2019 International Conference on Power, Energy, Environment and Material Science (PEEMS 2019)
ISBN: 978-1-60595-669-5

Anomaly Detection of Hyperspectral Images Based on Linear Unmixing and Low-Rank Representation

Chen-guang PAN[1,2], Ting-fa XU[1,2,*], Jian-hua HAO[1,2], A-xin FAN[1,2] and Chen HUANG[1,2]

[1]Key Laboratory of Photoelectronic Imaging Technology and System, School of Optoelectronics, Beijing Institute of Technology, Beijing, China

[2]Chongqing Innovation Center, Beijing Institute of Technology, Chongqing, China

*Corresponding author

Keywords: Hyperspectral anomaly detection, Kernel-PCA, Linear spectral unmixing, Low-rank representation, Dictionary learning.

Abstract. Anomaly detection is an important branch in hyperspectral detection. There are some challenging problems mainly due to the highly mixture nature and noise corruption. In this paper, we propose a hyperspectral anomaly detection method based on Linear Unmixing and Low-Rank Representation (LU-LRR) with construction of a more robust dictionary. Firstly, the Kernel-PCA (KPCA) is used to estimate the numbers of endmembers and mine high-order correlation among spectral bands of original hyperspectral data. Secondly, data are decomposed into a product form of endmember matrix and abundance matrix based on linear spectral unmixing method. Abundance matrix usually possesses more distinctive feature to distinguish anomaly from background compared with original data. Finally, LRR is exploited to decompose abundance matrix into low-rank component and sparse component indicating background and anomalies separately, and furtherly introduce a more robust subspace basis dictionary based on dictionary learning into low-rank component. Experiments in real hyperspectral images have demonstrated that LU-LRR is more effective than traditional anomaly detection methods.

Introduction

Due to the wide spectral response range, high spectral resolution and abundant spectral information, hyperspectral images are more suitable for target detection. Conventional hyperspectral detection methods are established on the basis of priori spectral information of targets, which can be obtained by online spectral libraries and actual measurements, and then detect pixels whose spectrum is similar to priori information. However, the acquisition of priori information is usually difficult so that anomaly detection, which extracts anomalous targets by spectral difference between targets and background without priori spectrum, plays a more and more significant role in practice. There is no absolutely explicit definition of anomaly, which can be approximately identified as small, sparse and dense regions and showing high contrast compared with background.

An important branch in anomaly detection is based on low-rank representation method. The most classic method is Robust Principal Component Analysis (RPCA)[1], which can be formed by adding an unknown sparse matrix to an unknown low rank matrix, and sparse matrix denotes anomaly. RPCA aims at detecting small heterogeneous and high contrast regions effectively[3]. Therefore, isolated noise pixels could be judged as anomalies which may cause increase in false alarm rate. GoDec algorithm[4] is proposed to alleviate this problem and decomposes hyperspectral data into three matrices: low rank matrix representing background, sparse matrix representing anomalous and noise matrix. However, anomalous objects tend to be judged as noise using this method. As a result, Liu *et al* [5] proposes a method named Low-Rank Representation (LRR) to seek the lowest-rank representation among all the candidates that represents all hyperspectral data as the linear combination of the bases in a dictionary to separate anomalies from background and isolated noise exactly. However, the dictionary in LRR is constructed by directly using the whole original

hyperspectral data as atoms. This way may cause some problems including expensive computation cost, sensitivity on the tradeoff parameter and anomaly contamination.

In addition, highly mixed nature of hyperspectral data also has influence on detecting precision. The existence of mixed pixels may cause serious error whether pixels including anomaly are classified as background or target. It is necessary to project hyperspectral data into a single subspace i.e. which only contains anomaly or background by spectral unmixing methods. In addition, abundance matrix after unmixing can provide more distinctive features to distinguish anomaly from background compared with original hyperspectral data. As a result, we introduce the unmixing algorithm into this paper before detecting with LRR. It is noteworthy that the number of endmembers is hard to estimate, so we get inspiration from eigenvalues methods and introduce KPCA[6] to make an estimation of this number.

Generally, to overcome these problems in hyperspectral anomaly detection caused by the highly mixture nature and interference of isolated noise, we propose a novel algorithm LU-LRR based on spectral unmixing and low-rank decomposition with a more robust dictionary. The main innovations of this algorithm can be summarized as follows:

1) We exploit KPCA to estimate number of endmembers. In addition, KPCA also plays a significant role in fully mining high-order correlation between spectral bands and reduce dimensionality to simplify computation.

2) The unmixing method is applied in this paper to project highly mixed data into a single subspace and extract more distinctive feature from abundance matrix.

3) The dictionary of LRR method can be constructed by dictionary learning method instead of original input data as the atoms of dictionary in conventional LRR. More precisely, the number of atoms in dictionary can be determined by mean-shift clustering method, standing for the number of all ground materials classes. RPCA is used to coarsely estimate background component to learn a more robust background dictionary without interference of anomaly.

Linear Unmixing and Low-Rank Representation (LU-LRR)

The general framework of this method can be shown in Fig. 1. This anomaly detection method can be divided into three major parts: performing KPCA to obtain the number of endmembers, linear spectral unmixing by MVC-NMF and conducting LRR to detect anomalous targets based on a more robust dictionary.

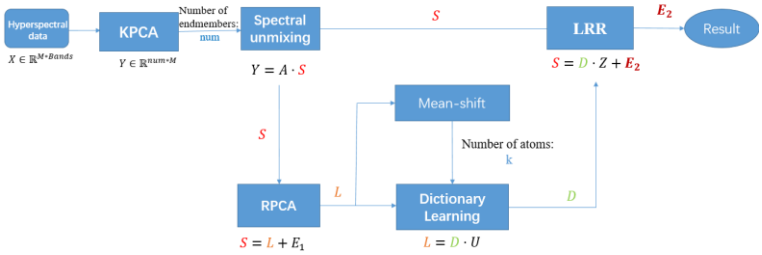

Figure 1. Overall framework of LU-LRR.

Spectral Unmixing Algorithm

The limitations of low spatial resolution and complex ground materials lead to the highly mixed nature in hyperspectral data. It means that there are some highly mixed pixels in hyperspectral images which contain anomaly and background partially. Unmixing methods aim at decomposing different ground materials (i.e. endmembers) from mixed pixels and obtaining mixing proportions of endmembers in each pixel (i.e. abundances).

Data mixture models used in hyperspectral unmixing algorithms are grouped into two categories: linear mixture model (LMM)[6] and nonlinear mixture model (NLMM)[7]. Linear mixture model is more popular in practical unmixing algorithm due to its simple structure and explicit physical meaning. LMM can be written as follows:

$$X = AS + N \tag{1}$$

Here, X is mixed hyperspectral data with each column standing for the spectrum of corresponding pixel, A is endmember matrix with num endmembers, S is abundance matrix, N is noise matrix.

According to whether there is endmember priori information or not, unmixing methods based on LMM can be furtherly divided into supervised unmixing methods and unsupervised unmixing methods. As mentioned above, anomaly detection lacks priori information, so we choose an unsupervised unmixing method MVC-NMF[9] to conduct spectral unmixing.

MVC-NMF combines Nonnegative Matrix Factorization (NMF)[8] with minimum volume constraint in endmembers which aims at minimizing volume of simplex. Constrained optimization problem, combining the goal of minimum reconstruction error with minimum simplex volume constraint in endmembers, can be written as:

$$\min f(A,S) = \frac{1}{2}\|Y - AS\|_F^2 + \mu J(A)\ s.t. A \geq 0, S \geq 0, 1_c^T S = 1_N^T \tag{2}$$

Here, Y is the output data of KPCA after nonlinear transformation and dimensionality reduction, 1_c (1_N) is a c (N)-dimensional column vector of all 1's. J(A) is the penalty function calculating the simplex volume determined by the estimated endmembers. The regularization parameter $\mu \in R$ is exploited to balance the tradeoff between the approximation error and the volume constraint. The minimum volume constraint makes MVC-NMF less sensitive to the initializations and more robust to noise corruptions.

[10] has proven that rows of abundance matrix corresponding to anomaly endmembers show distinctive patterns with three special characteristics: clustering in a small and compact region, high contrast and sparsity compared with noise pixels. It is simpler to distinguish anomaly from background and noise in abundance matrix rather than directly using original hyperspectral data.

Estimation of the Number of Endmembers Using KPCA

After spectral unmixing, mixed data have been projected into a single subspace. MVC-NMF needs to set the number of endmembers in advance. A classical and simple method in estimating number of endmembers is based on eigenvalues, which calculates eigenvalues of the covariance matrix after appropriate projection and retains the number of non-zero eigenvalues as intrinsic dimensionality.

The essence of KPCA is to map low-dimensional data to high-dimensional space through kernel trick and solve eigenvectors by implementing eigenvalue decomposition in kernel space. Therefore, benefiting from eigenvalues methods, we adopt KPCA to obtain covariance matrix and determine the number of endmembers according to distribution of eigenvalues.

The selection of kernel functions is crucial in KPCA. Gaussian Radial Basis Function (RBF) is widely popular due to its strong versatility. RBF can be written as:

$$\kappa_{RBF}(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{\sigma^2}) \tag{3}$$

Here, $\sigma > 0$ denotes the width of RBF.

However, RBF ignores effect of ground materials spectrum caused by natural factors including intensity of illumination and shadow. Interference of these factors will make the spectral energies of the same ground material quite different. As a result, we should consider spectral shape of ground materials in constructing a robust kernel function. Spectral Angle Mapper function (SAM) measures the difference of spectral shape by the intersection angle between spectral signals and is also robust to spectral energy. The kernel function of SAM can be formulated as follows:

$$\kappa_{SAM}(x_i, x_j) = cos^{-1}(\frac{x_i \cdot x_j}{\|x_i\| \cdot \|x_j\|}) \tag{4}$$

In this paper, we combine SAM with RBF in kernel function of KPCA algorithm to jointly take spectral energy and shape into consider. The mixed kernel function can be formulated as:

$$\kappa_{mix} = \beta \cdot \kappa_{RBF} + (1 - \beta) \cdot \kappa_{SAM} \tag{5}$$

Here, $\beta > 0$ is the tradeoff parameter between RBF and SAM kernel function.

Except for estimating the number of endmembers according to the number of non-zero eigenvalues i.e. num, KPCA is also exploited to reduce spectral dimensionality of input data X to num dimensions where num is smaller than spectral bands of X and carry out nonlinear transformation for deeply mining the nonlinear high-order spectral relevance in hyperspectral data.

LRR Based on Robust Dictionary Construction

LRR decomposes hyperspectral data into a low-rank matrix and a sparse matrix and introduces background dictionary into the corresponding low-rank component. Conventional LRR model can be formulated as follows:

$$\min_{Z,E}\{\|Z\|_* + \lambda\|E\|_{2,1}\} \ s.t. X = XZ + E \tag{6}$$

Here, X denotes original data, Z is the coefficient matrix and E is the sparse matrix only including anomalies. $\|\cdot\|_*$ denotes the nuclear form of a matrix, which is a convex relaxation of the rank function[11]. The $\ell_{2,1}$ norm minimization encourages the columns of E to be sparse. The tradeoff parameter $\lambda > 0$ is used to balance effects of the low-rank component and the sparse component.

However, using original hyperspectral data as the dictionary directly may cause three main problems: 1) X includes anomaly pixels, so using X as dictionary may cause anomaly contamination in background component; 2) a large hyperspectral image needs expensive computational cost to estimate Z and E; 3) applying input data itself as dictionary may cause sensitivity to the tradeoff parameter λ, which is difficult to choose an appropriate parameter in practice. One way for reducing computation burden is to randomly select some pixels from data to form the dictionary. Nevertheless, the dictionary should contain all kinds of background materials. Most of the scene is covered by a few kinds of major background materials, so random selection may ignore rare background materials of small probability in background component and leads to increase of the false alarm rate. [12] and [10] exploit clustering methods to estimate the combination of centroid vectors in each cluster as atoms in the dictionary, nevertheless, these methods are more or less subject to interference of anomaly. Considering aforementioned problems, a new strategy for dictionary construction is proposed.

Firstly, RPCA is exploited to roughly decompose abundance matrix S obtaining by spectral unmixing into low-rank component L and sparse matrix E_1:

$$S = L + E_1 \tag{7}$$

We construct a purer dictionary using the coarse background component L in which information of anomaly and noise are well suppressed for the purpose of eliminating interference of anomaly.

Secondly, we estimate the number of atoms in the dictionary using mean-shift clustering algorithm[13] which is a density-based non-parametric clustering method. Mean-shift method could automatically obtain the number of all ground materials classes excluding anomaly.

Lastly, a robust dictionary is constructed using Online Dictionary Learning method which is proposed in [14].The cost function of dictionary learning can be represented in the following form:

$$\min_{D,U} \frac{1}{2}\|L - DU\|_F^2 + \alpha\|U\|_1 \quad s.t. L = DU \tag{8}$$

Here, $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. L represents the low-rank component obtained from RPCA, D is dictionary and U is corresponding sparse coefficient matrix, α is regularization parameter. Further details of solving D and U can be found in Algorithm 1.

Algorithm 1 Online Dictionary Learning

Input: The background component after RPCA, L;
the regularization parameter, α;
the numbers of atoms in dictionary, k;
the number of iterations, T.

Step1: Initialization D: k column vectors randomly from L as atoms in initial D.
Step2: Sparse coding: compute using FISTA
  Take $X_0 = zeros(size(D,2), size(S,2))$
  $Y_1 \leftarrow X_0$, $t_1 \leftarrow 1$, Lipschitz constant $C \leftarrow max(eig(D^T \cdot D))$
  for $i = 1 : T$ **do**
    $x_i \leftarrow max(0, y_i - \frac{1}{C}\nabla y_i - \frac{\alpha}{C}) + min(0, y_i - \frac{1}{C}\nabla y_i + \frac{\alpha}{C})$
    $t_{i+1} \leftarrow \frac{1+\sqrt{1+4t_i^2}}{2}$
    $y_{i+1} \leftarrow x_i + \frac{t_i - 1}{t_{i+1}}(x_i - x_{i-1})$
  end for
  return U (learned sparse coefficient)
Step3: Update Dictionary
  $E = U \cdot U^T = [e_1, e_2, \cdots, e_k] \epsilon \mathbb{R}^{k \times k}$
  $F = S \cdot U^T = [f_1, f_2, \cdots, f_k] \epsilon \mathbb{R}^{M \times k}$
  repeat
  for $j = 1$ to k **do**
    update the $j - th$ column of D
    $a_j \leftarrow \frac{1}{F[j,j]}(e_j - D \times f_j) + d_j$
    $d_j \leftarrow \frac{1}{max(\|a_j\|_2, 1)} \times a_j$
  end for
  until convergence
  return D(updated dictionary)
Step4: **return** U,D

After constructing the robust dictionary D, we can solve sparse component E according to the above LRR model in this paper:

$$\min_{S,E}\{\|S\|_* + \lambda\|E\|_{2,1}\} \ s.t. S = DZ + E \tag{9}$$

We calculate the summation of each column in sparse matrix E and then reshape into the same spatial size of original hyperspectral data to get an anomaly image which larger values stand for the higher probability of existence of anomaly.

Experiment

Dataset

Two real hyperspectral images are used in this section to confirm the effectiveness of LU-LRR.

The first real hyperspectral image with the spatial size of 100 x 100 and 224 spectral bands was acquired by AVIRIS in San Diego airport. Fig. 2 (a) and (b) shows the image of sixth band in hyperspectral data and ground-truth image of anomaly target. The second test hyperspectral data obtained by the airborne HYDICE sensor with 162 bands. We choose a region whose spatial size is 80×100 pixels in the whole image with a size of 307×307 pixels. Fig. 2 (c), (d) and (e) shows the false-color image of the whole scene, the experiment area and ground-truth image of anomaly target.

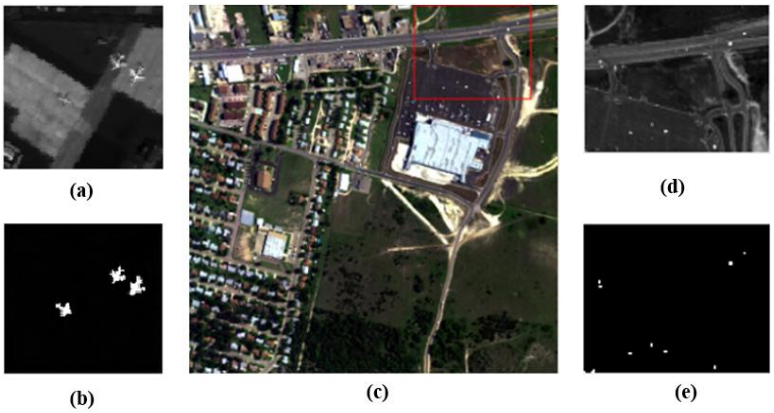

Figure 2. (a)AVIRIS image of the 6-th band; (b) ground-truth AVIRIS image; (c) false-color image of the whole scene in HYDICE image; (d) the 6-th band HYDICE image of selected area; (e) ground-truth image.

Performance Measure

To objectively evaluate results of different anomaly detection methods qualitatively and quantitatively, two common evaluation indexes are exploited in this section: Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC).

Table 1. AUC of different detection methods in test hyperspectral images.

Detection methods	Test image 1	Test image 2
Global RX[15]	0.839	0.841
CBAD[16]	0.922	0.974
RPCA[1]	0.790	0.916
ADLR[10]	0.904	0.801
LU-LRR	**0.970**	**0.985**

Here, we exploit four anomaly detection algorithms for comparison as follows: Global RX[15], CBAD[16], RPCA[1], ADLR[10]. Fig. 3 and Fig. 4 show ROC curves of different algorithms in first and second hyperspectral image respectively. Compared with other detection algorithms, LU-LRR, can distinguish all the anomalies while keeping a low false alarm rate. Table 1 reveals AUC values in two hyperspectral data. LU-LRR obtains the highest AUC values among five anomaly detection methods.

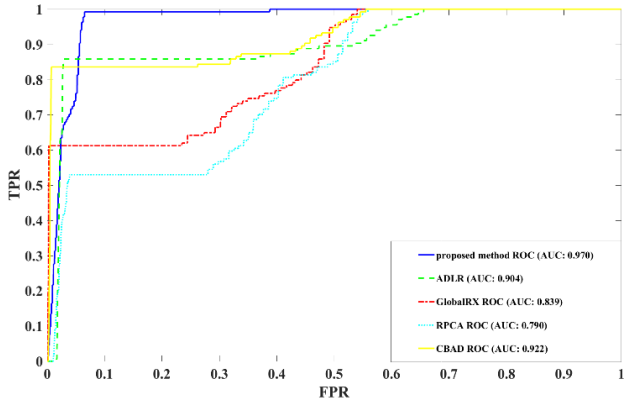

Figure 3. ROC curves of different detection methods in the first test data.

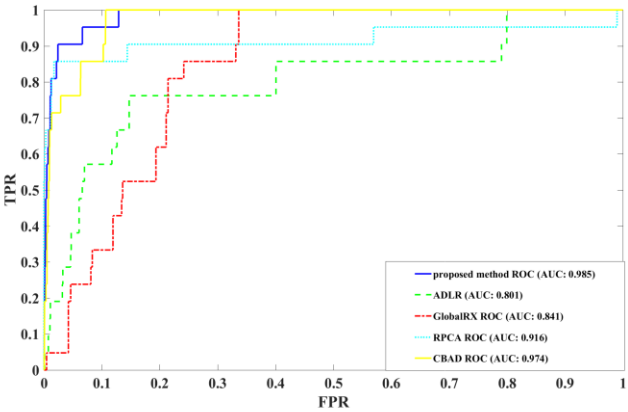

Figure 4. ROC curves of different detection methods in the second test data.

Conclusion

In this paper, we propose an effective anomaly detection method in hyperspectral imagery. This method utilizes KPCA to estimate the number of endmembers and mine high-order correlation among spectral bands. Then an unsupervised linear spectral unmixing method, MVC-NMF, is applied to obtain the abundance matrix, which effectively extracts features of clustered, high contrast and sparse anomaly targets. We introduce a robust dictionary in LRR to separate anomaly

from low-rank background component. This dictionary construction method solves anomaly contamination, expensive computation cost and sensitivity to tradeoff parameter compared with other dictionaries.

Acknowledgement

This research was financially supported by the National Science Foundation of China (Grant Nos.61527802).

References

[1] Zhao Q, Meng D, Xu Z, et al. Robust principal component analysis with complex noise[C]//International conference on machine learning. 2014: 55-63.

[2] Chen S Y, Yang S, Kalpakis K, et al. Low-rank decomposition-based anomaly detection[C]//Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX. International Society for Optics and Photonics, 2013, 8743: 87430N.

[3] Bitar A W. Exploitation of Sparsity for Hyperspectral Target Detection[D]. Centrale Supélec, Université Paris-Saclay, 2018.

[4] Zhou T, Tao D. Godec: Randomized low-rank & sparse matrix decomposition in noisy case[C]//Proceedings of the 28th International Conference on Machine Learning, ICML 2011. 2011.

[5] Liu G, Lin Z, Yu Y. Robust subspace segmentation by low-rank representation[C]//ICML. 2010, 1: 8.

[6] Schölkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 2008, 10(5):1299-1319.Hu Y H, Lee H B, Scarpace F L. Optimal linear spectral unmixing[J]. Geoscience & Remote Sensing IEEE Transactions on, 1999, 37(1):639-644.

[7] José M. P. Nascimento, José M. Bioucas-Dias. Nonlinear mixture model for hyperspectral unmixing[C]// Image and Signal Processing for Remote Sensing XV. International Society for Optics and Photonics, 2009.

[8] Lee D D, Seung H S. Algorithms for non-negative matrix factorization[C]//Advances in neural information processing systems. 2001: 556-562.

[9] Miao L, Qi H . Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2007, 45(3):765-777.

[10] Qu Y, Wang W, Guo R, et al. Hyperspectral Anomaly Detection Through Spectral Unmixing and Dictionary-Based Low-Rank Decomposition[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018:1-15.

[11] Wright J, Ganesh A, Rao S, et al. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization[C]//Advances in neural information processing systems. 2009: 2080-2088.

[12] Yang X, Wu Z, Li J, et al. Anomaly Detection in Hyperspectral Images Based on Low-Rank and Sparse Representation[J]. IEEE Transactions on Geoscience & Remote Sensing, 2016, 54(4):1990-2000.

[13] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002 (5): 603-619.

[14]Mairal J, Bach F, Ponce J, et al. Online Learning for Matrix Factorization and Sparse Coding[J]. Journal of Machine Learning Research, 2009, 11(1):19-60.

[15]Reed I S, Yu X . Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1990, 38(10):1760-1770.

[16]Carlotto, M. J. A cluster-based approach for detecting man-made objects and changes in imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(2):374-387.