ROBUST PRINCIPAL COMPONENT ANALYSIS FOR HYPERSPECTRAL ANOMALY DETECTION

Alex Sumarsono, Qian Du

Department of Electrical and Computer Engineering, Mississippi State University, USA

ABSTRACT

In this paper, we investigate the use of robust principal component analysis (RPCA) for anomaly detection. It is assumed that the resulting low-rank matrix corresponds to background, and the sparse matrix to targets. Intuitively, anomaly detection performance of a same detector on the sparse matrix is better than that on the original data matrix. To improve the efficiency of low rank and sparse matrix decomposition as well as the following anomaly detection, we propose to apply anomaly detection on the sparse matrix of each highly-correlated spectral segment, and then conduct decision fusion using the Choquet fuzzy integral, to produce the final detection output. The experimental result confirms the excellent performance of the proposed method.

***Index Terms*—**hyperspectral data, anomaly detection, robust principal component analysis, low rank and sparse matrix decomposition.

1. INTRODUCTION

Hyperspectral imagery (HSI) contains a wealth of spectral information, and target detection from HSI [1] has drawn lots of attention due to its important surveillance applications. With or without *a priori* knowledge, target detection algorithms can be generally grouped into two categories. When the target spectral signature is known, matched filter strategy is always considered. Known target spectral signature can come from the spectral library or available training samples. Spectral matched filter (SMF) [2] is a well-known detection approach using the spectral signature of target. The approach estimates covariance matrix of the background samples, and maximizes the signal-to-background ratio to distinguish targets from the background. Its extension, regularized SMF, has already been introduced in [3]. Subspace-based methods—matched subspace detector (MSD) [4] and adaptive subspace detector (ASD) [5]—use training data to estimate the covariance structure.

However, for practical reasons, the target spectral signature is always unknown, or untrustworthy due to additional noise. Under such a case, anomaly detection algorithms, which do not need any *priori* information, are commonly considered. The Reed-Xiaoli detector (RXD), which was introduced in [6], is actually based on the Mahalanobis distance between the input pixel and surrounding samples. The RXD has become the benchmark of anomaly detection algorithms. Two different versions of RXD have already been studied—Global RXD, which estimates the covariance matrix using the entire image, and Local RXD, which estimates the covariance matrix using local background samples (neighbors). In many literatures [3, 7], the local strategy has been proved to be better than the global one. Local RX employs a dual window strategy when calculating the Mahalanobis distance for each testing pixel. The inner window is slightly larger than the pixel size, the outer window is even larger than the inner one, and only samples in the outer region are adopted to estimate the covariance matrix.

Recently, robust principal component analysis (RPCA) [8] has been proposed for low-rank and sparse matrix decomposition. Typical PCA provides the optimal low-rank representation in an l_2-sense when data are corrupted as long as the magnitude of noise is small. However, it breaks down under arbitrary corruption which affects only very few of the additional observations [9, 10]. Comparatively, RPCA seeks to exactly recover underlying low-rank structure (background modeling) and a sparse structure (residual error) from the original data, even in the presence of large noise. Due to its characteristics, it is straightforward to apply RPCA for anomaly detection in hyperspectral imagery which usually assumes that the background image is low-rank while anomalies are preserved in the residual (sparse) image. As shown in [11], RPCA can produce accurate estimation of background. However, the sparse matrix from RPCA is not the detection output. Actually, it has the same size of the original data matrix. Thus, an anomaly detector, such as local or global RXD algorithm, should be applied. This is also suitable to the case when background (low rank matrix) and target (sparse matrix) separation is not ideal in practical applications. Intuitively, anomaly detection performance of a same detector on the sparse matrix is better than that on the original data matrix.

In this paper, we propose an approach to further improve the performance of RPCA-based anomaly detection. The basic idea is to spectrally partition the data into several highly-correlated groups, and operate RPCA to each spectral group, due to the fact that low rank and sparse decomposition is more efficient to highly-correlated spectral data. Anomaly detection is implemented in each group, and the final detection is the decision fusion output. The experimental result confirms the excellent performance of the proposed approach.

2. PROPOSED METHOD

2.1 Robust Principal Component Analysis

Let a data matrix $\mathbf{D} \in R^{m \times n}$ be composed of a low-rank matrix $\mathbf{A}_0$ and a small perturbation matrix $\mathbf{N}_0$:

$$\mathbf{D} = \mathbf{A}_0 + \mathbf{N}_0 . \tag{1}$$

Principal Component Analysis (PCA) can successfully recover $\mathbf{A}_0$ via the singular value decomposition (SVD) assuming the noise $\mathbf{N}_0$ is small. If this assumption is violated, i.e., $\mathbf{N}_0$ is replaced by $\mathbf{E}_0$ whose entries have arbitrary large magnitude, then the result of PCA could be quite inaccurate. However, this problem can be solved by Robust Principal Component Analysis (RPCA). If $\mathbf{A}_0$ is not sparse and the elements in $\mathbf{E}_0$ are uniformly random, then the low-rank matrix $\mathbf{A}_0$ and the sparse matrix $\mathbf{E}_0$ can be recovered by casting $\mathbf{D} = \mathbf{A}_0 + \mathbf{E}_0$ as a convex optimization problem [8]:

$$\begin{aligned} &\text{minimize} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \\ &\text{subject to}\, \mathbf{D} = \mathbf{A} + \mathbf{E} \end{aligned} \tag{2}$$

where λ is a positive regularizing parameter. For small problems, this convex optimization can be solved using off-the-shelf tools such as CVX. For large scale problems, a number of algorithms have been developed in recent years. Inexact Augmented Lagrange Multiplier (IALM) algorithm is chosen here due to its accuracy, stability and fast convergence [12].

2.2 Anomaly Detection

The sparse matrix **E**, which has the same size of the original data **D**, includes targets or anomalies, where background information is largely removed. To achieve anomaly detection, the well-known RXD algorithm can be applied to the **E** matrix, which can be expressed as:

$$\delta(\mathbf{r}) = (\mathbf{r} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \tag{3}$$

where **K** is data covariance matrix and $\boldsymbol{\mu}$ is the data mean vector [6].

The RXD basically evaluates the norm of a pixel after background whitening, and the one with very large norm is claimed as an anomaly. It works well when the data covariance matrix **K** can well represent background statistics. A modified algorithm, known as Uniform Target Detector (UTD), is defined as [13]:

$$\delta(\mathbf{r}) = (\mathbf{1} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \tag{4}$$

where **1** is the unity vector of length *L*, $(1,1,...,1)^{\mathrm{T}}$. The rationale behind choosing the unity vector is that the detector can be less sensitive to the pixel norm itself. It is particularly useful when image background is too complicated to be easily modeled by the **K**, resulting in less accurate background whitening.

2.3 Spectral Partitioning and Decision Fusion

It turns out the detector performance can be improved by employing the following methodology:

a) Partition the data into *N* spectrally correlated groups;
b) Perform RPCA on each group;
c) Apply anomaly detection to each sparse matrix; and
d) Fuse the detector outputs using Choquet fuzzy integral.

In other words, a hyperspectral image is first partitioned into *N* groups whose bands are highly correlated. Partitioning is based on the correlation coefficients of the spectral bands whose absolute value is greater than some threshold [14]. RPCA anomaly detection is performed on the sparse matrix of each group. The detector outputs are fused using the Choquet fuzzy integral to obtain a final image which is compared against the ground truth.

2.4 Choquet Fuzzy Integral

Fuzzy integral is a powerful tool to combine partial support from different sources of information such as the case in a multisensory image fusion. It is based on the concept of fuzzy measures [15]. A fuzzy measure g for a set X is a real-valued function $g: 2^X \to [0,1]$ which satisfies the following properties:

a) Boundary conditions: $g(\varnothing) = 0, g(X) = 1$;
b) Monotonicity: $A \subseteq B \Rightarrow g(A) \le g(B)$.

The discrete Choquet fuzzy integral of a function $h: X \to [0, \infty)$ with respect to fuzzy measure g is defined as [15]

$$C_g(h) = \sum_i h(x_{\pi(i)})(g(A_{\pi(i)}) - g(A_{\pi(i+1)})) \tag{5}$$

where π is a permutation of X such that $h(x_{\pi(1)}) \le ... \le h(x_{\pi(n)})$ and $A_{\pi(i)} = [x_{\pi(1)},...,x_{\pi(i)}]$. Among the aggregation operators used in practical applications of Choquet integral, the widely used is the ordered weighted averaging (OWA) defined by Yager [16]. It is given as

$$OWA = \sum_i w_i h(x_{\pi(i)}) \tag{6}$$

where $w_i = g(A_{\pi(i)}) - g(A_{\pi(i+1)})$. This operator allows the adjustment of the degree of "anding" and "oring" in the aggregation as a way to formulate multiple criteria decision. The weighting factor w satisfies the following properties: $w_i \in [0,1]$ and $\sum w_i = 1$.

For decision fusion of target detection in this research, $h(x_{\pi(i)})$ and w_i represent the detection result and the weight assigned to the i-th detector (i.e., i-th spectral segment), respectively.

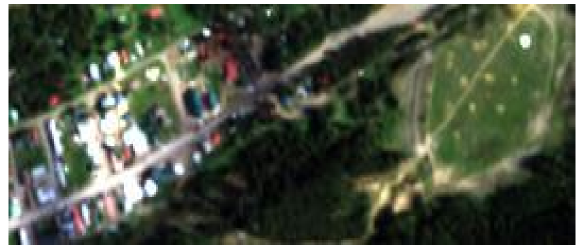
3. EXPERIMENTAL RESULTS

The real HyMap radiance image was captured over the Cooke City, Montana with approximately 3m spatial resolution. The dataset contains 126 bands covering 0.4-2.5 µm, and regions of interests (ROIs) of the seven targets are provided. 113 bands were kept for post processing after bad band removal. When converting radiance data from real numbers to integer, the data was first rescaled to preserve dynamic range, especially in the short wave infrared range. This rescaling involves multiplying bands 1 – 62 by 1000 and bands 63 – 126 by 4000. After scaling, the data is converted to integer format. The HyMap reflectance image is in units of reflectance factor scaled by 10,000. In the experiment, the original image data was cropped to a small one of size 100×300 as shown in Fig. 1. As shown in Fig. 1(b), there are 7 targets including 145 pixels.

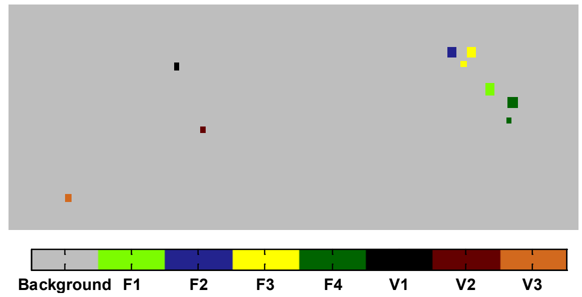
In Fig. 2, the data covariance matrix is displayed as a 2D image. Obviously, the data set can be partitioned into 4 groups: bands 1-15, 16-56, 57-85, and 86-113. The average of the absolute value of the correlation coefficients is 0.71, 0.99, 0.94, 0.90, respectively. The OWA operator is used with $w = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$ for the fuzzy measure.

In this paper, only the performance of global RXD and UTD is presented. As shown in Fig. 3, using the sparse matrix, the performance of RXD and UTD is improved, compared to the case of using the original image. The combination of RPCA sparse matrix + partitioning + fusion again shows better overall results. The overall performance improvement is due to the fact that spectrally correlated bands create stronger contrast between the background data and the targets (which should be spectrally different in the first place). This enables RPCA to better extract the targets into the sparse matrix which in turn makes it easier for a detector to identify the anomalies. Note that the performance of UTD is better than RXD, in particular, for low false alarm rates.

Table I further computes the area under ROC, and a larger value means better overall performance. For low false alarm rates, the proposed spectral partitioning and decision fusion with UTD can offer better performance.



(a) True color map of the image scene



(b) Ground truth map of the image scene

Fig. 1 Real data used in the experiment.

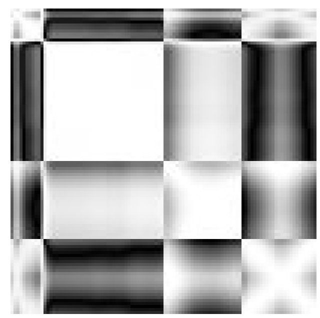
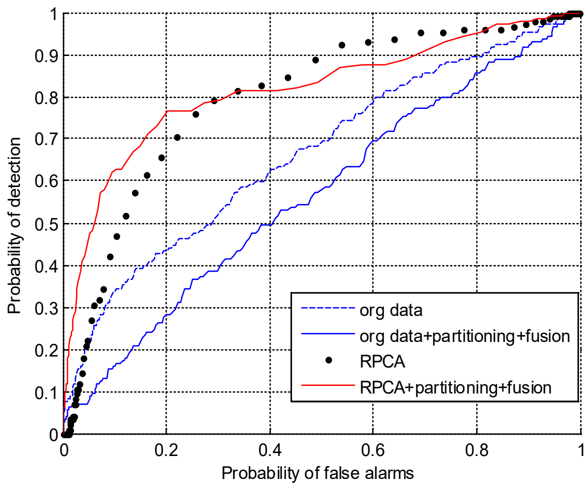


Fig. 2. Data covariance matrix displayed as a 2D image.

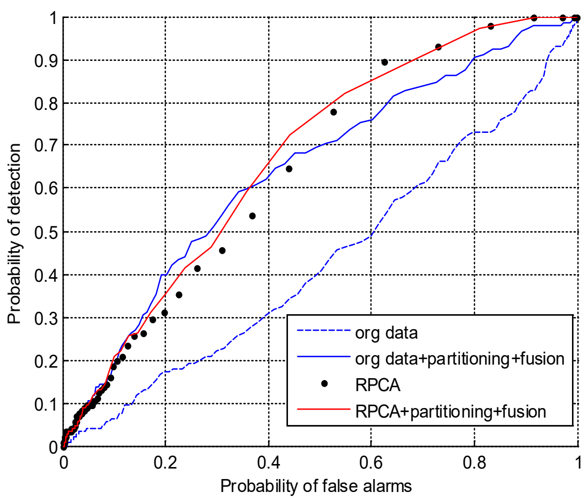
TABLE I
AREA UNDER ROC CURVES OF DIFFERENT DETECTION APPROACHES

	UTD	RXD
original data	0.6647	0.4318
original data, partitioning, then fusion	0.5666	0.6421
RPCA sparse matrix	0.7996	0.6555
Partitioning, RPCA sparse matrices, detection, fusion	**0.8167**	**0.6716**

It is also worth mentioning that when using the original data, the strategy of spectral partitioning and decision fusion does not necessarily produce better performance (as in Fig. 3(a)). This is because an entire spectral signature often contains more information for the separation of background and targets.



(a) UTD



(b) RXD

Fig. 3 ROC curves of different detection approaches.

4. CONCLUSIONS

In this paper, we investigate the use of sparse matrix from RPCA for anomaly detection. To improve the efficiency of low rank and sparse matrix decomposition as well as the following anomaly detection, we propose to apply anomaly detection on the sparse matrix of each highly-correlated spectral segment, and then conduct decision fusion using the Choquet fuzzy integral to produce the final detection output. The experimental result confirms the excellent performance of the proposed method. Interestingly, the improvement on a modified RXD, called UTD, is very significant.

5. REFERENCES

[1] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, January 2002.

[2] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 58–69, January 2002.

[3] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: Clutter adaption in anomaly detection," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1855–1871, August 2000.

[4] P. Bajorski, "Target detection under misspecified models in hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 470–477, April 2012.

[5] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, January 1992.

[6] I. S. Reed and X. Yu. "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech and Signal Process.*, vol. 38, no. 10, pp. 1760-1770, Ocotber 1990.

[7] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2146–2157, August 1994.

[8] E. J. Cand`es, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, May 2011.

[9] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," University of Illinois at Urbana-Champaign, Tech. Rep., October 2010.

[10] B. Bao, G. Liu, C. Xu, and S. Yan, "Inductive robust principal componenet analysis," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3794–3800, August 2012.

[11] W. Li and Q. Du, "Unsupervised nearest regularized subspace for anomaly detection in hyperspectral imagery," *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Melbourne, Australia, July 2013.

[12] Z. Lin, M. Chen, L. Wu. "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Mathematical Programming*, 2009.

[13] C.-I Chang and S.-S. Chiang. "Anomaly detection and classification for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol.40, no. 6, pp. 1314-1325, June 2002.

[14] X. Jia and J. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 538-542, January 1999.

[15] T. Havens, D. Anderson, J. Keller, "A fuzzy choquet integral with an interval type-2 fuzzy number-valued integrand," *Proceedings of IEEE International Conference on Fuzzy Systems*, 2010.

[16] R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decision making," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 1, pp. 183-190, January 1988.