Hyperspectral Anomaly Detection Based on Low-Rank Structure Exploration

Shizhen Chang[1], Pedram Ghamisi[1,2]

[1]*Institute of Advanced Research in Artificial Intelligence (IARAI), Landstraßer Hauptstraße 5, 1030 Vienna, Austria.*

[2]*Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, Chemnitzer Str. 40, D-09599 Freiberg, Germany*

Abstract

As one of the typical research area in unsupervised hyperspectral image learning, anomaly detection needs to accomplish the abnormal pixels separation process without prior spectral knowledge. Recently, the representation-based detectors which can find the spectral similarity between pixels under no statistical distribution assumption have attracted extensive attention and been frequently used. To this end, low-rank regularization methods can approximately decompose the hyperspectral data into a low-rank background part and a sparse anomaly part. Based on the theory of representation and self-representation, this paper proposed a double low-rank regularization (DLRR) model for hyperspectral anomaly detection. To further explore the reconstructed structure differences between the original data and the assumed background, the residual of their corresponding low-rank coefficient matrices are computed and utilized as a part of the detection output together with the column-wise ℓ_2 norm of the sparse matrix. Experiments carried out on two real-world hyperspectral datasets show promising performances compared with other state-of-the-art detectors.

Keywords

Hyperspectral imagery, anomaly detection, low-rank representation, sketched-subspace clustering

1. Introduction

With continuous and redundant spectral bands, hyperspectral images (HSIs) carry a wealth of spectral and spatial information of land-covers [1, 2]. This promotes military and civilian applications utilizing the spectral characteristics of different materials. And lots of research works have been conducted, such as feature extraction [3], noise reduction [4], unmixing [5], classification [6], and detection [7], etc. As a special branch of HSIs researches, anomaly detection aims to extract potential abnormal pixels without any prior knowledge [8]. Therefore, suitable methods need to be designed.

Traditionally, classic anomaly detectors have been developed mostly based on the assumption of data statistical distributions. The benchmark RX detector, the cluster-based anomaly detection (CBAD) [7] algorithm, the blocked adaptive computationally efficient outlier nominators (BACON) [9] and the random selection-based anomaly detector (RSAD) [10] assume that the data follows the Gaussian or Gaussian mixture distributions, then they implement the detection task according to the Mahalanobis distance between the pixel-under-test and the background. However, this hypothesis has obvious

limitations in practical applications.

To overcome the insufficient accuracy happened in the distribution-based models, the representation-based detectors have been proposed and shown intended performances. Representative methods are the collaborative representation detector (CRD) [11], the background joint sparse representation detector (BJSRD) [12], etc. A dual concentrate window is utilized to extract the possible background information as the confidence dictionaries of the background at each test pixel. And the detection result is approximately derived by calculating the representation residual of the pixel. Nowadays, the low-rank representation is widely used in hyperspectral anomaly detection which takes advantage of the repeatability of the background spectrum and decomposes the original data matrix. Considering that the anomalies are usually rare and sparse, Chen et. al. [13] first utilized the low-rank decomposition model for anomaly detection. Later, many types of research have been carried out based on the low-rankness of the background subspace and the sparsity of the anomaly subspace [14, 15]. However, after doing the low-rank decomposition, the detection decision of these methods either focus on analyzing the sparse matrix or back to the statistical estimation, a better combination of the assumed background component and anomaly component may let the detection more reasonable.

As is well known, subspace clustering (SC) is gradually developed for unsupervised HSIs interpretation [16]. It can learn the similarity between pixels through self-dictionary learning. Inspired by the sketched-SC and representation theory, we propose a double low-rank

regularization model for hyperspectral anomaly detection. The proposed method is solved via the alternating direction method of multipliers (ADMMs) method [17] and the anomalies are finally detected by computing a background-related residual matrix and an anomaly-related distance matrix.

2. Proposed method

2.1. Related works

Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N \in \mathbb{R}^{L \times N}$ denotes the data collections, where L is the number of bands and N is the total pixel number of the image. Then, the low-rank representation model aims to decompose the data into a lowest-rank matrix $\mathbf{L} \in \mathbb{R}^{L \times N}$ and a sparse matrix $\mathbf{E} \in \mathbb{R}^{L \times N}$. The optimization problem is given by:

$$\min_{\mathbf{L},\mathbf{E}} ||\mathbf{L}||_\star + \alpha||\mathbf{E}||_0 \quad s.t. \ \mathbf{X} = \mathbf{L} + \mathbf{E}, \qquad (1)$$

where α is the regularization parameter, $|| \cdot ||_\star$ and $|| \cdot ||_0$ represents the nuclear norm and the ℓ_0 norm, respectively.

For a given dictionary D, the low-rank matrix $\mathbf{L}$ can be rewritten as a linear combination of the D and its corresponding coefficient matrix Z. And the NP-hard problem eq. (1) can be represented as:

$$\min_{Z,\mathbf{E}} ||Z||_\star + \alpha||\mathbf{E}||_{2,1} \quad s.t. \ \mathbf{X} = DZ + \mathbf{E}, \quad (2)$$

where $D \in \mathbb{R}^{L \times n}$ has n dictionary samples, $Z \in \mathbb{R}^{n \times N}$, and $|| \cdot ||_{2,1}$ is the $\ell_{2,1}$ norm of the matrix. $\ell_{2,1}$ norm can be regarded as the ℓ_1 norm of the ℓ_2 norm of matrix columns.

2.2. Problem formulation

Assume that the data can be self-represented:

$$\mathbf{X} = \mathbf{X}C,$$

where C is the coefficient matrix. Then for a sketched data $\tilde{\mathbf{X}} = \mathbf{X}T$, we also have:

$$\mathbf{X} = \tilde{\mathbf{X}}A,$$

where $T \in \mathbb{R}^{N \times t}$ is defined as a random projection matrix to compress $\mathbf{X}$ while preserving its main information, and $A \in \mathbb{R}^{t \times N}$.

By means of the sketched data $\tilde{\mathbf{X}}$, the proposed double low-rank regularization (DLRR) model based on sketched-SC can be formulated as:

$$\min_{A,Z,\mathbf{E}} ||A||_\star + \beta||Z||_\star + \lambda||\mathbf{E}||_{2,1}$$
$$s.t. \ \ \mathbf{X} = \tilde{\mathbf{X}}A, \ \mathbf{X} = DZ + \mathbf{E}, \qquad (3)$$

where β and λ are the regularization parameters, A is the self-representation coefficient matrix, Z is the assumed background coefficient matrix, and $\mathbf{E}$ denotes the sparse part indicating the anomalies.

Considering that the difference of A and Z may reflect the abnormal information of the image, then the residual of these two matrices can be utilized as the reference of detection output when $t = n$. And the final detection result is formulated by the sum of the column-wise ℓ_2 norm of the low-rank coefficient matrices residual and the sparse coefficient matrix:

$$D_{DLRR}(\mathbf{x}_i) = ||A_{:,i} - Z_{:,i}||_2 + ||E_{:,i}||_2. \qquad (4)$$

2.3. Problem optimization

To solve the proposed DLRR model, the ADMM method is employed and the detailed optimization process is described as follows.

First, B and H are introduced as the auxiliary variables for the coefficient matrix A and Z, respectively:

$$\min_{B,A,H,Z,\mathbf{E}} ||B||_\star + \beta||H||_\star + \lambda||\mathbf{E}||_{2,1}$$
$$s.t. \quad \mathbf{X} = \tilde{\mathbf{X}}A, \ \mathbf{X} = DZ + \mathbf{E}, \qquad (5)$$
$$A = B, \ Z = H.$$

Then, the augmented Lagrangian function of (5) can be constructed:

$$\min_{B,A,H,Z,\mathbf{E},Y_1,Y_2,Y_3,Y_4} ||B||_\star + \beta||H||_\star + \lambda||\mathbf{E}||_{2,1}$$
$$+ \frac{\rho}{2}||\mathbf{X} - \tilde{\mathbf{X}}A + Y_1/\rho||_F^2 + \frac{\rho}{2}||\mathbf{X} - DZ - E + Y_2/\rho||_F^2$$
$$+ \frac{\rho}{2}||A - B + Y_3/\rho||_F^2 + \frac{\rho}{2}||Z - H + Y_4/\rho||_F^2, \tag{6}$$

where $Y_1 \in \mathbb{R}^{L \times N}$, $Y_2 \in \mathbb{R}^{L \times N}$, $Y_3 \in \mathbb{R}^{t \times N}$, and $Y_4 \in \mathbb{R}^{n \times N}$ are the Lagrangian multipliers, and $\rho > 0$ is the penalty parameter.

Then the equation (6) can be divided into five optimization problems and be updated one by one with iterative procedures. The updating rules of these variables are:

1) B step with fixed A and Y_3:

$$\min_B ||B||_\star + \frac{\rho}{2}||A - B + Y_3/\rho||_F^2. \quad (7)$$

2) H step with fixed Z and Y_4:

$$\min_H \beta||H||_\star + \frac{\rho}{2}||Z - H^{(k+1)} + Y_4/\rho||_F^2. \ (8)$$

3) $\mathbf{E}$ step with fixed Z and Y_2:

$$\min_{\mathbf{E}} \lambda||\mathbf{E}||_{2,1} + \frac{\rho}{2}||\mathbf{X} - DZ - E + Y_2/\rho||_F^2. \ (9)$$

4) A step with fixed B, Y_1 and Y_3:

$$\min_A \frac{\rho}{2}||\mathbf{X}-\tilde{\mathbf{X}}A+Y_1/\rho||_F^2+\frac{\rho}{2}||A-B+Y_3/\rho||_F^2. \tag{10}$$

5) Z step with fixed H, $\mathbf{E}$, Y_2 and Y_4:

$$\min_Z \frac{\rho}{2}||\mathbf{X}-DZ-E+Y_2/\rho||_F^2+\frac{\rho}{2}||Z-H+Y_4/\rho||_F^2. \tag{11}$$

6) The Lagrangian multipliers and the penalty parameter are updated as:

$$Y_1 = Y_1 + \rho(\mathbf{X} - \tilde{\mathbf{X}}A), \tag{12}$$

$$Y_2 + \rho(\mathbf{X} - DZ - E), \tag{13}$$

$$Y_3 = Y_3 + \rho(A - B), \tag{14}$$

$$Y_4 = Y_4 + \rho(Z - H), \tag{15}$$

$$\rho = \min\{1.1\rho, \rho_{max}\}. \tag{16}$$

The solutions of (7) and (8) are calculated by $B = \Theta_{(1/\rho)}(A + Y_3/\rho)$ and $H = \Theta_{(\beta/\rho)}(Z + Y_4/\rho)$, respectively, where Θ is the singular value thresholding (SVT) operator. Then $\mathbf{E}$ is updated by $\mathcal{S}_{(\lambda/\rho)}(\mathbf{X} - DZ + Y_2/\rho)$ where $\mathcal{S}$ is a $\ell_{2,1}$-min thresholding operator [18]. A and Z are respectively solved by finding the partial derivative and setting it to zero. Their optimized solutions are $A = (I + \tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}^\top\mathbf{X} + \tilde{\mathbf{X}}^\top Y_1/\rho + B - Y_3/\rho)$ and $Z = (I + D^\top D)^{-1}(D^\top\mathbf{X} - D^\top\mathbf{E} + D^\top Y_2/\rho + H - Y_4/\rho)$.

The initial settings of this optimization process are: $A_0 = Z_0 = H_0 = 0$, $\mathbf{E}_0 = 0$, $Y_0 = Y_1 = 0$, $Y_3 = Y_4 = 0$, $\rho_0 = 0.01$, $\rho_{max} = 10^6$. And the convergence conditions are $||X - \tilde{\mathbf{X}}A||_F < \epsilon$, $||X - DZ + E||_F < \epsilon$, $||A - B||_F < \epsilon$, $||Z - H||_F < \epsilon$, or the iteration times exceeds the predefined upper limit. Empirically, the predefined value of the error tolerance is $\epsilon = 10^{-6}$ and the maximum iteration time is 100.

3. Experiments

In this section, the performance of the proposed DLRR method is assessed on two real-world HSI scenes: the San Diego dataset and the Urban dataset. Four classical anomaly detection methods, which are RX, BACON, Kernel-RX (KRX) [19], and the low probability anomaly detector (LPAD) [20], respectively, are applied for comparable analysis. The regularization parameters β and λ are set as 5 and 10, respectively. The background dictionaries are collected by the mean vector of the K-means clusters, and the dictionary number n is set as 400.

1) *San Diego dataset*: This dataset was captured by the AVIRIS sensor, which has a spatial resolution of 3.5 m and a spectral resolution of 10 nm. This dataset has 224 original spectral bands in total,
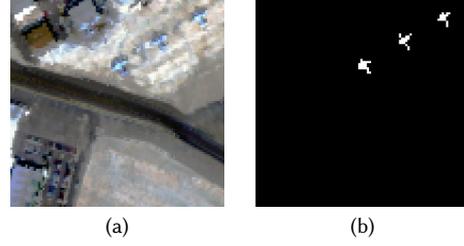
Figure 1: The San Diego dataset. (a) Image scene. (b) Ground-truth.
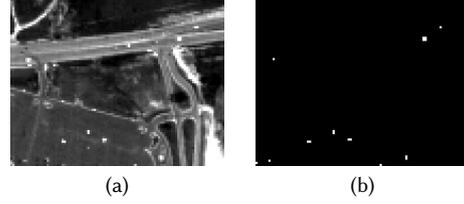
Figure 2: The Urban dataset. (a) Image scene. (b) Ground-truth.

and 189 bands are utilized for the detection task after eliminating the noisy bands. It records the area of the San Diego airport, CA, USA in 100×100 pixels, three aircrafts including 58 pixels are selected as the anomaly target. The visualized 2-D image scene and the ground-truth map of this dataset are shown in Figure 1.

2) *Urban dataset*: This dataset was collected by the HYDICE airborne sensor, which has a spatial resolution of 1 m and a spectral resolution of 10 nm. After removing low quality bands, 162 bands are left for anomaly detection. This image scene contains 80×100 pixels, and 17 small objects are considered as anomaly targets. The 2-D visualization map and the ground-truth of this dataset are shown in Figure 2.
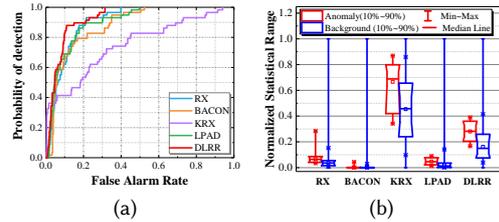
Figure 3: The San Diego dataset. (a) ROC curves. (b) Normalized background-anomaly statistical range.

The proposed DLRR model together with other comparable algorithms are conducted in the aforementioned

San Diego and Urban datasets, and the detection performances evaluated by the ROC curves and the normalized background-anomaly statistical range are shown in Figure 3 and Figure 4, respectively. It can be seen that in the San Diego dataset, the proposed DLRR model has the smallest false alarm rate when the detection probability reaches 1. And for the Urban dataset, our method has the largest separation range between the backgrounds and the anomalies. For further evaluation, the area under ROC curve (AUC) values are computed and the results are shown in Table 1. The results show that the proposed DLRR method has the largest AUC values compared with other methods in both two datasets.
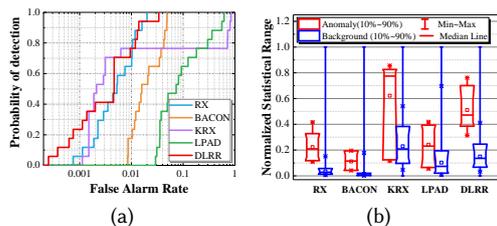
Figure 4: The Urban dataset. (a) ROC curves. (b) Normalized background-anomaly statistical range.

Table 1
The AUC Values of Five Algorithms in Two Datasets

Methods	San Diego Dataset	Urban Dataset
RX	0.8742	0.9919
BACON	0.8768	0.9720
KRX	0.7490	0.7836
LPAD	0.8973	0.8137
DLRR	**0.9261**	**0.9927**

References

[1] P. Ghamisi, M. Dalla Mura, J. A. Benediktsson, A survey on spectral–spatial classification techniques based on attribute profiles, IEEE Trans. Geosci. Remote Sens. 53 (2015) 2335–2353.

[2] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, J. Chanussot, Graph convolutional networks for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. (2020).

[3] Y. Xu, L. Zhang, B. Du, F. Zhang, Spectral–spatial unified networks for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 56 (2018) 5893–5909.

[4] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, J. Chanussot, Noise reduction in hyperspectral imagery: Overview and application, Remote Sens. 10 (2018) 482.

[5] R. Heylen, M. Parente, P. Gader, A review of nonlinear hyperspectral unmixing methods, IEEE J. Sel. Topics Appl. Earth Observa. Remote Sens. 7 (2014) 1844–1868.

[6] Y. Xu, B. Du, L. Zhang, Beyond the patchwise classification: Spectral-spatial fully convolutional networks for hyperspectral image classification, IEEE Trans. Big Data. 6 (2020) 492–506.

[7] M. J. Carlotto, A cluster-based approach for detecting man-made objects and changes in imagery, IEEE Trans. Geosci. Remote Sens. 43 (2005) 374–387.

[8] S. Chang, B. Du, L. Zhang, A subspace selection-based discriminative forest method for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 58 (2020) 4033–4046.

[9] N. Billor, A. S. Hadi, P. F. Velleman, BACON: blocked adaptive computationally efficient outlier nominators, Comput. Statist. Data Anal. 34 (2000) 279–298.

[10] B. Du, L. Zhang, Random-selection-based anomaly detector for hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 49 (2010) 1578–1589.

[11] W. Li, Q. Du, Collaborative representation for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 53 (2014) 1463–1474.

[12] J. Li, H. Zhang, L. Zhang, L. Ma, Hyperspectral anomaly detection by the use of background joint sparse representation, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 8 (2015) 2523–2533.

[13] S.-Y. Chen, S. Yang, K. Kalpakis, C.-I. Chang, Low-rank decomposition-based anomaly detection, in: Proc. SPIE, volume 8743, Art, 2013, p. 87430N.

[14] Y. Zhang, B. Du, L. Zhang, S. Wang, A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 54 (2015) 1376–1389.

[15] L. Li, W. Li, Q. Du, R. Tao, Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection, IEEE Trans. Cybern. (2020).

[16] H. Zhai, H. Zhang, L. Zhang, P. Li, Nonlocal means regularized sketched reweighted sparse and low-rank subspace clustering for large hyperspectral images, IEEE Trans. Geosci. Remote Sens. 59 (2020) 4164–4178.

[17] S. Boyd, N. Parikh, E. Chu, Distributed optimization and statistical learning via the alternating direction method of multipliers, Now Publishers Inc, 2011.

[18] L. Zhang, L. Peng, T. Zhang, S. Cao, Z. Peng, Infrared small target detection via non-convex rank approximation minimization joint $\ell_{2,1}$ norm, Remote Sens. 10 (2018) 1821.

[19] H. Kwon, N. M. Nasrabadi, Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery, IEEE Trans. on Geosci. Remote Sens. 43 (2005) 388–397.

[20] Z. Li, L. Wang, S. Zheng, Applied low dimension linear manifold in hyperspectral imagery anomaly detection, in: Proc. SPIE, volume 9142, Art, 2014, p. 91421P.