Salinas scene*, and the *Lakeshore scene*, respectively. The second row represents the corresponding ground truths.

50.5% to 100% with 0.25% increase per line. Finally, three spectra, i.e., sage brush, dry long grass, and tumblewee, are separately used as anomalies. The corresponding anomaly targets are constructed with three different sizes (i.e., 1×1 , 2×2 , and 3×3) and are sparsely placed at the top background, the bottom background, and the border between the two regions, respectively. Please refer to Fig. 4 for the specific locations.

- 2) *Salinas scene*. This scene was gathered by the AVIRIS system with 224 bands in 1998 over Salinas Valley, CA. The original data set consists of 512×217 pixels with a spatial resolution of 3.7 m and comprises vegetables, bare soils, and vineyard fields. Conventional experiments usually exclude 20 atmospheric and water bands (108–112, 154–167, and 224) due to the low signal-to-noise ratio (SNR). In this paper, the employed *Salinas scene* is a subset with 180×180 pixels cropped from the one downloaded from the GIC website.²
- 3) *Lakeshore scene*. This data set was acquired by the SAMSON sensor, which is a push broom sensitive to the near-infrared bands, and was collected by the Florida Environmental Research Institute as part of the GOES-R-sponsored experiment. The SAMSON sensor is developed by FERI and generates 156 bands covering the spectral range from 400 to 900 nm with a bandwidth of 3.2 nm. The collected scene consists of 952×952 pixels with a geometric resolution of 1.0 m. In this paper, the employed *Lakeshore scene* is a subset of the one provided by the Opticks project,³ and is composed of 200×200 pixels with atmospheric correction.

B. Experimental Setup

Preparatory to the experimental comparison, there are three key points that need to be stated: 1) ground truth construction of the data sets; 2) evaluation metrics; and 3) benchmark methods.

Since the first image is manually synthesized, its ground truth labels of anomalies are precisely known. As for the second and third data sets, their ground truths are not provided. Some researchers claim a ground truth labeled by themselves, but how the procedure is conducted is not introduced. To establish our ground truths as objective as possible, this work turns to hyperspectral classification techniques. To be specific, for a hyperspectral image with known category information of the background objects, a classification training set is constructed by selecting some adequate pixels for each background category and is further refined through the constraint of random sample consensus [35] to restrain the contamination by the possible anomalies. After performing the classification by *support vector machine*, *classification and regression trees*, *k-nearest neighbors*, and *Gaussian mixture model*, the pixels that have rare labels with respect to their surrounding neighbors are selected as candidates. Then, the true anomaly pixels are specified as those rare in more than half of the classification processes. The finalized ground truths of the employed data sets are shown in Fig. 4. The pixel sizes of the anomaly targets contained in these two real-world data sets are approximately in the range of 1×1 to 3×3 .

The second question is how to evaluate the anomaly detectors. In the related literatures, the detectors are typically evaluated by plotting their receiver operating characteristic (ROC) curves referring to ground truth masks [9], [10], [17]. This curve is sketched by varying the threshold used to dichotomize the detection results, according to which a detected pixel could be judged as true anomaly or false alarm. Based on the ROC curves, a derivative indicator, i.e., the area under

²<http://www.ehu.es/ccwintco/>.

³<http://opticks.org/confluence/display/opticks/>.

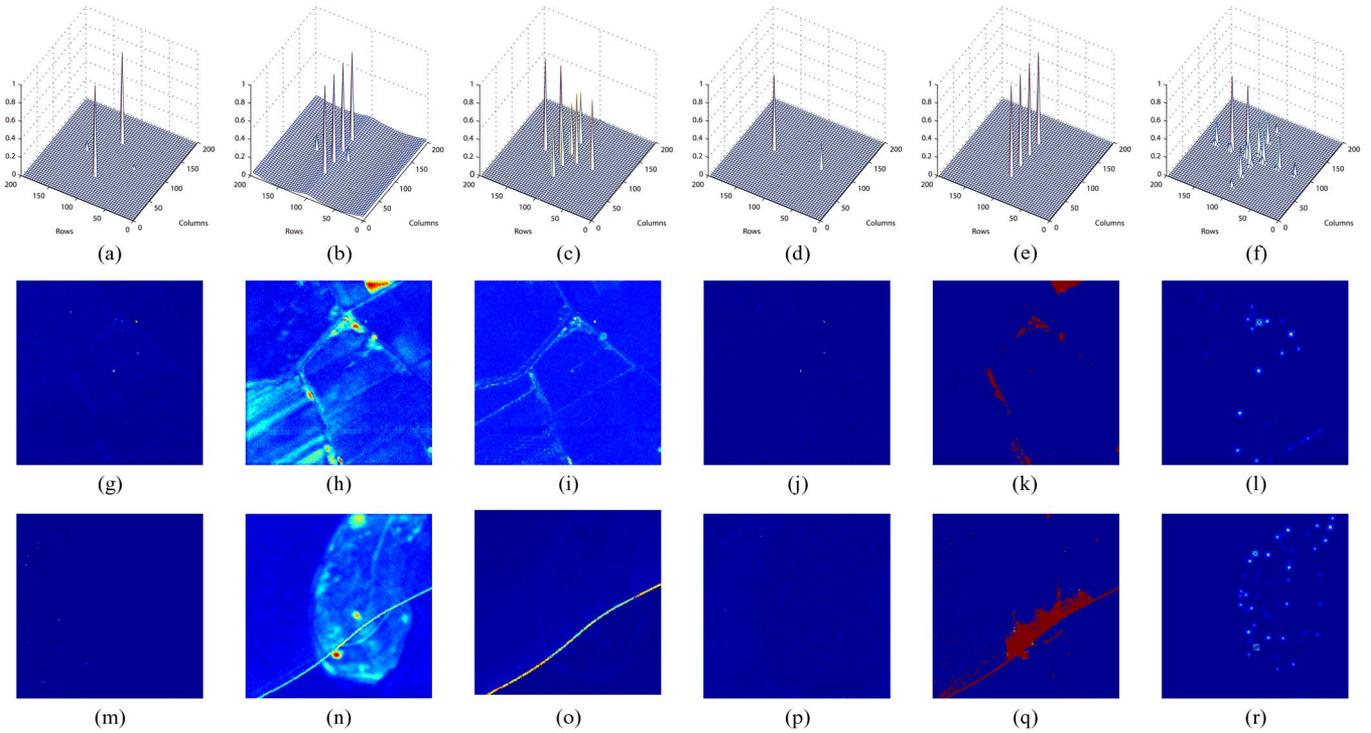


Fig. 5. Visualization of the anomaly detection results. From top to bottom, the first row represents the comparison of the 3-D illustrations on the *Simulated scene*, and the remaining rows respectively illustrate the comparisons of the normalized detection maps on the *Salinas scene* and the *Lakeshore scene*. From left to right, the columns respectively represent the detection results of RX, WSCF, CBAD, SVDD, RSAD, and 2DCAD.

the curve (AUC), is calculated for further evaluation [12], [19]. Additionally, the abilities of the detectors to intuitively separate anomaly targets from the background have also aroused interest in practice. Toward this direction, this paper introduces a criterion measuring the Bhattacharyya distance between the detection score histograms of the true anomaly pixels and the background pixels. Such evaluation metric is denoted with BD_{hist} and calculated by

$$BD_{\text{hist}}(H_1, H_2) = \sqrt{1 - \frac{\sum_i \sqrt{H_1(i)H_2(i)}}{\sum_i H_1(i) \sum_i H_2(i)}} \quad (10)$$

where H_1 and H_2 are two histograms, with $H_1(i)$ and $H_2(i)$ denoting their i th components. Larger BD_{hist} values indicate better performances.

In regard to the benchmark methods, there is a consensus that the employed methods should guarantee popularity, recency, and variety. In our experiments, the performance of the proposed method is verified by comparing it with RX, SVDD, CBAD, WSCF, and RSAD. The first two benchmark methods are cited with the highest frequency in anomaly detection literatures and lead to a lot of variants. CBAD is a modified version of RX, which uses RX in a global sense through vector-quantization-based clustering. WSCF is also a fast anomaly detector based on 2-D spatial filtering. RSAD has been developed very recently and can enrich the variety of competitors. Since there are four methods (RX, WSCF, SVDD, and 2DCAD) involving a sliding hollow window, the window size should be identified. To adapt to the employed hyperspectral data sets, the

dimensions of the inner and outer regions are fixed to 3×3 and 7×7 , respectively. The quantization level of CBAD is set to 4 for the *Simulated scene*, as there are four spectral signatures used to form the backgrounds, and to 8 for the two real-world data sets according to [30]. The kernel parameter for SVDD is determined by an approximate minimax technique formulated in [12]. As for 2DCAD, the order parameter K is set to 2 in the comparison between 2DCAD and the competitors. All the detection results are presented by the anomaly maps with the range $[0, 1]$, in which the high value of a pixel indicates the high possibility of being an anomaly.

C. Comparison Results

The comparison experiment is first conducted on the *Simulated scene*. Fig. 5(a)–(f) and Fig. 6(a) and (d) illustrate the detection results on this data set. In Fig. 5(a)–(f), it is found that the 3-D displayed detection results of 2DCAD greatly differs from those of the benchmark detectors. Verified by the ground truth mask, RX and RSAD are successful in finding the anomaly targets composed of spectral signatures greatly differing from all the others. These anomalies are located on the border between the two background regions. However, as for the anomalies that have relatively slight distinctiveness and are embedded in the mixed top and bottom backgrounds, these two methods do not perform well. SVDD and WSCF have the accuracy nearly equivalent to RX, but the detection results are opposite. These two detectors have the ability to detect the true anomalies in the strongly homogenous backgrounds but are ineffective even for the high discriminative anomalies when

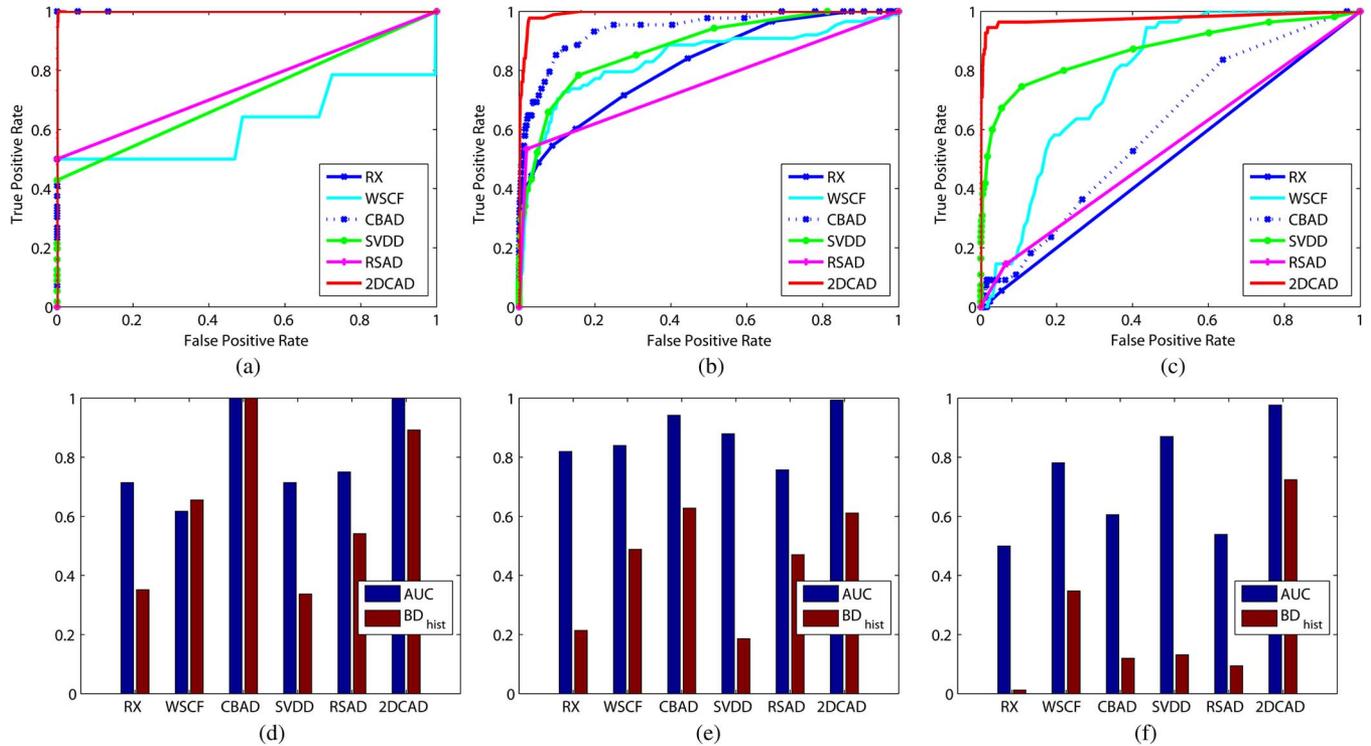


Fig. 6. Quantitative comparison of the anomaly detection results. The first row shows the ROC curves, and the second row illustrates the AUC and BD_{hist} bars. From left to right, the columns respectively correspond to the comparisons on the *Simulated scene*, the *Salinas scene*, and the *Lakeshore scene*.

the background is relatively complex. Compared with these competitors, the proposed detector can highlight more anomaly pixels, without losing the ability to suppress the background. As shown in Fig. 6(a) and (d), 2DCAD demonstrates the significantly better ROC curve, as well as the AUC and BD_{hist} bars, than RX, WSCF, SVDD, and RSAD. It should be noted that CBAD also performs well on the *Simulated scene*. Therefore, two more difficult tasks have been conducted on the employed real-world data sets for further comparison.

The detection results for the *Salinas scene* are illustrated in Fig. 5(g)–(l) and Fig. 6(b) and (e), where Fig. 5(g)–(l) presents the normalized anomaly maps produced by RX, WSCF, CBAD, SVDD, RSAD, and the proposed 2DCAD, respectively. As observed from these figures, the global method RSAD tends to assign high anomalous values to the pixels that have the low proportions of occurrence in this scene. This judgement is too arbitrary and inaccurate. Among the other three detectors, RX and SVDD yield visually similar results. They both have many omissions at which the spectral signatures are distinguishable but embedded in the relatively highly divergent backgrounds. WSCF and CBAD mainly highlight the boundaries and fail in suppressing the background. Differently, the performance of the proposed 2DCAD is satisfactory on this scene. As can be seen in Fig. 5(l), the anomaly map of 2DCAD effectively assigns the high values for most of the anomalies while suppressing the background. Therefore, for all the quantitative results plotted in Fig. 6(b) and (e), the proposed detector clearly dominates the other competitors.

In regard to *Lakeshore scene*, the comparison results are shown in Fig. 5(m)–(r) and Fig. 6(c) and (f). For this hyperspectral image, as proved in Fig. 6(c) and (f), the pro-

TABLE I
COMPARISON OF RUNNING TIMES (s)

	Simulated scene	Salinas scene	Lakeshore scene
RX	110.27	366.47	246.01
WSCF	3.65	3.74	3.15
CBAD	1.43	3.81	3.40
SVDD	2512.90	2330.31	2925.96
RSAD	769.29	132.15	236.08
2DCAD	3.96	3.37	4.00

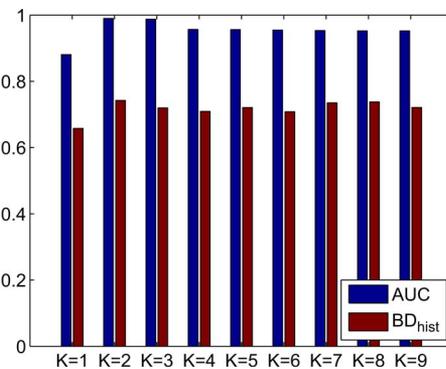


Fig. 7. Comparison between the averaged AUC and BD_{hist} bars under different settings of K .

posed detector also takes a significant advantage on the two employed evaluation metrics compared with the competitors. A careful observation on the produced anomaly maps reveals that RX mainly selects the pixels from the calm lake region

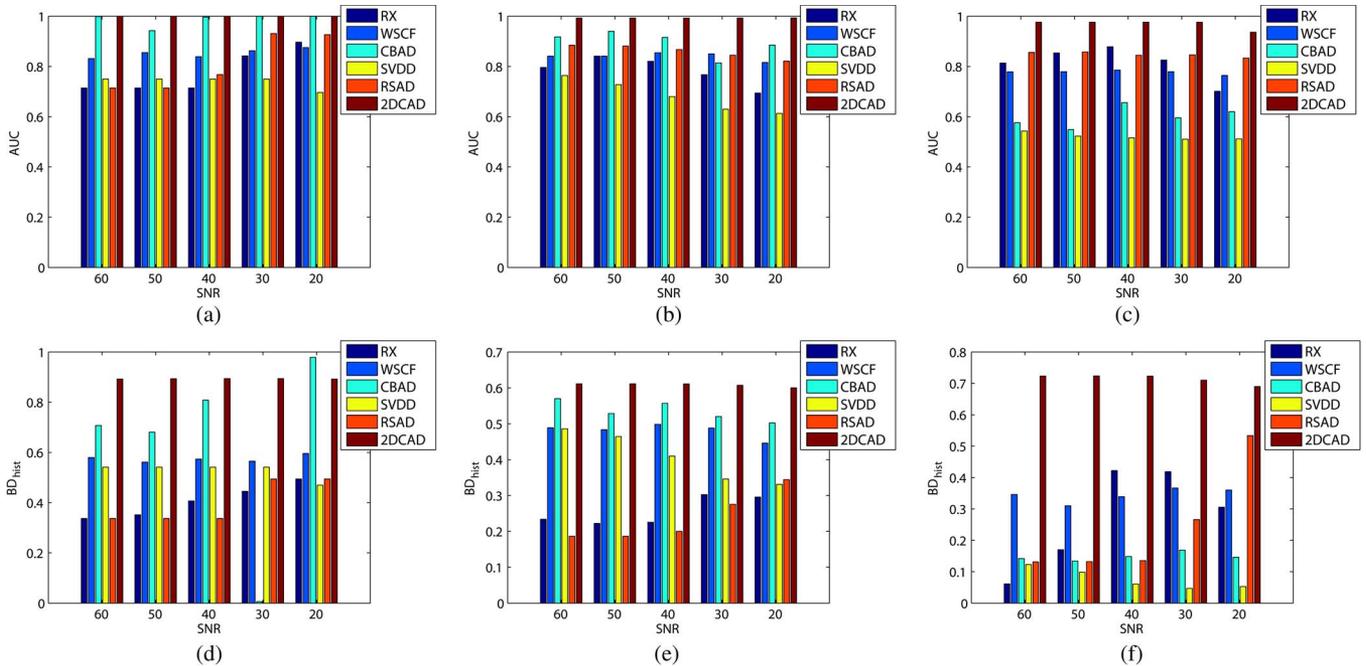


Fig. 8. Comparison of the performances on the noisy images. From left to right, the columns respectively represent the comparisons on the *Simulated scene*, the *Salinas scene*, and the *Lakeshore scene*.

TABLE II
AMOUNTS OF RELATIVE CHANGES ON THE PERFORMANCES BEFORE AND AFTER ADDING NOISE

	60dB		50dB		40dB		30dB		20dB	
	AUC	BD _{hist}								
RX	0.1126	0.0281	0.1253	0.0555	0.1266	0.1590	0.1688	0.1961	0.1699	0.1725
WSCF	0.0720	0.2370	0.0638	0.2330	0.0658	0.2336	0.0584	0.2491	0.0705	0.2507
CBAD	0.0183	0.1243	0.0387	0.1445	0.0267	0.0974	0.0458	0.3836	0.0239	0.0578
SVDD	0.0037	0.0148	0.0152	0.0033	0.0333	0.0312	0.0519	0.0572	0.0749	0.0839
RSAD	0.0065	0.0000	0.0049	0.0003	0.0302	0.0060	0.0916	0.1271	0.1023	0.2390
2DCAD	0.0000	0.0000	0.0000	0.0005	0.0000	0.0008	0.0002	0.0065	0.0134	0.0149

as anomalies, while losing awareness to the truly anomalous ones embedded in the complex lands. SVDD produces a result better than RX; however, the overall detection statistics for the entire anomaly positions are not conspicuous enough within the background regions. RSAD performs consistent on the *Salinas scene*. Its detection result lacks interpretability for the anomalies. WSCF and CBAD are seriously interfered with by the interface between two remote scanning series. On the contrary, the proposed 2DCAD is more effective for the employed *Lakeshore scene*.

D. Comparison of Running Times

To evaluate the efficiency of the implemented detectors, the running times of each method are compared in Table I. Timings have been taken on an Intel Core i3-550 3.2-GHz CPU with 3-GB RAM. All these detectors are implemented in the MATLAB platform. Assuming there are a total of N pixels contained in the outer region of the sliding hollow window and K orders that are taken into account, for each examined pixel, the proposed detector needs $O(4KN + N)$ computation time to assign a detection degree.

As a result, for all the employed simulated and real-world hyperspectral images, 2DCAD costs less time of two or three orders of magnitude than most of the competitive algorithms, except for the nearly equivalent running time compared with WSCF and CBAD, while guaranteeing the superior anomaly maps. As for WSCF and CBAD, it is found that there is always an unsatisfying upper bound on the performance in practice. For example, WSCF is limited by $AUC = 0.81$ on the employed *Lakeshore scene*. In such a case, claiming the true-positive rate at 90.91% entails the false-alarm rate to exceed 41%. Compared with WSCF, the proposed method achieves $AUC = 0.93$ on the same scene, with which a high true-positive rate of 92.73% only sacrifices the false-alarm rate to be 1.28%. Therefore, from an overall perspective, the proposed 2DCAD can claim to be most predominant in practical applications.

E. Parameter Discussion

As presented in Section IV-C, there is a tremendous advancement in both effectiveness and efficiency when taking the proposed 2DCAD for hyperspectral anomaly detection. However, there is still an uncertainty concerning the selection

of parameter K , which is worthy of further explanation. This section answers this question by taking an analysis of the performances of 2DCAD under different settings of K . Without loss of generality, in this analysis, the averaged AUC and BD_{hist} value over all the employed data sets are used as the validation metrics.

The results are illustrated in Fig. 7. It can be observed that employing only the first-order information is poor at discrimination ability. With the increasing of K , which means the effect of the higher-order information is incorporated in the detection procedure, the performance will improve. After K reaches 2, setting higher K cannot get a larger AUC value, or even leads to a deterioration, while the computational cost is more expensive. These results indicate that the high-order 2-D crossing is effective for hyperspectral anomaly detection, but it is necessary to make an appropriate selection of K . Considering the two employed evaluation metrics simultaneously, K is finally fixed as 2 in our practical implementations.

F. Robustness to Noise

Generally, the actual requirements in the practical conditions not only involve the satisfying performances on efficiency and effectiveness but also consider the robustness to significant levels of noise. To evaluate the robustness of all the implemented anomaly detectors, each employed hyperspectral image is extended to five copies containing certain levels of spectral correlated noise by using the strategy in [36]. For these copies, the SNR ranges from 60 to 20 dB with an interval of 10 dB, whereas the shape parameter η is fixed to 0.18 ($\eta \rightarrow \infty$ leads to white noise; $\eta \rightarrow 0$ generates one-band noise).

The performances of the implemented anomaly detectors on the noisy images are illustrated in Fig. 8. It is clear that the superiority of 2DCAD can hold for significant levels of noise. Moreover, the amounts of relative changes on the performances before and after adding noise are also compared. The comparison is shown in Table II, by which the averaged relative changes over the three employed data sets are presented. As can be seen from the table, the relative changes of 2DCAD on both AUC and BD_{hist} are very slight (less than 0.01) and smaller than those of the competitors in most of the cases. All these clues altogether demonstrate that the proposed 2DCAD has remarkable robustness to the levels of spectral correlated noise compared with the benchmark methods.

V. CONCLUSION

Anomalies in the hyperspectral image often represent crucial occurrences worthy of further investigation. Therefore, reliably detecting these anomalies is important in both academia and industry. In this regard, traditional hyperspectral anomaly detectors are far from satisfying due to their excessive time costs and high FPRs. In this paper, a novel hyperspectral anomaly detector has been proposed based on the formulation of the high-order 2-D crossing analysis. The proposed detector is termed 2DCAD, which can allow for fast examination of the testing pixels with respect to their neighborhoods, without losing accuracy.

2DCAD is fully unsupervised and computationally acceptable, the benefits of which are as follows: 1) Formulation of the high-order 2-D crossing approach dedicates to multichannel 2-D signal processing. The proposed high-order 2-D crossing approach can seek regions of rapid change in the spectrum, which can help decrease the FPR for hyperspectral anomaly detection without any *a priori* assumption regarding the background. 2) Low-complexity discrimination allows for a unitive filtering framework. This facilitates a conceptually simple and efficient implementation conducted with linear time cost.

The experimental results demonstrate that 2DCAD takes advantage of effectiveness, efficiency, and robustness for the examined data sets containing several pixel-level anomalies, with respect to the competitors RX, WSCF, CBAD, SVDD, and RSAD. The proposed detector has the ability to detect the true pixel-level anomalies in a greater probability and is sufficiently efficient for real-time application.

REFERENCES

- [1] J. Theiler and B. Wohlberg, "Local coregistration adjustment for anomalous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3107–3116, Aug. 2012.
- [2] A. Prieto, F. Bellas, R. J. Duro, and F. López-Peña, "An adaptive approach for the progressive integration of spatial and spectral features when training ground-based hyperspectral imaging classifiers," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 8, pp. 2083–2093, Aug. 2010.
- [3] H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 2011–2017, Aug. 2010.
- [4] B. Luo, C. Yang, J. Chanussot, and L. Zhang, "Crop yield estimation based on unsupervised linear unmixing of multivariate hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 162–173, Jan. 2013.
- [5] 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.
- [6] 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.
- [7] Q. Du and H. Ren, "Real-time constrained linear discriminant analysis to target detection and classification in hyperspectral imagery," *Pattern Recog.*, vol. 36, no. 1, pp. 1–12, Jan. 2003.
- [8] 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.
- [9] R. J. Johnson, J. P. Williams, and K. W. Bauer, "Autogad: An improved ICA-based hyperspectral anomaly detection algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3492–3503, Jun. 2013.
- [10] 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.
- [11] Q. Du and I. Kopriva, "Automated target detection and discrimination using constrained kurtosis maximization," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 38–42, Jan. 2008.
- [12] 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.
- [13] B. Kedem, *Time Series Anal. Higher Order Crossings*. Piscataway, NY, USA: IEEE Press, 1994.
- [14] B. K. Natarajan, V. Bhaskaran, and K. Konstantinides, "Low-complexity block-based motion estimation via one-bit transforms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 4, pp. 702–706, Aug. 1997.
- [15] B. Demir, A. Celebi, and S. Ertürk, "A low-complexity approach for the color display of hyperspectral remote-sensing images using one-bit-transform-based band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1-1, pp. 97–105, Jan. 2009.

- [16] 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.
- [17] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "Models and methods for automated background density estimation in hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2837–2852, May 2013.
- [18] P. Gurram and H. Kwon, "Support-vector-based hyperspectral anomaly detection using optimized kernel parameters," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1060–1064, Nov. 2011.
- [19] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 646–650, Jul. 2011.
- [20] 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.
- [21] D. W. Stein *et al.*, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.
- [22] 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.
- [23] 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.
- [24] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.
- [25] D. M. J. Tax, A. Ypma, and R. P. W. Duin, "Pump failure detection using support vector data descriptions," in *Proc. Adv. Intell. Data Anal.*, 1999, pp. 415–426.
- [26] 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.
- [27] J. Gaucel, M. Guillaume, and S. Bourennane, "Whitening spatial correlation filtering for hyperspectral anomaly detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 333–336.
- [28] T. Veracini, S. Matteoli, M. Diani, and G. Corsini, "Nonparametric framework for detecting spectral anomalies in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 666–670, Jul. 2011.
- [29] S. P. Catterall, "Anomaly detection based on the statistics of hyperspectral imagery," in *Proc. SPIE Conf. Imagery Spectroscopy X*, 2004, pp. 171–178.
- [30] 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.
- [31] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 733–740.
- [32] A. Adams, J. Baek, and M. A. Davis, "Fast high-dimensional filtering using the permutohedral lattice," *Comput. Graph. Forum*, vol. 29, no. 2, pp. 753–762, May 2010.
- [33] O. Urhan and S. Ertürk, "Constrained one-bit transform for low complexity block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 4, pp. 478–482, Apr. 2007.
- [34] A. Sheinker *et al.*, "Magnetic anomaly detection using high-order crossing method," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1095–1103, Apr. 2012.
- [35] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [36] 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.

Yuan Yuan (M'05–SM'09) is a Full Professor with the Chinese Academy of Sciences, Beijing, China. She has published over 150 papers, including about 100 in reputable journals such as the IEEE TRANSACTIONS AND PATTERN RECOGNITION, and conference papers in the IEEE Conference on Computer Vision and Pattern Recognition, British Machine Vision Conference, IEEE International Conference on Image Processing, and the IEEE International Conference on Acoustics, Speech, and Signal Processing. Her current research interests include visual information processing and image/video content analysis.



Qi Wang 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 Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His current research interests include computer vision and pattern recognition.



Guokang Zhu is currently working toward the Ph.D. degree in 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, China.

His current research interests include computer vision and machine learning.