IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 11, 2018, accepted January 16, 2019, date of publication February 1, 2019, date of current version February 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894590

Hyperspectral Anomaly Detection via Sparse Dictionary Learning Method of Capped Norm

YUAN YUAN[1,2], (Senior Member, IEEE), DANDAN MA[3,4], AND QI WANG[1,2], (Senior Member, IEEE)

[1]School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China
[2]Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, China
[3]Center for Optical Imagery Analysis and Learning, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China
[4]University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Qi Wang (crabwq@gmail.com)

This work was supported in part by the National Key R&D Program of China under Grant 2017YFB1002202, in part by the State Key Program of National Natural Science Foundation of China under Grant 61632018, in part by the National Natural Science Foundation of China under Grant 61773316, in part by the Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, in part by the Fundamental Research Funds for the Central Universities under Grant 3102017AX010, in part by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences, and in part by the Projects of Special Zone for National Defense Science and Technology Innovation.

ABSTRACT Hyperspectral anomaly detection is a research hot spot in the field of remote sensing. It can distinguish abnormal targets from the scene just by utilizing the spectral differences and requiring no prior information. A series of anomaly detectors based on Reed–Xiaoli methods are very important and typical algorithms in this research area, which generally have the hypothesis about background subject to the Gaussian distribution. However, this assumption is inaccurate to describe a hyperspectral image with a complex scene in practice. Besides, due to the unavoidable existence of abnormal targets, background statistics will be affected which will reduce the detection performance. To address these problems, we propose a sparse dictionary learning method by using a capped norm to realize hyperspectral anomaly detection. Moreover, a new training data selection strategy based on clustering technique is also proposed to learn a more representative background dictionary. The main contributions are concluded in threefold: 1) neither making any assumptions on the background distribution nor computing the covariance matrix, the proposed method is more adaptive to all kinds of complex hyperspectral images in practice; 2) owing to the good qualities of the capped norm, the learned sparse background dictionary is resistant to the effect of anomalies and has stronger distinctiveness to anomalies from background; 3) without using the traditional sliding hollow window technique, the proposed method is more effective to detect different sizes of abnormal targets. The extensive experiments on four commonly used real-world hyperspectral images demonstrate the effectiveness of the proposed method and show its superiority over the benchmark methods.

INDEX TERMS Anomaly detection, hyperspectral images, sparse, dictionary learning, capped norm.

I. INTRODUCTION

Hyperspectral image (HSI), as a 3-D cube data, can deliver both rich spectral and spatial information [1]. Its high spectral resolution with hundreds of narrow and approximately continuous spectral bands can provide a strong guarantee for discriminating the subtle differences of surface substances [2]. Owing to this advantage, hyperspectral image (HSI) processing techniques have been widely applied in different research fields, such as hyperspectral target detection [3], [4], hyperspectral image classification [5], [6], band selection [7], [8], and hyperspectral unmixing [9], [10].

Hyperspectral anomaly detection aims at locating the pixel as an abnormal target when it deviates significantly from the given reference background [11]. Since during the detection process any prior information about both the spectra of background and target is unknown [11], anomaly detection can be regarded as an unsupervised target detection problem. Benefiting from this unsupervised characteristic, hyperspectral anomaly detection technique has a high value of practical application in agriculture [12], geology [13], etc.

In more recent decades, lots of hyperspectral anomaly detection methods have been proposed [14]–[19].

Generally, the basic idea of these methods mainly follows some core procedures containing background information obtainment, anomaly measurement definition, and deviation estimation between background and anomalies. Among all these detectors, the Reed-Xiaoli (RX) [20] method is widely acknowledged as one classic algorithm. It has the hypothesis that background obeys a multivariate normal distribution. The mahalanobis distance is consequently used to measure the difference between the test pixel and its designated reference background. Depending on the scale of the background, RX method has two versions: local RX (LRX) (The local neighborhoods are usually defined as background), and global RX (GRX) (The whole image pixels or a large part of them are commonly regarded as background).

However, some intrinsic problems of RX method will result in high false alarms [21]. The original gaussian distribution hypothesis is not optimum to accurately describe different hyperspectral images, especially for the real image scenes which is usually complex in practice. When implementing RX, researchers are accustomed to directly calculating background statistics depending on the hyperspectral pixels in local regions or the whole image. But these used pixels may contain some potential abnormal pixels. As a result, the mahalanobis distance mainly based on the computation of covariance matrix will be affected by the contamination of anomaly targets. Despite the fact that a sliding-dual-window technique [22] is commonly adopted to mitigate the effect of anomalies as much as possible, however, it is hard to get rid of all the possible abnormal pixels from the defined background region. At the same time, it will cause the problem of sensitiveness to the window size.

To tackle the related problems of RX method, a large number of methods have been presented. Some of them are with the purpose of restraining the effect of abnormal targets on the background estimation. For example, Du and Zhang [23] propose a random-selection-based anomaly detector (RSAD), which designs a random selection approach to pick some representative background pixels out of the hyperspectral image. When this random selection is performed in a sufficient number of times, the proper and accurate background will be obtained. Zhao *et al.* [24] present an algorithm named robust nonlinear anomaly detection (RNAD), which is capable of constantly purifying the background through adopting a regression strategy and consequently realizes suppression of contamination of anomalies. Some methods take advantage of kernel technique to nonlinearly project the original hyperspectral image into a high dimensional feature space to strengthen the separability between anomalies and background. The Kernel-RX (KRX) [25] is a typical nonlinear form of RX algorithm. As for the support vector data description (SVDD) [26], it is a non-parametric kernel-based anomaly detection method. Through constructing an minimum enclosing hypersphere to envelop the background pixels, it recognizes the pixels falling outside this hypersphere as anomalies in the high dimensional feature space. In addition, the background suppression is also a good strategy to improve anomaly detection ability. For example, the subspace RX [27] adopts the orthogonal subspace projection approach to project the data onto a subspace that is orthogonal to the background clutter in order to suppress the undesired background.

A number of representation-based methods [28]–[33] have been proposed of late years which commonly have the philosophy that hyperspectral pixels can be well represented by a dictionary. For example, Wei and Du [34] propose a collaborative-representation-based detector (CRD) to detect anomalies through considering that a background pixel commonly has a approximately linear representation of its surrounding samples, which is based on a sparse representation-based detector (SRD). Yuan *et al.* [29] propose an anomaly detection method based on the local sparsity divergence, which directly adopts a sliding window to compute the sparsity difference between anomalies and background. Li *et al.* [32] use the background joint sparse representation to detect anomalies in the hyperspectral image. This method aims to select the most active dictionary bases regarded as the representative background. In general, background is also supposed to have low rank property because of lying in a low-dimensional subspace; anomalies usually have a sparse property for their low occurrence probabilities. Therefore, some researchers also utilize these properties to design their detectors. For example, Zhang *et al.* [30] propose a low-rank and sparse matrix decomposition-based mahalanobis distance method (LSMAD) for Hyperspectral anomaly detection through using the low-rank characteristic to estimate the background and further applying mahalanobis distance to detect the probable anomalies.

Based on the aforementioned methods, it is easy to draw a conclusion that sparse representation is really a good model to describe the hyperspectral images. Owing to its virtue, in this paper we propose a novel hyperspectral anomaly detection method via sparse dictionary learning of capped norm (SDLCN) to overcome the problems involved in the RX based methods. It is formulated based on two facts: 1) for the background pixel in a hyperspectral image scene, it is commonly similar to its local surroundings, and 2) the total number of background classes contained in the hyperspectral image is limited. Consequently, the background usually contains a great deal of redundant information [32], and a dictionary can be definitely learned to well characterize the background knowledge. Specifically, when finishing learning the background dictionary, a background pixel will be represented sparsely by some bases, but abnormal pixels can not satisfy this condition. Generally, in order to learn the dictionary, we need firstly build the training data. It is a simple and direct way to construct this data set through picking a great number of pixel patches out of the entire hyperspectral image. Nevertheless, in spite of anomalies having low probabilities to be selected to contribute to learning the background dictionary for their low population, it is still hard to completely avoid the existence of anomalies in the training set. Consequently, the representation ability of dictionary will be

affected by the potential abnormal targets which can further result in the reduction of detection performance. Therefore, in this work, we impose the capped norm [35], [36] on the learning process, which can be understood as a weight constraint capable of suppressing anomalies' effect. To conclude, our main contributions can be identified as follows.

1) Neither making any assumptions on the background distribution nor computing the covariance matrix, the proposed method is more adaptive to all kinds of complex hyperspectral images in practice.

2) Owing to the good qualities of capped norm, the learned sparse background dictionary is resistant to the effect of anomalies and has stronger distinctiveness to anomalies from background.

3) Without using the traditional sliding hollow window technique, the proposed method is more effective to detect different sizes of abnormal targets.

In addition, we need to state that this work is actually an extension version of our previous published conference paper [36]. The major extensions consist of the improvement on the method technique, substantial experiments and new promising results. Overall, this work contains more than 60% new material over the original version. Here, we will briefly introduce the main extensions. In terms of method innovation, we propose a new training data selection strategy based on clustering technique in order to learn a more representative background dictionary. Some important formula deductions are added to make it more rigorous and comprehensible as well. In the experiment, we add other three real-world hyperspectral images with higher used frequency to make the performance evaluation more convincing. We also add parameters setting discussion, convergence speed analysis and time consumption comparison. Besides, many more state-of-the-art methods are used to compare with the proposed method and demonstrate its performance. In this paper, we make sufficient experiments, in-depth analyses, and comprehensive discussions. We have obtained many novel results.

The remainder of this paper is organized as follows. Section II introduces the proposed SDLCN method in detail. In Section III, experimental results on four commonly used real-world hyperspectral images and the corresponding analyses are reported. Finally, we conclude this paper in Section IV.

II. SDLCN

We propose a novel anomaly detection method via sparse dictionary learning of capped norm named SDLCN in this work. The philosophy of our method is that a background pixel can be well represented by a sparse background dictionary, while an anomaly pixel can not satisfy this condition for it will have a large reconstruction error. In order to avoid the effect of anomalies on the background dictionary learning, the capped norm is employed to deal with anomalies existing in the original background training set. Consequently, the anomalies with the larger reconstruction errors will be penalized in the dictionary learning process. The learned background dictionary can only code well for the definite background pixels but not find codings for anomalies. We also adopt a novel strategy to construct the training data for dictionary learning. In order to learn a dictionary with a stronger representative ability, a clustering method is applied in this work. A number of samples are selected from each clusters to construct the training data and further to generate the initial dictionary. The proposed method will be introduced in detail in the following parts.

A. BACKGROUND DICTIONARY

This part will elaborate the proposed background dictionary learning process. Based on the sparse representation framework, it is commonly thought that the background pixel can be sparsely linearly represented by some dictionary bases. For a 3-D hyperspectral image, with height of m pixels, width of n pixels, and B spectral bands, we reshape it into a 2-D representation $I \in R^{mn\times B}$. Here mn is the total number of whole hyperspectral pixels. We select N pixels from the entire image in order to build the training data $H = [h_1, ..., h_N] \in R^{B\times N}$, and each column of matrix H denotes a hyperspectral pixel. Thus, the classic sparse dictionary learning [37] technique can be formulated as:

$$\min_{D,A,\|d_i\|\leq 1} \|H - DA\|_F^2 + \lambda\|A\|_1. \tag{1}$$

Here, $A \in R^{K\times N}$ denotes the sparse coefficient matrix, and $D \in R^{B\times K}$ denotes the background dictionary matrix consisting of K bases to learn. In addition, λ is the balance parameter. Generally, the constraint that the l_2 norm value of each column vector in D is not larger than 1, is usually used to ensure the l_2 norm of D not being extremely larger and consequently the element values in matrix A not being arbitrarily small.

In this work, we aim to propose a novel anomaly detection method from the perspective of learning a good dictionary having a superb ability to characterize the background. However, since the quadratic loss function is used, the traditional dictionary learning method can be affected by anomalies to an extent. Specifically, when there are some anomalies, dictionary learning process will be deeply disturbed and degenerate. Even though anomalies in a hyperspectral image scene usually have a very low population, to completely ensure their absence from the training data is still impossible. As a result, the final performance of anomaly detection will be decidedly affected. To address this problem, we impose the capped l_1 norm on the loss function to learn a superb background dictionary. Through penalizing anomalies in the learning process, the dictionary will be resistant to the effect of anomalies in the training set.

In this work, we define the capped l_1 norm based loss function as follows:

$$l_\varepsilon(r_i) = \min(|r_i|, \varepsilon), \tag{2}$$

where r_i is the recovery error of the i-th sample, and ε denotes a boundary parameter. When $|r_i|$ is bigger than ε, the loss

function is directly equal to ε. This means that if a sample has a large error, then it will be penalized by capped l_1 norm. Therefore, the proposed sparse dictionary learning based on capped l_1 norm is formulated as follows:

$$\min_{D,A,\|d_i\|\le 1} \sum_{i=1}^{N} \min\left(\|h_i - Da_i\|_2, \varepsilon\right) + \lambda\|A\|_1. \tag{3}$$

Our objective function is composed of a concave function (the first item) and a convex function (the second item). This function is equivalent to (4) as follows.

$$\min_{D,A,\|d_i\|\le 1} \sum_{\{i|\|h_i - Da_i\|_2 \le \varepsilon\}} \|h_i - Da_i\|_2 + \sum_{\{i|\|h_i - Da_i\|_2 > \varepsilon\}} \varepsilon + \lambda \|A\|_1. \tag{4}$$

The second term of (4) is a constant, which will not affect solving the optimal values of parameters. Thus (4) is further equivalent to the following function:

$$\min_{D,A,\|d_i\|\le 1} \sum_{\{i|\|h_i - Da_i\|_2 \le \varepsilon\}} \|h_i - Da_i\|_2 + \lambda \|A\|_1. \tag{5}$$

From (5), it can be clearly seen that our method can adaptively select the samples with lower reconstruction errors to conduct the dictionary learning. Consequently, the background pixels will be selected from the original training data set to learn the dictionary. But anomalies contained in the training data will not participate in the dictionary learning due to their larger reconstruction errors. Therefore, the learned dictionary has a good ability to represent background through directly restraining the effect of anomalies. By introducing the auxiliary variables s_i, (5) can be rewritten as (6) which is expressed as:

$$\min_{D,A,\|d_i\|\le 1} \sum_{i=1}^{N} s_i \|h_i - Da_i\|_2^2 + \lambda\|A\|_1$$
$$s_i = \begin{cases} \dfrac{1}{2\|h_i - Da_i\|_2}, & if\ \|h_i - Da_i\|_2 \le \varepsilon \\ 0, & otherwise \end{cases} \tag{6}$$

We make use of a re-weighted technique to solve (6). In this case, we alternatively update D, A, and s_i to get its solution. Specifically, with A fixed, the above mentioned problem becomes:

$$\min_{D,\|d_i\|\le 1} \sum_{i=1}^{N} s_i \|h_i - Da_i\|_2^2. \tag{7}$$

It is equivalent to the following function:

$$\min_{D,\|d_i\|\le 1} \sum_{i=1}^{N} \|x_i - Dz_i\|_2^2, \tag{8}$$

where $x_i = \sqrt{s_i}h_i$ and $z_i = \sqrt{s_i}a_i$. We adopt the same method proposed in [37] to solve this problem.

Through fixing D, we can decompose (6) into N independent subproblems as follows:

$$a_i = \arg\min_a s_i\|h_i - Da_i\|_2^2 + \lambda\|a\|_1. \tag{9}$$

When $s_i \neq 0$, it can be rewritten as :

$$a_i = \arg\min_a \|h_i - Da_i\|_2^2 + \frac{\lambda}{s_i}\|a\|_1. \tag{10}$$

Obviously, as a Lasso problem, the solution of formulation (10) will be obtained efficiently. The convergence analysis of the proposed SDLCN is similar to [35].

B. DATA PREPARATION FOR LEARNING

This paper aims to learn a sparse dictionary which has an excellent ability to express the background and the stronger distinctiveness to anomalies. In order to learn the dictionary, some data preparations should be made for training. In the aforementioned method, both training data H and initial dictionary D play very important roles in the learning process.

A direct approach is to randomly select some pixels from the hyperspectral image to build the training data H, and further pick out some samples from H to generate the initial dictionary D. Although our method does not need to consider the contamination of anomalies probably contained in the training set, because it can effectively suppress their effects on learning the background dictionary, random selection may result in the problem of inadequate representativeness of the selected samples. The reason is that random selection tends to choose the major materials while other materials with relatively low population will be missed. This problem may not be significant when the image scene is relatively simple, but it becomes serious for a complex image scene covering many kinds of materials that is quite common in practice. As a result, it will lead to a high false alarm rate because of easily recognizing the material with the low population as an anomaly.

Inspired by the traditional clustering, a simple K-means++ technique [38], [39] is used to prepare the training data for its good stability. A hyperspectral image is firstly divided into C clusters. We set the number of training samples as N. Then from each cluster we select some samples which are nearest to each cluster center. The number of the selected samples from each clustering is directly in proportion to the amount of each cluster with insuring that the sum of samples in H is N. As for the initial background dictionary, when the value of dictionary size K is given, the selection way is the same as the construction of training set.

C. ANOMALY DETECTION

In this part, the final anomaly detection process will be briefly introduced. Using the above proposed dictionary learning technique with capped norm, we have obtained the background dictionary D. Therefore, according to the reconstruction representation shown in (1), the sparse reconstruction error of each hyperspectral pixel can be computed, which can be regarded as the corresponding anomaly probability.

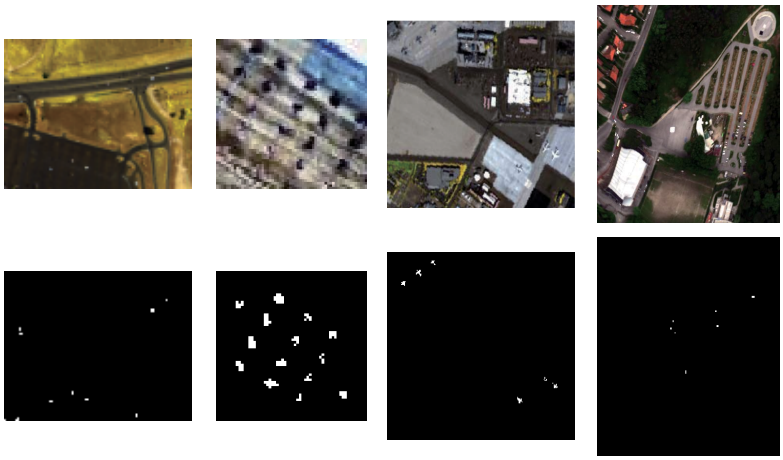


FIGURE 1. The visualization figures and the corresponding ground truth maps of different HSIs. The upper row presents the false color pictures of Urban, AVIRIS1, AVIRIS2, and D1F12H1 respectively. The lower row shows the ground truths.

In this case, the entire hyperspectral image matrix I is used to construct sample data H, D is the learned background dictionary and the same λ is used to detect anomalies. The sparse coefficient matrix A is firstly computed, and then reconstruction errors are obtained.

III. EXPERIMENTS AND ANALYSES

In this section, we carry out several experiments and analyze the performance of the proposed method deeply. First, four commonly used real hyperspectral images are introduced. Then, we elaborate experimental setup, containing evaluation metrics, related competitors, and parameter setting. Afterward, the parameter selection rules and their effects are discussed. Finally, we illustrate the experimental results and analyze the effectiveness of the proposed method in detail.

A. DATA SETS

In the experiment, four commonly used real-world hyperspectral images are used to evaluate the performance of our proposed SDLCN. We will introduce the detailed information of these images as follows.

The first data set is the HYDICE Urban image. It can be downloaded from the website of U.S. Army Engineer Research and Development Center.[1] This data is collected by HYDICE on an airborne platform which scans an urban scene. Its spectra range from 400 nm to 2500 nm with the resolution of 10 nm. Its spatial resolution is about 1 m. By convention, researchers commonly crop a relatively smaller sub-image from the upper right part of the original hyperspectral image with a larger size of 307×307, because of the undefined ground truth for the whole scene. Therefore, in this work, we also obtain the sub-image having a size of 80×100 according to the same criterion. As for the spectral bands, considering the low-SNR and water vapor absorption, we finally remain 160 bands after removing totally 50 bands in accordance with [40]. The ground truth map for the sub-image is defined in consistent with literature [41], in which several cars and roofs are defined as abnormal targets. The visualization picture and ground truth map of this cropped hyperspectral image are shown in the first column of Fig. 1.

The second and the third data sets are called as AVIRIS1 and AVIRIS2 respectively referring to the work [42]. They are different sub-images of one same hyperspectral image which is collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) from San Diego, CA, USA. Their spatial resolution is 3.5 m, and the spectral resolution is about 10 nm. As for the spectral information, considering that the original 224 bands with spectra ranging from 370 nm to 2510 nm contain some useless bands due to water absorption regions or low-SNR, we remove several bands and finally retain 189 bands in the experiment according to [4]. The size of AVIRIS1 is 60×60 and it contains 14 planes as anomalies. As for AVIRIS2, its size is 200×240 including 6 planes as abnormal targets. Their visualization pictures and ground truth maps are shown in the second and third columns of Fig. 1 respectively.

The fourth data named D1F12H1 is collected by visible near infrared (VNIR) camera of SIM.GA sensor from Viareggio, Italy [43]. Its spectral range is from 400 nm to 1000 nm. Its spectral resolution is up to 2.3 nm, and spatial resolution is about 0.6 m. This data set has 511 spectral channels with the size of 450×375 mainly covering parking lot, suburban, and vegetated area. Five kinds of different targets are defined as anomalies including a blue vehicle, a green vehicle, a white vehicle covered with red tarp, three green panels, and a bright green panel in this image. The fourth column of Fig. 1 presents its visualization picture and ground truth map.

B. EXPERIMENTAL DETAILS

In this section, we introduce the evaluation criteria, the competitors and parameter setup in detail as follows.

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

1) EVALUATION CRITERIA

It is very important to utilize a valid evaluation criterion to impartially estimate and analyze one anomaly detection method's performance. The receiver operating characteristic (ROC) curve is a classic comparison measurement. It is plotted by a set of values of target detection rates and false alarm rates reflecting the relationship between these two rates. The two rates can be effectively computed under a given discrimination threshold. The ROC curve can illustrate the qualitative analysis. For an accurate quantitative analysis, the area under the curve (AUC) is further calculated through integrating the area under the ROC curve. The AUC value can intuitively reflect the performance of one detector.

2) COMPETITORS

In order to accurately demonstrate the effectiveness of the proposed method, our method is compared with eight state-of-the-art detectors by comprehensively taking popularity, recency, and variety into consideration: global RX (GRX), local RX (LRX), RSAD, SVDD, SRD, CRD, LSMAD, and the traditional sparse dictionary learning method (DL) [37] that is used to directly demonstrate the effectiveness of our method using capped norm to restrain the effect of anomalies on background. These competitors are usually used as benchmark detectors in the experiment of hyperspectral anomaly detection research area. Therefore, it is convincing to compare with them for analyzing and evaluating one method's performance.

3) PARAMETER SETUP

Some critical parameters involved in the experiments are elaborated in this part. For the proposed SDLCN, the main parameters containing the number of training data N, the dictionary size K, the number of clusters C, the boundary parameter ε, and the balance parameter λ should be given. For simplicity, we fix N at 1000 for each hyperspectral image. The K is determined according to the number of spectral bands B of each HSI. In order to obtain a complete dictionary, it requires that $K > B$. Therefore, the values of K are set as 300, 300, 500, and 600 for Urban, AVIRIS1, AVIRIS2 and D1F12H1 respectively. Their corresponding number of clusters C are set as 10, 3, 2, and 10 mainly based on the complexity of each image scene. Due to the low occurrence probability and small population of abnormal targets, we empirically take 99.5 percentile of all the reconstruction errors as the value of ε during iterations in the training process. In addition, λ is fixed at 0.01 for all the data sets. The detailed parameter selection rules will be further discussed in section III-C.

For other competitors, since LRX, SVDD, SRD, and CRD require the sliding window technique, the sizes of outside window and inner window should be set. Due to different hypespectral images having different targets, proper sizes are pretty necessary because different sizes will greatly affect the detection performance. In order to make a fair and convincing comparison, different pairs of window sizes are selected to fully test these four methods and obtain their best detection performance. Considering the largest expected size of anomalies in all the image scenes, two different inner window sizes, 7×7 and 9×9, are defined in the experiment. Consequently outer window sizes with larger values are required for better detection performance. Since the former three hyperspectral images have similar spectral band number, we simply set the same window size for these data sets. In this case, the detailed window sizes for different competitors are introduced as follows. For LRX method, 4 pairs of window sizes $(wout, win)$ containing $(17, 7)$, $(17, 9)$, $(19, 7)$, and $(19, 9)$ are used in the experiments by comprehensively taking all the spectral dimensions of the three images into consideration and simultaneously avoiding covariance matrix singular problem. As for SVDD, owing to its heavy computation burden, we finally select $(13, 7)$, $(13, 9)$, $(15, 7)$, and $(15, 9)$. In addition, six pairs of different sizes consisting of $(13, 7)$, $(15, 7)$, $(17, 7)$, $(17, 9)$, $(19, 7)$, and $(19, 9)$ are defined for SRD and CRD. The regularization parameter λ of CRD is fixed at 10^{-6} consistent with the original work [34]. Since D1F12H1 data set has a larger number of bands, we consistently select $(25, 7)$ and $(25, 9)$ for LRX, SRD, and CRD. SVDD is not compared on this data because of the foreseeable unaffordable time consumption. As for RSAD, we set the size of the randomly selected image block as 70. We define the parameter of LSMAD in strict accordance with the original literature [30], through setting the maximal rank of background matrix r as 2 and the cardinality of the sparse matrix k as 0.005. For the implementation of DL, all the processing setup is exactly the same as our SDLCN except for without using the capped constraint.

C. PARAMETERS SETTING DISCUSSION

In this part, the problems of parameters setting related to the proposed SDLCN method are discussed. The first one is about the dictionary size K and the number of clusters C. The second one is to analyze and determine the proper value of the balance parameter λ.

1) SENSITIVITY TO CLUSTERS AND DICTIONARY SIZES

The proposed method adopts a clustering technique and sparse representation to construct the anomaly detection approach. In this process, two parameters including the number of cluster C, and the number of background dictionary bases K will be set with different values to analyze the parameter selection problem. We carry out the parameter experiment on the four hyperspectral images to completely and accurately discuss parameters' effect. Considering the different complexities of different hyperspectral image scene, the number of cluster C is set as 2, 3, 4, 5, 6, 8, 10, 15, and 20 changing in a wide range. The dictionary size K is respectively set as different values from 200 to 1000 for the former three images, and from 600 to 1000 for the fourth image at the interval of 100. The corresponding results are orderly shown from Fig. 2 to Fig. 5.

The results of Urban image are shown in Fig. 2. When C varies within 10, the AUC values nearly remain unchanged along with the increasing of K. Then the AUC value gradually becomes larger as C increases at each dictionary bases number K. When C becomes larger than 10, AUC still achieves better performance with K within 700. Overall, for the Urban data set, when the number of bases K is limited within 700 and the number of clusters takes a relatively bigger value, our SDLCN can stably achieve good detection performance. Therefore, it can be concluded that SDLCN can perform well over a wide range of K values and different C values will have some impacts on performance. The number of clusters C tends to be set as a larger value when the tested hyperspectral image is complex covering many categories of ground materials. So we finally set C as 10 and K as 300 for Urban image.

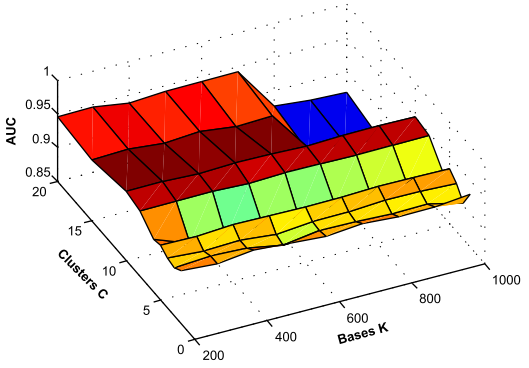


FIGURE 2. Performance of the proposed method under different values of C and K on Urban.

The results in Fig. 3 generally show that the proposed method is assuredly robust to parameter K on AVIRIS1 image. As for C, the detection performance of SDLCN improves as the number of clusters C increases to 3. After that, the detection performance firstly obviously decreases and then almost stably changes with minor fluctuations as C increases. The behind reason for this phenomenon may be explained as follows. For one thing, since the proportion of abnormal targets is very large, anomaly detection for AVIRIS1 is more difficult. Consequently, when C is smaller, a few abnormal targets may be contained in the training data set and SDLCN can effectively restrain their effects on background dictionary. But when C becomes larger, too many anomalies will be contained in the training data set and some of them tend to be regarded as background which results in the reduction of restraining. For another, the whole scene of AVIRIS1 is relatively simple and it has two or three dominate categories. Therefore, a smaller value for C is better. Therefore, we finally set C as 3 and K as 300 for AVIRIS1.

From the results of AVIRIS2 in Fig. 4, there is a same conclusion that the changes of K do not have obviously effects on detection performance. As for C, the AUC values tend to go down firstly then gradually up as C increases within 15. After that, AUC decreases again as C increases. However, it is worth noting that all the AUC values are gradually changing limited in a higher value range, which demonstrates the good robustness to parameters' effects. Since the scene of AVIRIS2 is relatively simple, we finally fix C at 2 and K at 500 for a promising detection result. Fig. 5 shows the results of D1F12H1 image. It can be seen that the analysis about K on this data is also similar. Our method is really robust to the variation of K. As for the effect of C, the AUC values tend to increase firstly as C increases to 10. After that, the AUC

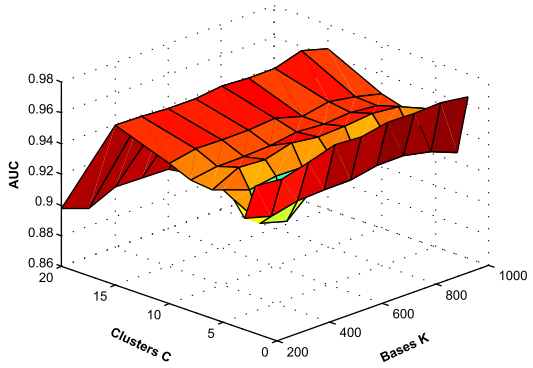


FIGURE 4. Performance of the proposed method under different values of C and K on AVIRIS2.

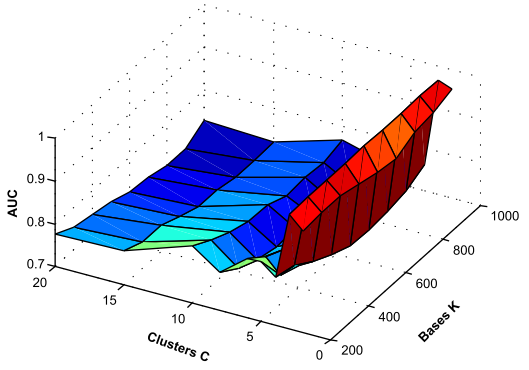


FIGURE 3. Performance of the proposed method under different values of C and K on AVIRIS1.

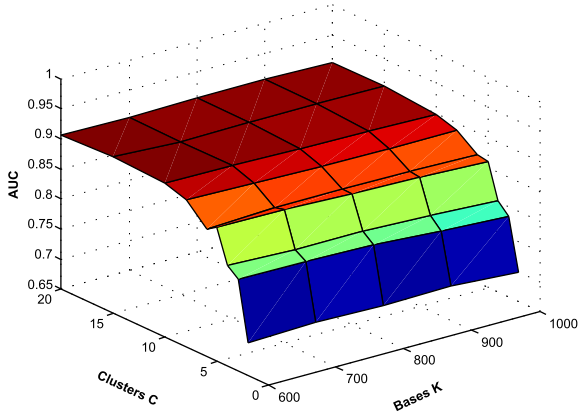


FIGURE 5. Performance of the proposed method under different values of C and K on D1F12H1.

values remain almost unchanged as C increases. We finally set C as 10 and K as 600 because the scene of D1F12H1 is relatively complex.

Although different images have different proper parameter values for better performance, it is difficult to find a perfect value to satisfy all the hyperspectral images. However, through analyzing all the results, our method has nearly obtained the decent performance for all the images even on different parameter values instead of finding the optimal parameter values. Considering conveniently implementing our method in practice, we have some good suggestions of parameter selection. For the number of clusters C, it can be selected from the range of 2 to 10. A smaller value will be set when the hyperspectral contains less categories, while a larger value is required for a complex scene. Since our method is really robust to the number of bases K, it can be assigned to some relatively smaller values under the condition of complete dictionary for efficiency.

2) SELECTION OF BALANCE PARAMETER λ

The balance parameter λ reflects the strength of the sparse constraint. A proper λ is also important for the detector. This experiment is also conducted on all the hyperspectral images. The λ is chosen from $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.2, 0.4, 0.6, 0.8\}$, while the other parameters are fixed. The results of the proposed method with different λ values are illustrated in Fig. 6. The trends of four curves are almost the same. They all begin to gradually improve, then slowly go down, and finally almost maintain stable only having some slightly floating changes as λ increases within 0.2. After that, they obviously descend as λ increases to 0.8. It can be seen that the detection performance is relatively stable and promising when $\lambda \in [0.01\ 0.2]$ for all the hyperspectral images. Therefore, we have empirically set λ as 0.01 instead of selecting the optimal valueS for all the employed data sets in our experiments for simplicity.

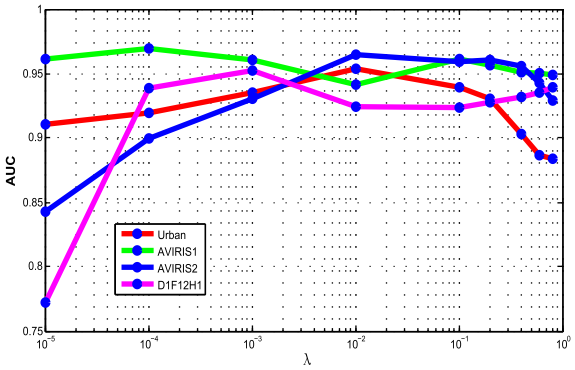


FIGURE 6. The AUC values of the proposed method SDLCN with different λ values on Urban, AVIRIS1, AVIRIS2, and D1F12H1.

D. COMPARISON RESULTS

Before analyzing the detection results, we experimentally discuss the speed of convergence of our proposed SDLCN firstly. This experiment is conducted on all the hyperspectral images. The involved parameters are set as the above mentioned. The sparse reconstruction errors of each iterations during the background dictionary learning period are shown in Fig. 7. For the four different images, our method can converge quickly. Considering efficiency and effectiveness, we fix the iteration time at 5 for all the experiments. In the following part, we will continue analyzing the detection performance.

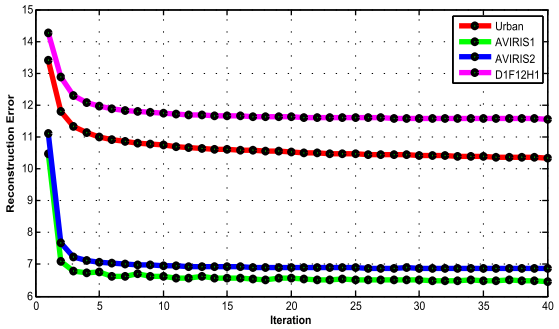


FIGURE 7. The reconstruction errors of each iteration during the background dictionary learning period on Urban, AVIRIS1, AVIRIS2, and D1F12H1.

We evaluate our method's performance through comparing with eight competitors. All the experiments are conducted on the machine with Intel Core i3-2130 3.4-GHz CPU and 16-GB RAM in the MATLAB platform. Through the qualitative and quantitative comparisons, we analyze the experimental results in depth. Considering the convenience of typography, for the competitors using the sliding window technique, we only illustrate their visualization results and ROC curves corresponding to the optimal window sizes and underline these values as shown in Table 1 and Table 3. Table 1 to Table 4 report AUC values and time consumption obtained by using all of window sizes. We also highlight the best AUC values in bold for each data set.

The visualization comparison results of Urban data set are presented in Fig. 8. The results in sequence correspond to GRX, RSAD, LRX, SVDD, SRD, CRD, LSMAD, DL, and our SDLCN. It can be observed visually that RSAD's performance is really poor with a very high false alarm rate. GRX and SVDD seem to have some omissions, and LRX is not intuitive to analyze. For other methods, each of them almost detects all the abnormal targets, only assigning different intensities to each anomaly intuitively. However, it is clear that an accurate estimation cannot be obtained by visual inspection. We need to make a further quantitative comparison. Fig. 10 illustrates the ROC curves of all the detectors. The trends of all the curves are complex and no one obviously stands out. Our method exhibits a slightly lower detection probability when the false alarm rate is within about 0.004; fortunately, it quickly goes up and nearly keeps above all other curves except for SRD and CRD in the case of a larger false alarm rate. But CRD is really sensitive to window sizes as shown in Table 1. If the size is not inappropriate,

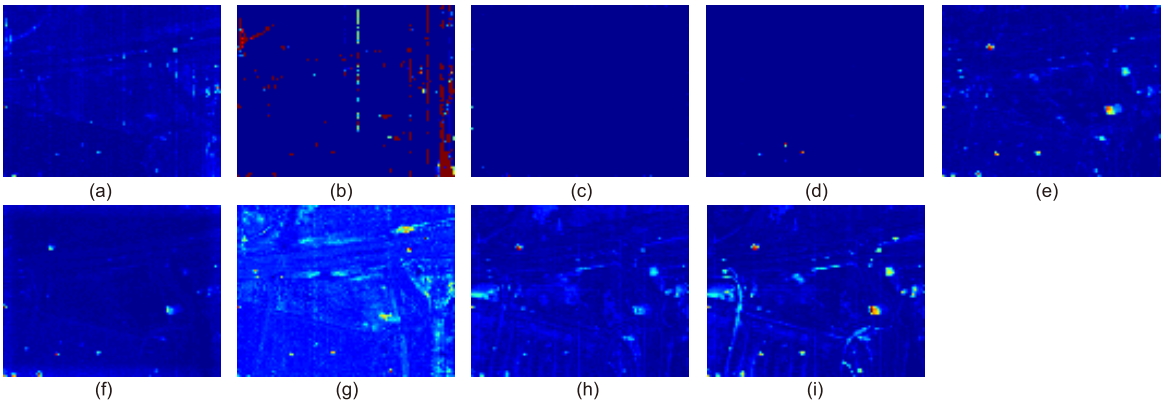


FIGURE 8. The visualization of the anomaly detection results of different detectors on Urban. (a) GRX. (b) RSAD. (c) LRX: window size (19,7). (d) SVDD: window size (13,7). (e) SRD: window size (15,7). (f) CRD: window size (13,7). (g) LSMAD. (h) DL. (i) SDLCN.

TABLE 1. AUC values of all the competitors on Urban, AVIRIS1, and AVIRIS2. We underline the better results corresponding to the optimal window sizes. The highest AUC value is highlighted in bold.

AUC		Urban	AVIRIS1	AVIRIS2
GRX		0.8906	0.7601	0.8343
RSAD		0.8786	0.4845	0.6717
LRX	(17,7)	0.8619	0.5849	0.7627
	(17,9)	0.8467	0.8104	0.6960
	(19,7)	0.9102	0.4843	0.8314
	(19,9)	0.8694	0.4516	0.8644
SVDD	(13,7)	0.8817	0.8867	0.8751
	(13,9)	0.8733	0.8742	0.9212
	(15,7)	0.8476	0.8370	0.9088
	(15,9)	0.8512	0.8567	0.9555
SRD	(13,7)	0.9456	0.7809	0.8034
	(15,7)	0.9475	0.7616	0.8093
	(17,7)	0.9438	0.7625	0.8112
	(17,9)	0.9359	0.7614	0.8400
	(19,7)	0.9459	0.7476	0.8152
	(19,9)	0.9423	0.7447	0.8425
CRD	(13,7)	**0.9694**	0.4904	0.5013
	(15,7)	0.5709	0.6288	0.4183
	(17,7)	0.4902	0.4515	0.4872
	(17,9)	0.5135	0.4639	0.4794
	(19,7)	0.3372	0.4663	0.4901
	(19,9)	0.4499	0.4521	0.5031
LSMAD		0.9318	0.8329	0.9039
DL		0.9276	0.8025	0.9640
SDLCN		0.9533	**0.9659**	**0.9652**

TABLE 2. Time consumption of all the competitors on Urban, AVIRIS1, and AVIRIS2 (s).

Times		Urban	AVIRIS1	AVIRIS2
GRX		0.75	0.28	4.12
RSAD		18.70	10.33	183.31
LRX	(17,7)	84.59	54.21	756.45
	(17,9)	97.69	62.93	915.49
	(19,7)	73.48	47.35	647.34
	(19,9)	103.03	67.01	958.67
SVDD	(13,7)	3774.72	1542.70	24784.29
	(13,9)	2118.98	877.72	13802.01
	(15,7)	7358.77	3064.08	48979.88
	(15,9)	5295.20	2154.86	38947.73
SRD	(13,7)	5.23	2.43	35.44
	(15,7)	7.43	3.60	52.29
	(17,7)	9.32	4.80	67.89
	(17,9)	11.72	7.67	92.47
	(19,7)	11.29	5.79	94.21
	(19,9)	16.81	9.21	119.14
CRD	(13,7)	14.17	6.913	99.47
	(15,7)	29.00	11.64	178.60
	(17,7)	53.73	22.87	404.53
	(17,9)	48.22	22.44	344.06
	(19,7)	87.37	41.11	616.60
	(19,9)	91.96	36.97	617.99
LSMAD		17.12	9.59	127.27
DL		4.14	4.34	30.33
SDLCN		78.19	175.04	412.20

CRD even fails to detect targets. As for SRD, our AUC is higher than all of its results in different window sizes. Our method also performs better than DL, which verifies the good ability of capped norm to suppress anomaly. Moreover, the time consumption of SDLCN is relatively low as shown in Table 2. In conclusion, the good performance of tour proposed method has been demonstrated by experimental results. It can accurately characterize background resistant to anomalies and has a good ability to detect different abnormal targets.

Fig. 9 illustrates the detection results on AVIRIS1. This hyperspectral image is relatively more difficult to detect anomalies because the number of targets is larger and their locations are relatively intensive. It can be seen that RSAD method tends to regard the local background region as anomalies but fails to detect the real targets. SRD and CRD nearly lose the ability to identify anomaly from background because they have a serious problem of high false alarm rate. GRX, LRX and SVDD have many omissions and their detected targets are not remarkable compared with the intensities of background. The results of LSMAD, DL, and SDLCN are relatively similar, and the positions of their detected targets are nearly consistent. However, SDLCN assigns the higher scores to some targets and meanwhile shows the better ability to suppress the background than LSMAD and DL do. Fig. 11 plots the ROC curves of all the detectors on this image. It can be seen that among all the detection curves

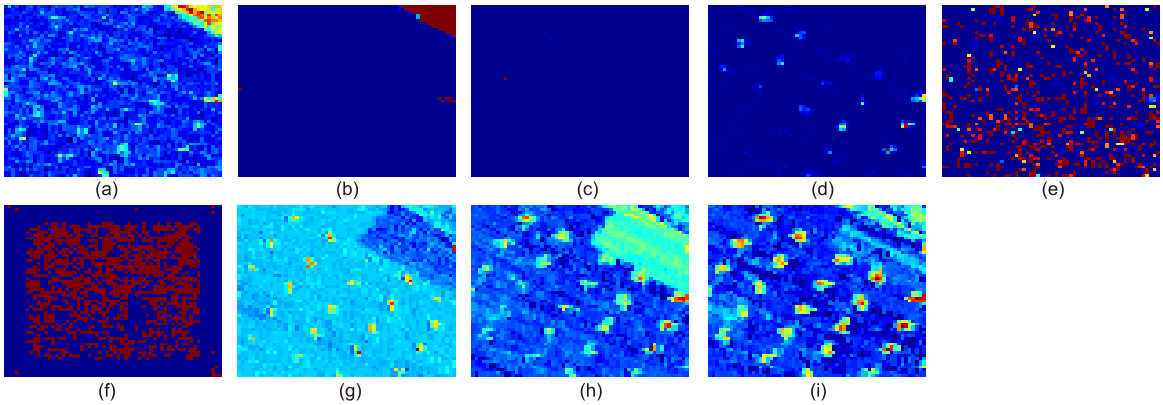


FIGURE 9. The visualization of the anomaly detection results of different detectors on AVIRIS1. (a) GRX. (b) RSAD. (c) LRX: window size (17,9). (d) SVDD: window size (13,7). (e) SRD: window size (13,7). (f) CRD: window size (15,7). (g) LSMAD. (h) DL. (i) SDLCN.

TABLE 3. AUC values of all the competitors on D1F12H1. We underline their better results corresponding to the optimal window sizes. The highest AUC value is highlighted in bold.

D1F12H1	GRX	RSAD	LRX		SRD		CRD		LSMAD	DL	SDLCN
			(25, 7)	(25, 9)	(25, 7)	(25, 9)	(25, 7)	(25, 9)			
AUC	0.8495	0.8190	0.8242	0.8520	0.8719	0.9020	0.5360	0.5830	0.6394	0.8788	**0.9240**

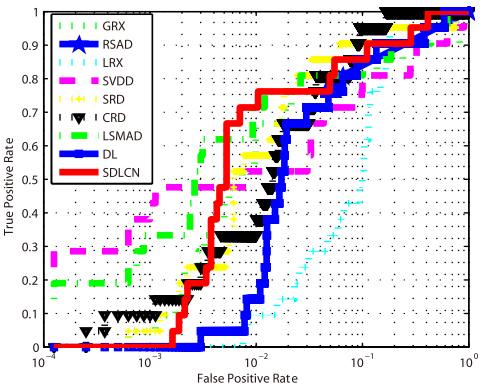


FIGURE 10. The ROC curves of the anomaly detection results on Urban.

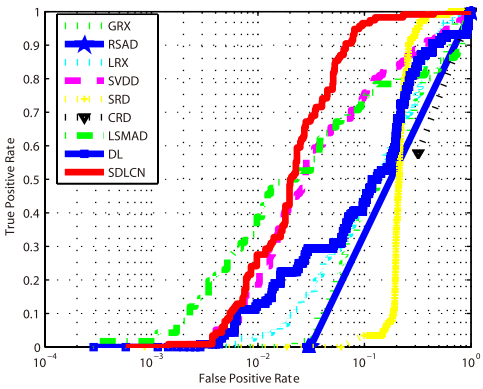


FIGURE 11. The ROC curves of the anomaly detection results on AVIRIS1.

when the false positive rate is relatively low, SDLCN only exhibits a slightly lower detection probability than that of LSMAD. But when false positive rate increases a little bit, SDLCN steeply goes upward and keeps staying over the other curves gaining a higher detection probability than all the other competitors. On the whole, the detection performance of our SDLCN is still better. The highest AUC value of SDLCN further demonstrates the excellent ability to detect anomalies as shown in Table 1. Compared with other competitors' barely satisfactory or even bad results, the proposed SDLCN apparently performs the best. Moreover, its significant superiority to DL demonstrates that the capped norm is useful to restrain anomalies' effects on the background dictionary learning which can enhance the discrimination between anomaly and background.

For AVIRIS2 image, the visualization results are shown in Fig. 12. The performance of RSAD is still poor for it has recognized many background pixels as anomalies. SRD and CRD also have the problem of high false alarm rate. As for GRX, LRX, SVDD, and LSMAD, they can locate some anomalies but the targets do not stand out saliently. The results of DL and our SDLCN are similar, and both of them can detect all of the abnormal targets with high detection values. The ROC curves of all the methods are shown in Fig. 13. Compared with SVDD, although our method exhibits a lower detection probability as the false alarm rate increases to about 0.02, however, it completely defeats SVDD after that. The curve of SDLCN almost coincides with that of DL, and they are obviously above the curves of another seven competitors. Both of them nearly achieve the highest detection probabilities for all false alarm rates. The corresponding AUC values are presented in Table 1. Our method obtains the highest AUC value. When compared with DL, our SDLCN shows the slight superiority. Nevertheless, it significantly outperforms other competitors with a convincing margin. Although SVDD also has achieved a relatively good AUC

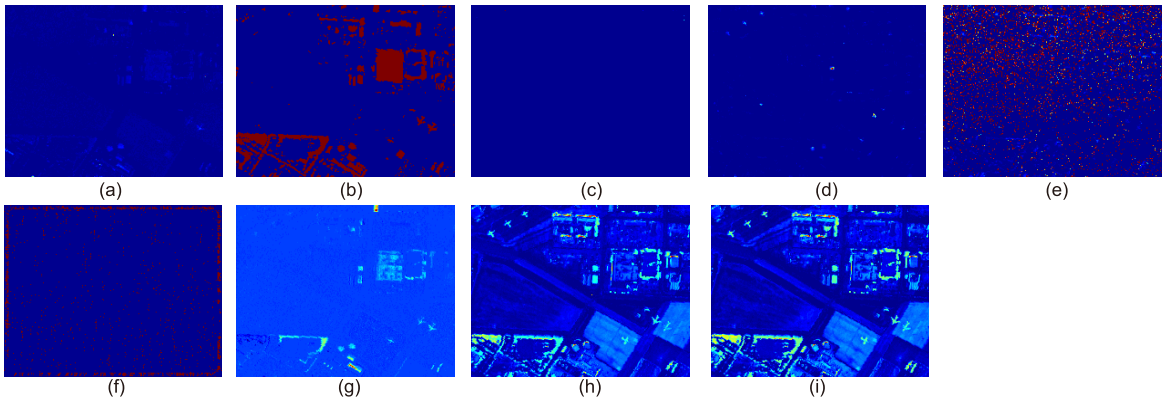


FIGURE 12. The visualization of the anomaly detection results of different detectors on AVIRIS2. (a) GRX. (b) RSAD. (c) LRX: window size (19,9). (d) SVDD: window size (15,9). (e) SRD: window size (19,9). (f) CRD: window size (19,9). (g) LSMAD. (h) DL. (i) SDLCN.

TABLE 4. consumption of all the competitors on D1F12H1 (s).

Time	GRX	RSAD	LRX		SRD		CRD		LSMAD	DL	SDLCN
			$(25, 7)$	$(25, 9)$	$(25, 7)$	$(25, 9)$	$(25, 7)$	$(25, 9)$			
D1F12H1	27.14	26998.36	25390.99	26188.31	2333.57	2215.96	10835.04	10145.21	977.44	629.78	1118.33

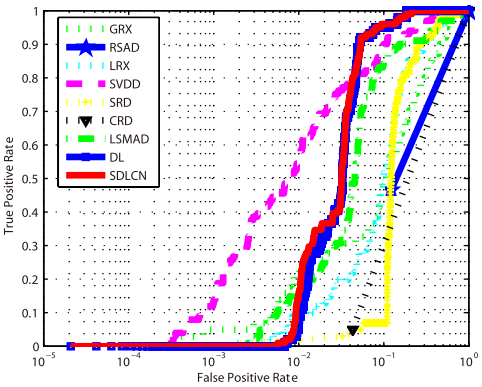


FIGURE 13. The ROC curves of the anomaly detection results by different algorithms on AVIRIS2.

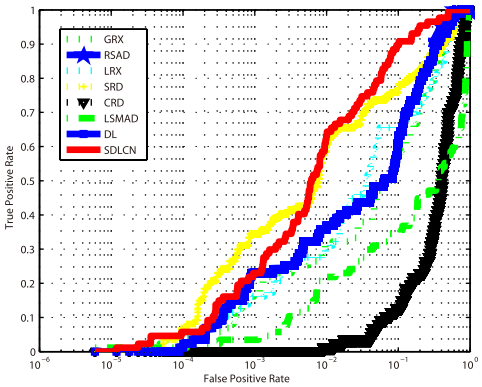


FIGURE 14. The ROC curves of the anomaly detection results by different algorithms on D1F12H1.

value 0.9555 with window size (15, 9), its time consumption is actually almost 11 hours as shown in Table 2. However, our method is much more efficient. To conclude, all the results and analyses have demonstrated the superiority of our proposed method. The SDLCN has the stronger distinctiveness to anomalies.

Fig. 15 illustrates the detection results on D1F12H1 image. It can be seen that RSAD and LSMAD have poorer performance because of a very high value of false alarm rate. The results of LRX and CRD are not intuitive to analyze because there are not many remarkable detected pixels through observing their visualization pictures. As for other methods, each of them almost detects all the abnormal targets, only assigning different intensities to each anomaly intuitively. The ROC curves of all the methods are shown in Fig. 14. We can observe that the curve of SDLCN keeps above the curves of all the other competitors except for SRD. When compared with SRD, our method also has the comparable and even better performance. It can be seen that when the false positive rate is in the range of approximately less than 10^{-4}, the proposed SDLCN obtains higher true positive rate than SRD does for each false positive rate. When the false positive rate continues increasing, the detection probability of SDLCN is slightly lower than that of SRD with the false positive rate changing in a small range at the beginning, then our ROC curve quickly goes up and gets the 100% detection rate more rapidly than SRD does. On the whole, the performance of our method is still outstanding. The highest AUC value achieved by SDLCN as shown in Table 3 further demonstrates its effectiveness. The good advantage of capped norm makes the learned dictionary have a good ability to represent background because the abnormal target is suppressed in the dictionary learning process. Consequently, the discrimination between anomaly and background is significantly enhanced.

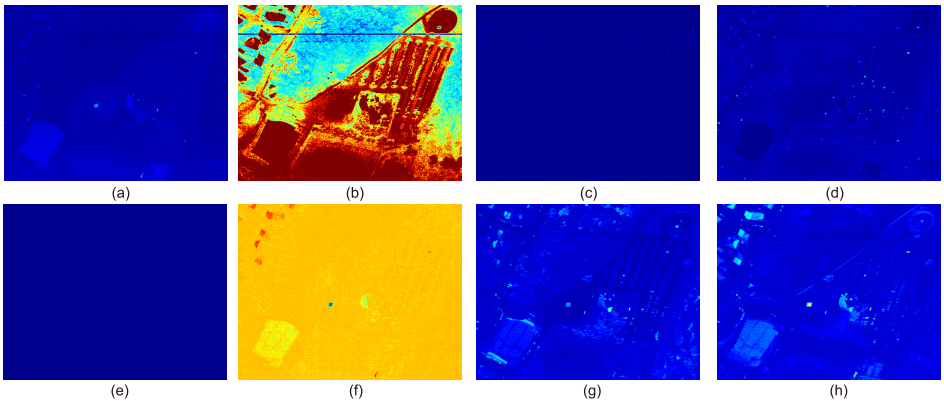


FIGURE 15. The visualization of the anomaly detection results of different detectors on D1F12H1. (a) GRX. (b) RSAD. (c) LRX: window size (25,9). (d) SRD: window size (25,9). (e) CRD: window size (25,9). (f) LSMAD. (g) DL. (h) SDLCN.

All the experimental results demonstrate the good performance of the proposed method. Compared with other state-of-the-art methods which may only perform well occasionally on one data set or be very sensitive to window sizes, the proposed method almost achieves the best detection results on all the four images, which shows its good adaptiveness to different hyperspectral image scenes and stronger capability to detect different kinds of anomalies. Besides, it is also efficient through the time consumption comparison shown in Table 2 and Table 4. It can be seen that GRX is the most efficient method, while SVDD is extremely inefficient. Compared with the methods using the sliding window technique, to a certain degree, our method still has the superiority in time consumption. On the whole, although the proposed method is not the most efficient detector, its time consumption is promising on all the data sets. To conclude, all the experimental results have demonstrated the superiority of the proposed SDLCN by using capped norm to restrain the effect of anomalies on background dictionary.

IV. CONCLUSION

In this paper, we propose a novel anomaly detection method via sparse dictionary learning of capped norm named SDLCN. In order to suppress the effect of potential anomalies on the background dictionary, the capped norm is employed to penalize the larger reconstruction errors during the dictionary learning process. Consequently, the learned background dictionary has a good expressive ability and is resistant to anomalies as well. In order to demonstrate the effectiveness of the proposed method fairly, several benchmark competitors are used to obtain the convincing comparison. Four commonly used real-word hyperspectral images are employed to test all the methods. Extensive experiments demonstrate the effectiveness of the proposed method and show its superiority over the benchmark methods. Without making any assumptions on the background distribution, our SDLCN method is more adaptive to all kinds of complex hyperspectral images. Using the Capped norm to restrain the effects of anomalies, our method can accurately characterize the background with stronger distinctiveness to anomalies through suppressing their effects. Moreover, it also shows a good ability to detect different sizes of abnormal targets. All of these desirable characteristics ensure that the proposed method has important application value.

REFERENCES

[1] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.

[2] A. Moghimi, C. Yang, and P. M. Marchetto, "Ensemble feature selection for plant phenotyping: A journey from hyperspectral to multispectral imaging," *IEEE Access*, vol. 6, pp. 56870–56884, 2018.

[3] T. Wang, H. Zhang, H. Lin, and X. Jia, "A sparse representation method for *a priori* target signature optimization in hyperspectral target detection," *IEEE Access*, vol. 6, pp. 3408–3424, 2017.

[4] J. Du and Z. Li, "A hyperspectral target detection framework with subtraction pixel pair features," *IEEE Access*, vol. 6, pp. 45562–45577, 2018.

[5] S. Jia, L. Shen, J. Zhu, and Q. Li, "A 3-D Gabor phase-based coding and matching framework for hyperspectral imagery classification," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1176–1188, Apr. 2018.

[6] H. Yuan and Y. Y. Tang, "Spectral–spatial shared linear regression for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 934–945, Apr. 2017.

[7] Q. Wang, J. Wan, F. Nie, B. Liu, C. Yan, and X. Li, "Hierarchical feature selection for random projection," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2018.2868836.

[8] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.

[9] Y. Zhou, A. Rangarajan, and P. D. Gader, "A Gaussian mixture model representation of endmember variability in hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2242–2256, May 2018.

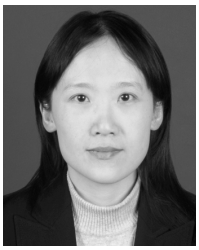
[10] B. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, "Hyperspectral unmixing using a neural network autoencoder," *IEEE Access*, vol. 6, pp. 25646–25656, 2018.

[11] 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.

[12] C. M. Gevaert, J. Suomalainen, J. Tang, and L. Kooistra, "Generation of spectral–temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for precision agriculture applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3140–3146, Jun. 2015.

[13] 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.

[14] 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.
[15] 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.
[16] 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.
[17] M. Díaz, R. Guerra, S. López, and R. Sarmiento, "An algorithm for an accurate detection of anomalies in hyperspectral images with a low computational complexity," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1159–1176, Feb. 2018.
[18] 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.
[19] Q. Wang, Z. Qin, F. Nie, and X. Li, "Spectral embedded adaptive neighbors clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2018.2861209.
[20] 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.
[21] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3123–3134, Dec. 2016.
[22] 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.
[23] 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.
[24] 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.
[25] 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.
[26] 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.
[27] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.
[28] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
[29] 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.
[30] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.
[31] J. Li, H. Zhang, and L. Zhang, "Background joint sparse representation for hyperspectral image subpixel anomaly detection," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 1528–1531.
[32] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.
[33] Y. Zhang, B. Du, L. Zhang, and T. Liu, "Joint sparse representation and multitask learning for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 894–906, Feb. 2017.
[34] L. Wei and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
[35] W. Jiang, F. Nie, and H. Huang, "Robust dictionary learning with capped l1-norm," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, Jul. 2015, pp. 3590–3596.
[36] D. Ma, Y. Yuan, and Q. Wang, "A sparse dictionary learning method for hyperspectral anomaly detection with capped norm," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 648–651.
[37] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Jun. 2009, pp. 689–696.
[38] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Berkeley Symp. Math. Statist. Probab.*, 1967, pp. 281–297.
[39] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting coherent groups in crowd scenes by multiview clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2018.2875002.
[40] 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.
[41] L. Ma, M. M. Crawford, and J. Tian, "Anomaly detection for hyperspectral images based on robust locally linear embedding," *J. Infr., Millim., Terahertz Waves*, vol. 31, no. 6, pp. 753–762, Jun. 2010.
[42] D. Ma, Y. Yuan, and Q. Wang, "Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation," *Remote Sens.*, vol. 10, no. 5, p. 745, 2018.
[43] N. Acito, S. Matteoli, A. Rossi, M. Diani, and G. Corsini, "Hyperspectral airborne 'Viareggio 2013 trial' data collection for detection algorithm assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2365–2376, Jun. 2016.



YUAN YUAN (M'05–SM'09) is currently a Full Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. She has authored or co-authored over 150 papers, including about 100 papers in reputable journals, such as the IEEE TRANSACTIONS and *Pattern Recognition*, and 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 intelligent science and technology from Xidian University, Xi'an, China, in 2013, and the M.S. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2016. She is currently pursuing the Ph.D. degree with the Center for Optical Imagery Analysis and Learning, 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 systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.