A Selective Kernel PCA Algorithm for Anomaly Detection in Hyperspectral Imagery

Yanfeng GU, Ying LIU, and Ye ZHANG

Department of Information Engineering, Harbin Institute of Technology, Harbin 150001, China

ABSTRACT

In this paper, a selective kernel principal component analysis algorithm is proposed for anomaly detection in hyperspectral imagery. The proposed algorithm tries to solve the problem brought by high dimensionality of hyperspectral images in anomaly detection. This algorithm firstly performs kernel principal component analysis (KPCA) on the original data to fully mine high-order correlation between spectral bands. Then, high-order statistics in local scene are exploited to define local average singularity (LAS), which is used to measure the singularity of each nonlinear principal component transformed. Based on LAS, one component transformed with maximum singularity is selected after KPCA. Finally, with RX detector, anomaly detection is performed on the component selected. Numerical experiments are conducted on real hyperspectral images collected by AVIRIS. The results prove that the proposed algorithm outperforms the conventional RX algorithm.

1. INTRODUCTION

Hyperspectral remote sensing exploits the fact that all materials reflect, absorb, and emit electromagnetic energy, at specific wavelengths, in distinctive patterns related to their molecular composition [1]. Thus, data exploitation of hyperspectral images (HSI) makes it possible to identify ground materials of interest based on their spectral signatures, such as target recognition and anomaly detection [1-3]. Relative to target detection, anomaly detector enables one to detect targets whose signatures are spectrally distinct from their surroundings with no a priori knowledge [4].

Anomaly detection is a pattern recognition scheme that is used to detect objects that might be of military interest. In spectral anomaly detection algorithms, materials that have a significantly different spectral signature from their neighboring background clutter pixels are identified as spectral anomalies. Spectral anomaly detection algorithms could also use spectral signatures to detect anomalies embedded within background clutter with a very low signal-to-noise ratio (SNR) [2,4]. A widely used anomaly detector for multispectral data, i.e., developed for sensors with a reduced number of spectral bands (ten or less) is the algorithm presented in [5], commonly referred to as the RX algorithm, after the initials of its proponents, Reed and Xiaoli Yu. The RX algorithm is a likelihood ratio detector based on a number of simplifying assumptions. Extending the RX algorithm to hyperspectral imagery suffers from two major limitations [6,7]. In substance, these limitations are both caused by the high dimensionality of the hyperspectral images.

In this paper, a selective kernel principal component analysis algorithm is proposed to improve performance of the conventional RX algorithm. In the proposed algorithm, KPCA is used to transform hyperspectral images and fully mine high-order correlation between spectral bands. According to high-order statistics in local scene, LAS is defined to select the most singular component transformed, which is the most effective one for anomaly detection. Final anomaly detection is performed with the selected component. The effectiveness of the proposed algorithm is proven in numerical experiments.

2. SELECTIVE KERNEL PCA

The main contribution of the proposed algorithm is to concentrate information about anomalous targets and select the component with maximum singularity for final detection. The proposed algorithm includes three techniques, i.e. KPCA, defining local average singularity (LAS) and RX detector.

2.1. Kernel PCA

KPCA has excellent ability to process nonlinear data [8,9]. Given a set of centered random samples $y_k(k=1,2,\cdots,l)$, $y_k \in \mathcal{R}^d$, $\sum_{k=1}^{l} y_k = 0$. Firstly, the sample set $\{y_k\}$ is mapped into a feature space $\mathcal{R}^{\mathcal{X}}$ by $\phi: \mathcal{R}^d \to \mathcal{R}^{\mathcal{X}}$ and the covariance matrix Σ_ϕ is computed in the feature space $\mathcal{R}^{\mathcal{X}}$,

$$\Sigma_\phi = \frac{1}{l}\sum_{j=1}^{l} \phi(y_j)\phi(y_j)^T \tag{1}$$

Let $\mathbf{V}$ $(\mathbf{V} \neq 0)$ be an eigenvector of Σ_ϕ that corresponds to a positive eigenvalue λ of Σ_ϕ. So the eigenvector is in

This paper is supported by National Natural Science Foundation of China (project No.60402025 and 60472048)

the space spanned by the mapped samples, i.e. $\mathbf{V} \in span\{\phi(y_1), \phi(y_2), \cdots, \phi(y_l)\}$. This can be described as

$$\mathbf{V} = \sum_{i=1}^{l} \beta_i \phi(y_i) \tag{2}$$

Thus, eigenvalue decomposition can be written as

$$\lambda \cdot \mathbf{V} = \sum_{\phi} \cdot \mathbf{V} \tag{3}$$

Furthermore, we multiply both sides of equation (3) by $\phi(y)$ and obtain the following equation as

$$\lambda(\phi(y) \cdot \mathbf{V}) = (\phi(y) \cdot \sum_{\phi} \cdot \mathbf{V}) \tag{4}$$

For all l eigenvectors, equation (4) can be also written as

$$\lambda(\phi(x) \cdot \mathbf{V}^k) = (\phi(x) \cdot \sum_{\phi} \cdot \mathbf{V}^k) \tag{5}$$

where $k = 1,2,\cdots,l$, $\mathbf{V}^k = \sum_{i=1}^{l} \beta_i^k \phi(\mathrm{y}_i)$. According to construction of the kernel function, define an $l \times l$ dimensionality of kernel matrix $\mathbf{K}$ in feature space as

$$(\mathbf{K})_{i,j} = (\phi(y_i) \cdot \phi(y_j)) = k(y_i, y_j) \tag{6}$$

and consider an eigenvalue decomposition for the expansion coefficients β^k by using kernel matrix $\mathbf{K}$ as

$$\lambda \beta^k = \mathbf{K} \beta^k, \left(\beta^k = (\beta_1^k, \beta_2^k, \cdots, \beta_l^k)^T \right) \tag{7}$$

The obtained solution (λ_k, β^k) has to be normalized by imposing $\lambda_k(\beta^k \cdot \beta^k) = 1$ and to be centered by substituting centered $\mathbf{K}_c$ for the $\mathbf{K}$. The $\mathbf{K}_c$ is given by

$$\mathbf{K}_c = \mathbf{K} - 1_l \cdot \mathbf{K} - \mathbf{K} \cdot 1_l + 1_l \cdot \mathbf{K} \cdot 1_l \tag{8}$$

where 1_l is an $l \times l$ matrix, of which all elements are equal to $1/l$. To obtain a new feature of the samples, the processing, which projects the mapped sample $\phi(y)$ on vector $\mathbf{V}^k$, is necessary, namely,

$$(\mathbf{V}^k \cdot \phi(y)) = \sum_{i=1}^{l} \beta_i^k (\phi(y_i) \cdot \phi(y)) = \sum_{i=1}^{l} \beta_i^k k(y_i, y) \tag{9}$$

Notice that the original hyperspectral images are data cube. The data have to be firstly converted to two-dimensionality form before they are transformed. Each band of the hyperspectral images is viewed as a sample vector, and number of total samples for transform is equal to number of original bands. Thus, after these samples are centered and normalized, KPCA is performed on them to fully mine high-order correlation between samples (i.e. spectral bands).

2.2. Definition of LAS

The third- and the four-order moment, i.e. skewness and kurtosis, are used to measure symmetry of data distribution and trailing respectively. Given a random variable t, let $\hat{m}$ and $\hat{\sigma}$ be mean and variance estimations of D observations t_i $(i = 1,2,\cdots,D)$ respectively. The skewness is defined as

$$\gamma_3 = \mathrm{E}\{[t - \mathrm{E}(t)]^3\} / (\mathrm{E}\{[t - \mathrm{E}(t)]^2\})^{3/2} \tag{10}$$

The estimation of the skewness can be written as

$$\hat{\gamma}_3 = \sum_{i=1}^{D} (t_i - \hat{m})^3 / (N-1)\hat{\sigma}^3 \tag{11}$$

The kurtosis is defined as

$$\gamma_4 = \left(\mathrm{E}\{[t - \mathrm{E}(t)]^4\} / (\mathrm{E}\{[t - \mathrm{E}(t)]^2\})^2 \right) - 3 \tag{12}$$

and its estimation is written as

$$\hat{\gamma}_4 = \left(\sum_{i=1}^{D} (t_i - \hat{m})^4 / (N-1)\hat{\sigma}^4 \right) - 3 \tag{13}$$

According to the skewness and the kurtosis, we define the LAS to measure the singularity of each nonlinear principal component one by one and denote it by symbol $\overline{N}_{\mathrm{LAS}}$. The detailed calculation of $\overline{N}_{\mathrm{LAS}}$ for one nonlinear component is as follows

1) Construct a local sliding window with size $n_1 \times n_2$, let moving interval be equal to half width of the window and let C be the total sliding times
2) Let $\overline{N}_{\mathrm{LAS}} = 0$ and $j = 1$
3) Calculate absolute values A_s^j and A_k^j of the skewness and the kurtosis in the $j-$th local window
4) Compare A_s^j and A_k^j with two thresholds T_s^j and T_k^j respectively; if both A_s^j and A_k^j are larger than T_s^j and T_k^j, then let $\overline{N}_{\mathrm{LAS}} = \overline{N}_{\mathrm{LAS}} + 1$; else hold $\overline{N}_{\mathrm{LAS}}$ invariant
5) If $j > C$, then the calculation is over and note the final value of $\overline{N}_{\mathrm{LAS}}$; else let $j = j + 1$ and return to 3)

In the anomaly detection, if data in the local windows follow Gaussian distribution, then the corresponding skewness and kurtosis are both equal to zero. If there are anomalies in the local windows, the Gaussian distribution is broken and the absolute values of the skewness and the kurtosis become large. In the calculation of LAS above, two thresholds are defined as follow

$$T_s^j = \tau_s \theta_s \text{ and } T_k^j = \tau_k \theta_k \tag{14}$$

where τ_s and τ_k are equal to $\sqrt{6/(n_1 \times n_2)}$ and $\sqrt{24/(n_1 \times n_2)}$ respectively, $n_1 \times n_2$ is the size of the local window, θ_s and θ_k are constant and are generally set to 11. If the absolute values of the skewness and the kurtosis are both larger than T_s^j and T_k^j, then this means that the data distribution in the local window is not Gaussian.

2.3. RX detector

In the conventional RX algorithm, two hypothesis are adopted as follow

$$H_0: \quad x = n$$

$$H_1: \quad x = s + n \tag{15}$$

where x is an observation test vector of single pixel target x, s is the spectral signature of the signal (target), and n is a vector that represents the background clutter noise process. H_0 represents that target is absent from the local scene and the local scene follows Gaussian distribution with zero-mean. H_1 represents that target is present in the scene and the local scene follows Gaussian distribution with such mean as being equal to target gray. Thus, under H_0, the data (local background clutter) are modeled as $N(0,\Gamma_x)$, and under H_1 the data are modeled as $N(s,\Gamma_x)$. In practice, the target signature s and background covariance Γ_x are assumed to be unknown. In the conventional RX algorithm, it is assumed that the background and target have the same covariance matrix in the model as shown in (15). Generally, the conventional RX detector is given by

$$RX(x) = (x-\hat{\mu})^T \left(\hat{\Gamma}_x\right)^{-1} (x-\hat{\mu}) \underset{<}{\geq} \eta \tag{16}$$

where $\hat{\mu}$ is the estimated background clutter sample mean, $\hat{\Gamma}_x$ is the background covariance estimated from the local background clutter data, and η is a detection threshold. If the detection result in the test pixel vector is bigger than or equal to η, there is target in this pixel; else there is no target in this pixel.

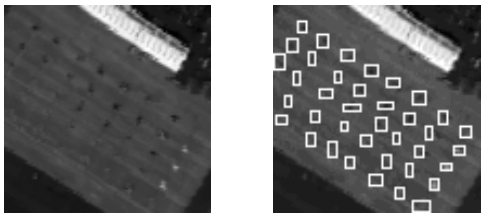
3. EXPERIMENTS

3.1. Data description

In order to verify the effectiveness of the proposed method for anomaly detection in hyperspectral imagery, the numerical experiments are performed on real hyperspectral data collected by AVIRIS. After bands that correspondings to the water absorption regions, low SNR and bad bands are removed, 126 available bands remain in the 0.4–1.8 μm wavelength range. The ground sampling distance of those hyperspectral images is 3.5m. A scene of 100×100 pixels was selected for our experiments, in which there are 38 panels as anomalous targets for our detection. Fig.1 (a) shows the sixth band of the image scene used in our experiments and Fig.1 (b) shows the ground distribution of 38 anomalous targets.

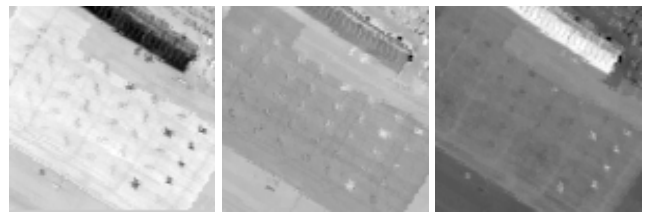
3.2. Experimental results

In KPCA, the Gaussian radial basis function (RBF) $k(x,y) = \exp\left\{-(x-y)^2/2\sigma^2\right\}$ was used as the nonlinear kernel. The choice of the parameter is very critical. It should be chosen such that the overall data variations can be fully exploited by the Gaussian RBF kernel. In this paper, the value of σ was determined experimentally and was set to 70. Fig.2 show the first six nonlinear principal components transformed by KPCA. It is easy to see that the fourth component as shown in Fig.2 (d) includes almost all information about these anomalous targets.

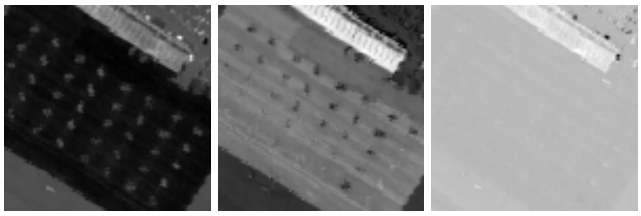


(a) sixth band (b) ground distribution

Fig.1 The sixth band of hyperspectral images and corresponding distribution of targets for anomaly detection experiments



(a) (b) (c)



(d) (e) (f)

Fig.2 The first six components transformed by KPCA

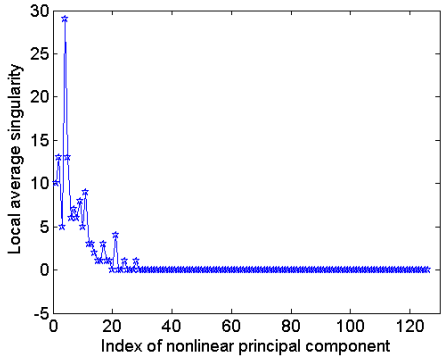
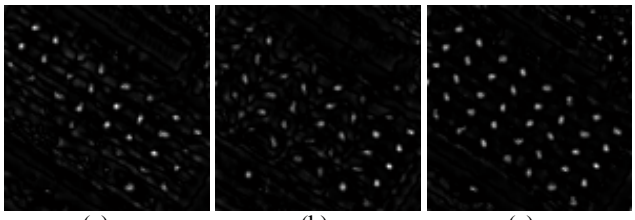


Fig.3 The varied curve of LAS with nonlinear component index

Fig.3 provides the varied curve of LAS accompanying with each nonlinear principal component. In this curve, the 4-th nonlinear principal component is the one with maximum LAS. So the fourth component was used in RX anomaly detection. In our experiments, the proposed algorithm is compared with the conventional RX algorithm.

In addition, the first component that has maximum energy was also used in anomaly detection to prove the effectiveness of LAS rule. Fig.4 provides the detection results from the conventional RX algorithm, the anomaly detection with the first nonlinear component transformed, and the proposed selection-based anomaly detection algorithm. It is easy to find that the detection result obtained by the proposed algorithm greatly outperforms other methods.



(a) (b) (c)

Fig.4 The comparison of detection results of three methods. (a) is of the conventional RX algorithm, (b) is of the detection with the first component transformed, and (c) is of the proposed algorithm

To quantitatively evaluate the three algorithms, their receiver operation characteristics (ROC) were researched in the experiments. The ROC represents varying relationship of detection probability and false alarm rate, and can provide quantitative comparison of the detection performance. Fig.5 shows the ROC comparison of the conventional RX algorithm and the proposed algorithm. From Fig.5, it is easy to prove that the proposed algorithm greatly improves the detection performance of the conventional RX algorithm and has low false alarm rate. Fig.6 provides the ROC comparison of the detection with the selected component and the detection with the first component transformed. Fig.6 forcefully proves the effectiveness of the anomalous component selection method based on LAS.

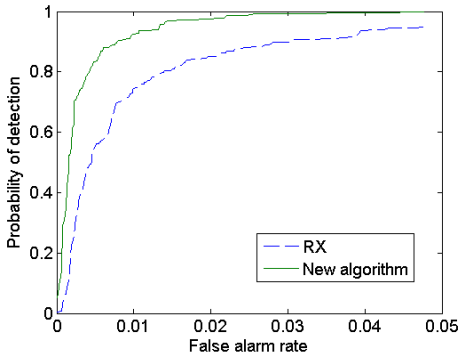


Fig.5 The ROC comparison of RX and the proposed algorithm

4. CONCLUSIONS

In this paper, a selective kernel PCA algorithm is proposed for anomaly detection in hyperspectral imagery. By using KPCA, this algorithm effectively concentrates almost all information about anomalous targets and fully mines the high-order correlation between spectral bands. A more important contribution of the proposed algorithm consists in that the LAS is defined by using the high-order statistics. With the LAS, the component with maximum singularity is properly selected, which is the most useful for anomaly detection. The experimental results prove that the proposed algorithm greatly improves the performance of the conventional RX algorithm.

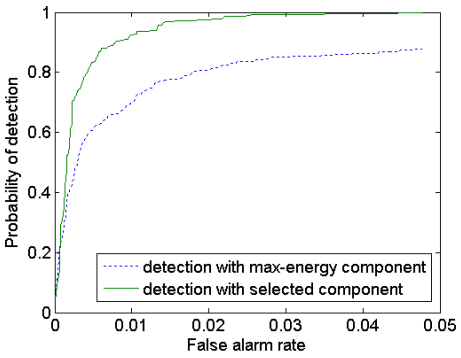


Fig.6 The ROC comparison of detection with maximum energy component and detection with the selected component

5. REFERENCES

[1] G.Shaw and D.Manolakis, "Signal Processing for Hyperspectral Image Exploitation," IEEE Signal Process. Mag., pp. 12-16, January 2002

[2] David W.J.Stein, Scott G.Beaven, Lawrence E.Hoff et al, "Anomaly Detection from Hyperspectral Imagery," IEEE Signal Process. Mag., pp. 58-69, January 2002

[3] David Landgrebe, "Hyperspectral Image Analysis," IEEE Signal Process. Mag., pp.17-28 January 2002

[4] C.-I Chang and Shao-Shan Chiang, "Anomaly Detection and Classification for Hyperspectral Imagery," IEEE Trans. Geosci. Remote Sens., vol. 40, no. 6, pp. 1314-1325, June 2001

[5] 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

[6] H.Kwon and N.M.Nasrabidi, "Kernel RX algorithm: A Nonlinear Anomaly Detector for Hyperspectral Imagery," IEEE Trans. Geosci. Remote Sens., vol. 43, no. 2, pp. 388–397, Feb 2005

[7] S.M.Schweizer and J.M.F.Moura, "Hyperspectral Imagery: Clutter Adaptation in Anomaly Detection," IEEE Trans. Information Theory, vol. 46, no. 5, pp. 1855–1871, Aug. 2000

[8] B. Schökopf, A. J. Smola, and K.-R. Müller, "Kernel principal component analysis," Neural Comput., no. 10, pp. 1299–1319, 1999

[9] A. Ruiz and E. Lopez-de Teruel, "Nonlinear kernel-based statistical pattern analysis," IEEE Trans. Neural Networks, vol. 12, no. 1, pp. 16–32, Jan. 2001