A SPARSE DICTIONARY LEARNING METHOD FOR HYPERSPECTRAL ANOMALY DETECTION WITH CAPPED NORM

Dandan Ma[1,2], *Yuan Yuan*[1], *Qi Wang*[3*],

[1]Xi'an Institute of Optics and Precision Mechanics of CAS, [2]University of Chinese Academy of Sciences

[3]School of Computer Science and Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University

[*]Corresponding author:crabwq@nwpu.edu.cn

ABSTRACT

Hyperspectral anomaly detection is playing an important role in remote sensing field. Most conventional detectors based on the Reed-Xiaoli (RX) method assume the background signature obeys a Gaussian distribution. However, it is definitely hard to be satisfied in practice. Moreover, background statistics is susceptible to contamination of anomalies in the processing windows, which may lead to many false alarms and sensitiveness to the size of windows. To solve these problems, a novel sparse dictionary learning hyperspectral anomaly detection method with capped norm constraint is proposed. Contributions are claimed in threefold: 1) requiring no assumptions on the background distribution makes the method more adaptive to different scenes; 2) benefiting from the capped norm our method has a stronger distinctiveness to anomalies; and 3) it also has better adaptability to detect different sizes of anomalies without using the sliding dual window. The extensive experimental results demonstrate the desirable performance of our method.

***Index Terms*—** Anomaly detection, Hyperspectral images, Sparse, Dictionary learning, Capped norm

1. INTRODUCTION

Hyperspectral image is a 3-D cube delivering rich information containing both on the spectral and spatial dimensions [1]. Hyperspectral anomaly detection is a kind of unsupervised target detection problem, where the prior information about both the background and the target is unknown [2]. It locates a hyperspectral pixel as an anomaly according to its significant spectrum deviation from its surrounding background or a given reference background. A large number of anomaly detection methods have been proposed in recent decades. The core idea behind these detectors is to acquire the background information and further estimate the difference between the anomalies and the background by defining some certain measurements in order to distinguish them. The Reed-Xiaoli (RX) [3] method is the most popular and typical detector, which assumes that the background conforms to a same multivariate normal distribution and the Mahalanobis distance is used to estimate the difference between the pixel under test and its reference background.

However, RX method has some inherent problems which may make it suffer from the high false alarms [4]. The simple distribution hypothesis is inaccurate to describe the real complex image scene. Besides, in the practical processing, the background statistics is calculated by all the pixels from a local region or a global image where the anomaly targets may exist. Therefore, the Mahalanobis distance involving the background mean and covariance matrix is susceptible to the contamination of anomaly targets. Although a sliding dual window [5] is usually used to alleviate the effects of anomaly as far as possible, it cannot completely eliminate all the potential anomalies from the designated background region, which also leads to the sensitiveness to the size of windows.

In order to restrain the abnormal target's effect on the estimation of background statistics, Du *et al.* proposed a RSAD method [6], which adopted a random selection process to pick out some representative background pixels from the hyperspectral image. Support vector data description (SVDD) [7] is also a type of method aiming at alleviating the contamination of anomalies. The hypothesis of SVDD is that the background is enveloped by an enclosing hypersphere in a high-dimensional feature space, and the pixels who fall outside the hypersphere are treated as anomalies.

In this paper, we propose a novel Sparse Dictionary Learning with Capped Norm (SDLCN) hyperspectral anomaly detection method to overcome these problems of RX based methods. The SDLCN has some remarkable advantages in threefold: neither requiring no assumptions on the background distribution nor estimating the covariance matrix, effectively restraining the effect of anomalies on the accurate dictionary learning, and being adaptive to detect different sizes of abnormal targets without using the sliding dual window.

The rest of this paper is organized as follows. In Section 2, the proposed SDLCN method is described in detail. In

This work was supported in part by the National Basic Research Program of China (Youth 973 Program) under Grant 2013CB336500, in part by the State Key Program of National Natural Science of China under Grant 60632018 and Grant 61232010, and in part by the National Natural Science Foundation of China under Grant 61379094.

Section 3, extensive experiments conducting on both the simulated and real hyperspectral images are reported. Finally, conclusions are drawn in Section 4.

2. SDLCN

In this paper, we propose a novel sparse dictionary learning with capped norm anomaly detection method named SDLCN. It is designed based on these two facts : 1) the background pixels are similar to their local surroundings in a hyperspectral image, and 2) the classes of background landscape covered by the hyperspectral image are generally limited. Therefore, there is a large amount of redundant information in the background pixels [8], and a dictionary which is able to well describe the background knowledge can be learned definitely. By this way, the background pixel can be sparsely represented by the learned background dictionary, while the anomaly target is not. However, to learn this dictionary, a direct approach is to select a large amount of image patches from the entire hyperspectral image to construct the training data. Despite the fact that the population of anomaly is very less and its selected probability is also low, and consequently the training data can be completely approximated as background, the anomaly is unavoidably included in the training set. It will affect the representation ability of dictionary and further reduce the performance of anomaly detection. Therefore, the capped norm [9] is imposed on the learning process, which plays the role of weight constraint to suppress the effect of anomalies.

2.1. Background Dictionary Learning

The key process is to learn a proper and accurate dictionary to represent the background knowledge. In the sparsity model[10], a background pixel is approximately regarded as a sparse linear combination of the basis elements from the learned dictionary. A 3-D hyperspectral image can be reshaped as a 2-D representation $\mathbf{I} \in R^{mn \times B}$, where m and n respectively denotes the height and width of the original image, mn denotes the total number of image pixels, and B is the number of bands. We randomly select a number of image patches $\mathbf{H}_i = [\mathbf{h}_1, ..., \mathbf{h}_N] \in R^{B \times N}$ with same size from the entire image to construct the training data, and each column of $\mathbf{H}_i$ represents a hyperspectral pixel. We stack all the selected patches from the scene as background training matrix $\mathbf{H} = [\mathbf{H}_1, ..., \mathbf{H}_i, ..., \mathbf{H}_c] \in R^{B \times cN}$ $(i = 1, ..., c)$, where c denotes the number of image patches. Therefore, the traditional sparse dictionary learning [11] process can be expressed as:

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

In the above formulation, $\mathbf{A} \in R^{K \times cN}$ is the sparse coefficient matrix, $\mathbf{D} \in R^{B \times K}$ is the background dictionary matrix containing K basis elements to learn, and λ is a balance parameter. In general, each column of $\mathbf{D}$ is constrained by its corresponding l_2 norm not more than 1 in order to avoid the l_2 norm of $\mathbf{D}$ being extremely larger and the elements of $\mathbf{A}$ consequently being arbitrarily small.

For the anomaly detection, we are aiming at learning a high quality dictionary with good ability to express the background. However, as we all know, the above dictionary learning method is sensitive to anomalies for using the quadratic loss function. The quality of dictionary learning will be greatly influenced and degenerated when the anomalies exist. Although the population of anomaly is very less, it is still impossible to ensure the inexistence of anomaly in the training data. Consequently, performance of anomaly detection will be greatly influenced. To this end, the capped l_1 norm based loss function is used to penalize the anomalies in the training set, which will learn a better dictionary resistant to anomalies.

The capped l_1 norm based loss function is defined as:

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

where r_i denotes the residual error of the i-th test sample with its corresponding estimated value, and ε is a boundary parameter. When $|r_i|$ is larger than ε, the loss function will be equal to the constant ε. It means that a test sample with large error will be definitely penalized by capped l_1 norm. Therefore, the new objective function of sparse dictionary learning with capped l_1 norm can be formulated as:

$$\min_{\mathbf{D},\mathbf{A},\|\mathbf{d}_i\| \le 1} \sum_{i=1}^{cN} \min(\|\mathbf{h}_i - \mathbf{D}\mathbf{a}_i\|_2, \varepsilon) + \lambda \|\mathbf{A}\|_1. \tag{3}$$

It is obvious that our objective function consists of a concave function (the first term) and a convex function (the second term). A re-weighted technique can be used to solve this problem by introducing the auxiliary variables s_i. Therefore, $\mathbf{D}$, $\mathbf{A}$ and s_i can be iteratively updated according to the following rules:

$$[\mathbf{D}, \mathbf{A}] = \arg \min_{\mathbf{D},\mathbf{A},\|\mathbf{d}_i\| \le 1} \sum_{i=1}^{cN} s_i \|\mathbf{h}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \lambda \|\mathbf{A}\|_1, \tag{4}$$

$$s_i = \begin{cases} \frac{1}{2\|\mathbf{h}_i - \mathbf{D}\mathbf{a}_i\|_2}, & if \ \|\mathbf{h}_i - \mathbf{D}\mathbf{a}_i\|_2 \le \varepsilon \\ 0, & otherwise \end{cases}. \tag{5}$$

Note that the formulation (4) is a weighted dictionary learning model, which is similar to the classical dictionary learning. Each of the samples is assigned a weight to change the importance of different samples. Specifically, a sample with lower sparse reconstruction error will be given a larger weight, which makes the dictionary learning insensitive to potential anomalies.

We adopt an iterative update strategy to solve formulation (4). When $\mathbf{A}$ is fixed, the above problem becomes:

$$\min_{\mathbf{D},\|\mathbf{d}_i\| \le 1} \sum_{i=1}^{cN} s_i \|\mathbf{h}_i - \mathbf{D}\mathbf{a}_i\|_2^2. \tag{6}$$

Algorithm 1 SDLCN

Input: $\mathbf{H}$, K, λ and ε.
Initialize: $\mathbf{D}$, $\mathbf{A}$, and $s_i = 1$ for $i = 1, ..., cN$.
Repeat
1. Solving the weighted dictionary learning model (4)
 Repeat
 a. Update $\mathbf{D}$: $\mathbf{P} = [\mathbf{p}_1, ..., \mathbf{p}_K] \in R^{K\times K} = \sum_{i=1}^{cN} s_i \mathbf{a}_i \mathbf{a}_i^T$,
 $\mathbf{Q} = [\mathbf{q}_1, ..., \mathbf{q}_K] \in R^{B\times K} = \sum_{i=1}^{cN} s_i \mathbf{h}_i \mathbf{a}_i^T$.
 Repeat
 for $j = 1$ **to** K **do**
 $\mathbf{u}_j \leftarrow \frac{1}{\mathbf{P}_{jj}} (\mathbf{q}_j - \mathbf{D}\mathbf{p}_j) + \mathbf{d}_j$
 $\mathbf{d}_j = \frac{1}{\max(\|\mathbf{u}_j\|_2, 1)} \mathbf{u}_j$
 end for
 Until *Converge*
 b. Update $\mathbf{A}$: Update $\mathbf{a_i}$ for $i = 1, ..., cN$ as
 if $s_i > 0$ $\mathbf{a}_i = \arg\min_{\mathbf{a}} \|\mathbf{h}_i - \mathbf{D}\mathbf{a}\|_2^2 + \frac{\lambda}{s_i}\|\mathbf{a}\|_1$
 else $\mathbf{a}_i = \mathbf{0}$
Until *Converge*
2. Update s_i for $i = 1, ..., cN$ as formulation (5)
Until Converge
Output: $\mathbf{D}$ and $\mathbf{A}$

It can be further written as:

$$\min_{\mathbf{D}, \|\mathbf{d}_i\| \leq 1} \sum_{i=1}^{cN} \|\mathbf{x}_i - \mathbf{D}\mathbf{z}_i\|_2^2, \tag{7}$$

where $\mathbf{x}_i = \sqrt{s_i}\mathbf{h}_i$ and $\mathbf{z}_i = \sqrt{s_i}\mathbf{a}_i$. We adopt the same method proposed in [11] to solve this problem.

When $\mathbf{D}$ fixed, the formulation (4) will be decomposed into cN independent subproblems as follows:

$$\mathbf{a}_i = \arg\min_{\mathbf{a}} s_i \|\mathbf{h}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \lambda\|\mathbf{a}\|_1. \tag{8}$$

When $s_i \neq 0$, this problem can be further rewritten as:

$$\mathbf{a}_i = \arg\min_{\mathbf{a}} \|\mathbf{h}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \frac{\lambda}{s_i}\|\mathbf{a}\|_1. \tag{9}$$

Obviously, the formulation (9) is a Lasso problem, which can be solved efficiently. Alg. 1 gives the detailed iteration procedures to solve the the objective function (3).

2.2. Anomaly Detection

After finishing the learning of background dictionary $\mathbf{D}$, the final anomaly detection procedure will be implemented. The entire hyperspectral image matrix constructs the testing data ($\mathbf{H} = \mathbf{I}^T$). Considering the effectiveness, the formulation (1) with fixed $\mathbf{D}$ is used to detect the abnormal targets. Through computing the representation coefficient matrix $\mathbf{A}$, the sparse reconstruction errors are further calculated, which can be regarded as the corresponding anomaly probability values.

Table 1. AUC values of all the competitors on all datasets.

AUC	Simulated scene	Salinas scene	HYDICE Urban
GRX	0.9937	0.8872	0.8906
RX	0.9814	0.9043	0.8762
RSAD	0.7500	0.7572	0.8782
SVDD	0.9356	0.8932	0.8552
SDLCN	**1.0000**	**0.9328**	**0.9422**

3. EXPERIMENTAL RESULTS

We conduct several experiments on three hyperspectral images containing both simulated and real datasets to evaluate the effectiveness of the proposed method. The simulated scene data [4] has the size of $200 \times 200 \times 105$, which contains 12 anomaly targets with different sizes. Salinas scene is $180 \times 180 \times 224$, which is cropped from the original downloaded image according to the literature [12]. The HYDICE Urban Data Set is also a sub-image with the size of $80 \times 100 \times 160$ [4], which contains several cars and roofs regarded as anomalies. In order to demonstrate the superiority of the proposed SDLCN method, four kinds of benchmarks containing GRX, LRX, RSAD, and SVDD are implemented as the competitors. The ROC curve and AUC value are the estimation criteria.

Fig. 1 shows the visualization of the anomaly detection results. It can be seen obviously that our SDLCN is significantly superior to the other competitors for its good effectiveness to recognize anomaly target. Comparing with other competitors which only perform well on some partial data, the SDLCN is capable of detecting all the abnormal targets with a high intensity for all the different images. The corresponding quantitative results ROC curves and AUC values are shown in Fig. 2 and Table 1 respectively. All the ROC curves of SDLCN keep staying over the other curves and it also achieves the highest AUC values on all the datasets.

These phenomena verify that the SDLCN is adaptive to different image scenes for its requiring no assumptions on the image distribution. It can successfully distinguish anomalies from the background. Benefiting from the capped norm, the learned dictionary has a good ability to express the background which is resistant to the anomalies in the learning process, and consequently it will have a stronger distinctiveness to anomalies. Moreover, it is worth noting that the abnormal targets have different sizes in the experiments and SDLCN is adaptive to all the sizes for no need of sliding window technique. All of the experimental results fully confirm that the ability of the proposed SDLCN to detect anomaly is indeed excellent.

4. CONCLUSION

This paper proposes a novel hyperspectral anomaly detection method named SDLCN based on dictionary learning and sparse representation. Considering the effects of potential

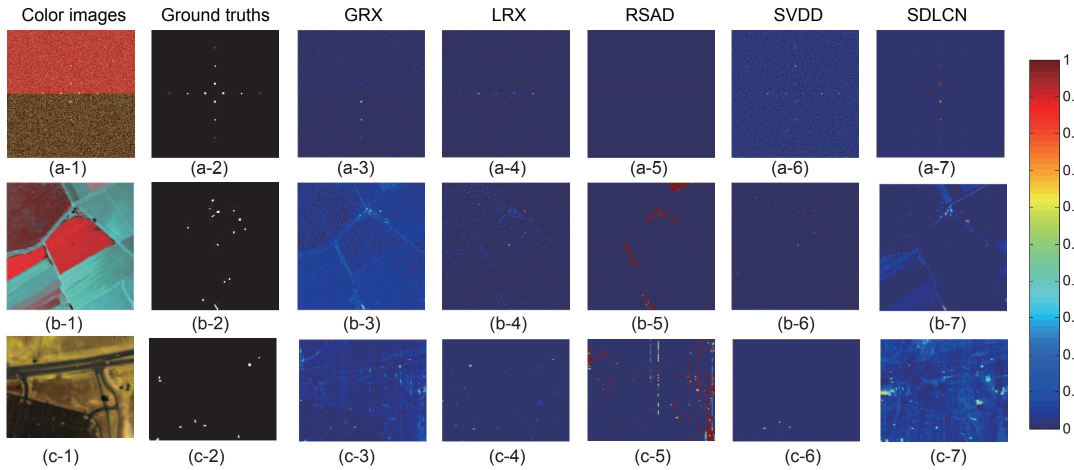
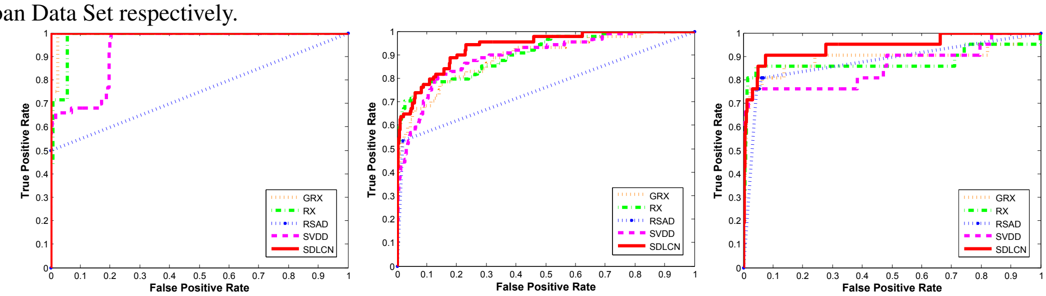


Fig. 1. The false color pictures, ground truths and detection results. Each row corresponds to the Simulated scene, the Salinas scene, and the HYDICE Urban Data Set respectively.



(a) Simulated scene (b) Salinas scene (c) HYDICE Urban Data Set

Fig. 2. The ROC curves of different detection methods on three experimental datasets.

anomalies contained in the training datasets on the expressive ability of the learned background dictionary, the capped norm is employed to suppress these effects and make the dictionary resistant to anomalies. The experimental results show that SDLCN is more superior than the state-of-the-art benchmarks. Our SDLCN method not only has good adaptability to different data scenes, but also possesses the stronger distinctiveness to anomalies from background. In addition, it also has the better ability to detect different sizes of abnormal targets.

5. REFERENCES

[1] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. and Learning Syst.*, vol. 27, no. 6, pp. 1279–1289, 2016.

[2] F. Li, Y. Zhang, L. Zhang, X. Zhang, and D. Jiang, "Hyperspectral anomaly detection using background learning and structured sparse representation," in *IEEE Int. Geosci. and Remote Sens. Symp.*, 2016, pp. 1618–1621.

[3] I. 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, 1990.

[4] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cyber.*, vol. 46, no. 12, pp. 3123–3134, 2016.

[5] 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, 2014.

[6] 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, 2011.

[7] 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, 2006.

[8] 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. 3140–3146, 2015.

[9] W. Jiang, F. Nie, and H. Huang, "Robust dictionary learning with capped l1-norm," in *Proc. Int. Joint Conf. AI*, 2015, pp. 3590–3596.

[10] X. Peng, S. Xiao, J. Feng, W. Yau, and Z. Yi, "Deep subspace clustering with sparsity prior," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 1925–1931.

[11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Int. Conf. Mach. Learning*, 2009, pp. 689–696.

[12] Y. Yuan, Q. Wang, and G. Zhu, "Fast hyperspectral anomaly detection via high-order 2-d crossing filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 620–630, 2015.